

令和 6 年 6 月 3 日現在

機関番号：24405

研究種目：若手研究

研究期間：2020～2023

課題番号：20K19886

研究課題名（和文）ラフ集合理論に基づく不確実性を考慮したクラスタリング手法の新展開

研究課題名（英文）New Development of Clustering Methods Considering Uncertainty Based on Rough Set Theory

研究代表者

生方 誠希（Ubukata, Seiki）

大阪公立大学・大学院情報学研究科 ・准教授

研究者番号：10755698

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：RCMdd法の改良をはじめとして、ラフクラスタリング手法の種々の改良に取り組み、頑健かつ柔軟な分類を実現した。また、ラフ協調フィルタリングとしてRCM-CFやRSCM-CF、RMCM-CF、さらに、ラフ共クラスタリングベースの協調フィルタリングとしてRCCMM-CFやRSCCMM-CF等を提案した。これらにより、実データに内在する不確実性に対処し、推薦性能を向上させることができた。また、逐次データを読み込みパラメータを更新するオンライン学習やミニバッチ学習を取り入れ、さらに、クラスター数を適応的に調節する機構の導入や次元削減の活用により、大規模データへの適用を可能とした。

研究成果の学術的意義や社会的意義

本研究では、ラフ集合理論に基づいて不確実性を考慮したクラスタリング手法の改良やその協調フィルタリングへの応用に取り組み、実データを用いた数値実験を通してそれらの有効性を実証した。提案するラフクラスタリングやラフ協調フィルタリングにより、実世界の多様なデータセットにおいて不確実性を考慮した適切な分類が可能となり、ユーザーが探しているコンテンツへのアクセスを容易にするなど、我々の生活をより豊かにするアプローチを提供した。

研究成果の概要（英文）：By improving RCMdd and various rough clustering techniques, robust and flexible clustering has been achieved. Additionally, rough collaborative filtering (CF) methods such as RCM-CF, RSCM-CF, and RMCM-CF have been proposed. Furthermore, rough co-clustering-based CF methods like RCCMM-CF and RSCCMM-CF have been proposed. These approaches address the inherent uncertainties in real-world datasets, enhancing recommendation performance. Moreover, by incorporating online learning and mini-batch learning, which sequentially read data and update parameters, as well as introducing mechanisms for adaptively adjusting the number of clusters and utilizing dimensionality reduction, the application to large-scale datasets has been made possible.

研究分野：ソフトコンピューティング

キーワード：ソフトコンピューティング ファジィ理論 ラフ集合理論 クラスタリング ファジィクラスタリング
ラフクラスタリング 共クラスタリング 協調フィルタリング

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

高度情報化社会において、インターネットやスマートフォンの普及に伴い、膨大なデータ(ビッグデータ)が蓄積されるようになり、その活用が求められている。近年では、自動的に特徴量の学習から判断までを行う深層学習が種々のタスクで人間を凌駕する分類性能を達成し、脚光を浴びている。一般的な深層学習のタスクでは膨大なデータに対して分類の教師情報が必要となり、多大なコストがかかる問題がある。膨大な「教師なし」のデータの活用が求められており、教師なしで自動的に類似した対象をグループ化するクラスター分析(クラスタリング)の需要が増している。伝統的なクラスタリング手法として HCM (k-means)法が広く活用されてきた。HCM 法はハードな分割を構成し、クラスターの境界付近の、本来帰属が不確実である対象についても強制的にいずれかのクラスターへ割り当てるため、誤分類のリスクが高いという問題がある。実世界のデータではクラスター構造は必ずしも明瞭でなく、対象がいずれのクラスターに帰属するかを一意に決定できない、不確実性の取り扱いが課題となる。不確実性を取り扱い、誤分類を低減するために、ファジィ理論やラフ集合理論等のソフトコンピューティングアプローチが有効である。ファジィ理論に基づく FCM 法を基礎としたファジィクラスタリングは世界中で研究が推進され、広く活用されている一方で、ラフ集合理論の観点を導入した RCM 法等のラフクラスタリングは十分に研究が進んでいない。筆者はこれまでに、RCM 法の基礎理論を整理することで RCM 法での不確実性の取り扱いの本質を明らかにし、理論的価値を高めた。一方、RCM 法ではラフ集合理論において最も重要な観点の一つである対象空間の粒状化が考慮されず、粒状化された空間の下で不確実性を扱うための近似概念(ラフ近似)が使用されないという問題があった。応募者はラフ集合理論に基づき、粒状化された空間において不確実性を考慮しながらクラスタリングを実行する RSCM 法を提案し、理論的優位性に加えて、分類精度の向上を実証した。

2. 研究の目的

膨大なデータを教師なしで自動的に分類するクラスタリングにおいて、クラスター割り当ての不確実性の扱いは重要な課題であり、ファジィクラスタリングの他、ラフ集合理論に基づくラフクラスタリングの研究が推進されている。本研究では、筆者が基礎理論を確立したラフクラスタリングの新たな展開を図る。目的関数の定式化によって基礎理論を補強し、ノイズロバストなアプローチを導入することで、頑健性と実用性を高める。さらに、文書解析での文書×単語の共起情報やマーケット分析での顧客×商品の購買履歴などの対象と項目の共起関係データから関連性の強い対象と項目からなる共クラスタ構造を抽出する共クラスタリングに対しラフクラスタリング手法を応用することで、ラフ共クラスタリングとラフ協調フィルタリングを提案し、共クラスタリング・協調フィルタリングにおけるラフ集合理論に基づく粒状性と不確実性の取り扱いの影響を明らかにする。

3. 研究の方法

(3.1) クラスタリング手法の改良

(3.1.1) Rough C-Medoids 法の改良

RCM 法ではクラスター領域のプロトタイプとして重心を使用するが、重心は二乗誤差の最小化に基づいているため、外れ値の影響を受けやすいという問題がある。そこで、クラスター領域のプロトタイプとしてメドイド(クラスター内の対象との誤差を最小化する対象)を使用した Rough C-Medoids (RCMdd)法が提案されている。本研究では、RCMdd 法に対し、粒状性の観点を導入し、RSCMdd 法を提案した。さらに、Euclid 距離を一般化した Minkowski 距離を導入した手法を開発し、柔軟かつ頑健なクラスタリングを目指した。

(3.1.2) ラフ共クラスタリングの提案

「対象×項目」の共起関係データの分析には共クラスタリングのアプローチが有効であり、代表的なファジィ共クラスタリングとして FCCMM 法などがある。本研究では、FCCMM 法を参考に、共クラスタリングにラフ集合理論の観点を導入し、ラフ共クラスタリングを提案した。通常のクラスタリングにおいてハードクラスタリングである HCM 法からラフクラスタリングである RCM 法を導くラフ化の仕組みを、FCCMM 法をクリस्प化した HCCMM 法に適用することで、ラフ共クラスタリングとして RCCMM 法を提案した。さらに、RCCMM 法に対し、粒状性の観点を導入し、RSCCM 法を提案した。

(3.2) 協調フィルタリングへの応用

(3.2.1) ラフ協調フィルタリングの提案

クラスタリングベースの協調フィルタリングにラフ集合理論の観点を導入し、ラフ協調フィルタリングを提案した。ラフクラスタリングである RCM 法を使用した協調フィルタリングとして RCM-CF を開発し、さらに、粒状性を考慮したラフクラスタリングである RSCM 法や RMCM 法を使用した発展的手法として RSCM-CF や RMCM-CF を開発した。また、様々な粒状化手法を検討し、

その効果を検証した。加えて、RMCM-CF に目的関数を導入した RMCM2-CF を提案した。さらに、確率的ラフ近似の概念を導入した Probabilistic RSCM 法に基づく CF として、PRSCM-CF を提案し、より柔軟な近似に基づいて推薦を行う手法を開発した。

(3.2.2) ラフ協調フィルタリングの種々の改良

RCM-CF において、アンサンブル学習を導入し、多様なパラメータ設定による複数のクラスタリング結果に基づく推薦度を統合する方式を提案し、多角的な推薦を行う手法を開発した。協調フィルタリングタスクにおけるデータは多数の欠測値を含んでおり、従来法では本来の値とは異なる単一の値を代入することによりデータに歪みが生じる問題があった。それに対処するため、欠測値を含む次元を無視する部分的距離戦略を採用した RCM 法に基づく CF を提案した。RSCM 法においてノイズ除去機構を導入した Noise RSCM 法に基づく CF を提案し、ノイズロバストな推薦を目指した。また、ユーザーではなくアイテムの類似性に基づき推薦を行うアイテムベースのアプローチの効果を検証した。

(3.2.3) ラフ共クラスタリングベースの協調フィルタリング

協調フィルタリングタスクにおけるデータは「ユーザー×アイテム」の共起関係データと考えられるため、ユーザーとアイテムを同時にクラスタリングする共クラスタリングのアプローチが有効であると考えられる。そこで、代表的なファジィ共クラスタリングである FCCMM 法を参考に、ラフ共クラスタリングとして RCCMM 法や RSCCMM 法を提案している。本研究では、ラフ共クラスタリングに基づく協調フィルタリングとして、RCCMM-CF や RSCCMM-CF を提案した。

(3.2.4) 大規模データへのアプローチ

従来のラフクラスタリング手法は、すべてのデータを一括で読み込んで分析を行うバッチ学習として実装されているため、メモリの容量を超えたデータや、逐次新しいデータが到着するような状況に対応できない。そこで、大規模データへの適用に向けて、RCM-CF に対しオンライン学習のアプローチを導入し、新しいデータを一行ずつ読みこんでクラスタ中心を徐々に更新する ORCM-CF を提案した。また、クラスタ数を適応的に調節する機構を導入した Adaptive ORCM-CF を提案し、利便性を向上させた。さらに、AORCM 法に対し、ミニバッチ学習を導入することで、AMBRCM 法とその CF への応用を提案し、計算効率や推薦性能の向上を実現した。その他、大規模データへのアプローチとして、ランダム射影によってデータの次元削減を行った上で類似ユーザーの分類を行う RP-RCM-CF を提案した。さらに、主成分分析や潜在意味解析、autoencoder、t-SNE 等の種々の次元削減手法を活用した RCM-CF を提案した。

4. 研究成果

上記の種々の提案について、実データを用いた数値実験を通して、その特性や効果を検証した。

Rough C-Medoids 法の改良によって、外れ値の影響を低減し、さらに不確実性を考慮することで、頑健かつ柔軟なクラスタリングを実現できることを確認した。ラフ共クラスタリングの提案によって「対象×項目」の共起関係データの分析において、従来のベクトルデータ用の手法と比較して、より適切に共クラスタ構造を抽出できることを確認した。ラフ集合理論の導入によって、オーバーラップした共クラスタを抽出することができ、実データに内在する不確実性に対処することができる。本課題を通して、ラフ共クラスタリングの理論的基礎を構築することができた。ラフ協調フィルタリングの提案によって、MovieLens データセット等の協調フィルタリングのベンチマークデータセットにおいて、従来のハードなクラスタリングに基づく手法よりも推薦性能を高められることが確認できた。これにより、ユーザーによる映画等のアイテムの評価データにおいて、オーバーラップ構造が存在することが示唆された。また、粒状化を考慮した手法が高い推薦性能を達成したことから、協調フィルタリングタスクにおいて、粒状性が重要な要素であることが示唆された。また、ラフ協調フィルタリングの種々の改良を通して、推薦性能を向上させることができ、ラフ協調フィルタリングの実用性を高めることができた。ラフ共クラスタリングベースの協調フィルタリングによって、推薦性能の向上が確認できたため、「ユーザー×アイテム」の共起関係データに共クラスタリングのアプローチが有効であり、さらに、ラフ集合理論によるクラスタのオーバーラップの取り扱いが有効であることがわかった。大規模データへのアプローチとして、データを逐次読み込む方式を導入し、バッチ学習と比較して推薦性能をある程度維持できたことから、大規模データでの有効性が示唆された。また、次元削減を活用した手法においても推薦性能をある程度維持できたため、計算コストの低減が可能となった。

ラフ集合理論に基づくクラスタリング手法とその応用により、データの不確実性を適切に処理し、クラスタリング精度および推薦性能の向上を実現した。特に、粒状性を考慮したクラスタリング手法により、ノイズロバストなクラスタリングと推薦が可能となった。これにより、さまざまな応用分野でのデータ解析がより精度高く、信頼性のあるものとなった。以上のように、本研究はラフ集合理論に基づくクラスタリング手法とその応用において重要な成果を上げ、不確実性を考慮したデータ解析手法の新たな可能性を示した。今後も、さらなる研究と応用の展開が期待される。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計25件（うち招待講演 0件 / うち国際学会 13件）

1. 発表者名 Ubukata Seiki, Mouri Kenryu, Honda Katsuhiro
2. 発表標題 Basic Consideration of Collaborative Filtering Based on Rough Co-clustering Induced by Multinomial Mixture Models
3. 学会等名 2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (国際学会)
4. 発表年 2022年

1. 発表者名 Ubukata Seiki, Kawakami Tomohiro, Honda Katsuhiro
2. 発表標題 Adaptive Online Rough C-Means Clustering and Its Application to Collaborative Filtering
3. 学会等名 2022 IEEE Symposium Series on Computational Intelligence (国際学会)
4. 発表年 2022年

1. 発表者名 三栖 翔, 生方 誠希, 本多 克宏
2. 発表標題 Rough C-Meansクラスタリングに基づくアンサンブル協調フィルタリングに関する一検討
3. 学会等名 第39回ファジィシステムシンポジウム
4. 発表年 2023年

1. 発表者名 毛利 憲竜, 生方 誠希, 本多 克宏
2. 発表標題 粒状性を考慮したラフ集合ベースの混合多項分布型共クラスタリングに基づく協調フィルタリング
3. 学会等名 第39回ファジィシステムシンポジウム
4. 発表年 2023年

1. 発表者名 川上 知宏, 生方 誠希, 本多 克宏
2. 発表標題 Adaptive Mini-Batch Rough C-Meansクラスタリングとその協調フィルタリングへの応用に関する一検討
3. 学会等名 第66回自動制御連合講演会
4. 発表年 2023年

1. 発表者名 Ubukata Seiki, Mouri Kenryu, Honda Katsuhiko
2. 発表標題 Basic Consideration of Collaborative Filtering Based on Rough Co-clustering Induced by Multinomial Mixture Models
3. 学会等名 2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (国際学会)
4. 発表年 2022年

1. 発表者名 Ubukata Seiki, Kawakami Tomohiro, Honda Katsuhiko
2. 発表標題 Adaptive Online Rough C-Means Clustering and Its Application to Collaborative Filtering
3. 学会等名 2022 IEEE Symposium Series on Computational Intelligence (国際学会)
4. 発表年 2022年

1. 発表者名 村上 雄大, 生方 誠希, 本多 克宏
2. 発表標題 Rough Set C-Meansクラスタリングに基づく協調フィルタリングにおける種々の粒状化手法の効果
3. 学会等名 第38回ファジィシステムシンポジウム
4. 発表年 2022年

1. 発表者名 畑中 浩貴, 生方 誠希, 本多 克宏
2. 発表標題 Rough C-Meansクラスタリングに基づく協調フィルタリングにおける種々の次元削減手法の活用
3. 学会等名 第38回ファジィシステムシンポジウム
4. 発表年 2022年

1. 発表者名 キム ヘラン, 生方 誠希, 本多 克宏
2. 発表標題 目的関数ベースのRough Membership C-Meansクラスタリングに基づく協調フィルタリング
3. 学会等名 第30回インテリジェント・システム・シンポジウム
4. 発表年 2022年

1. 発表者名 生方 誠希, 得原 一馬, 本多 克宏
2. 発表標題 Probabilistic Rough Set C-Meansクラスタリングに基づく協調フィルタリングに関する基礎的検討
3. 学会等名 第65回自動制御連合講演会
4. 発表年 2022年

1. 発表者名 畑中 浩貴, 生方 誠希, 野津 亮, 本多 克宏
2. 発表標題 Rough C-Means クラスタリングに基づく協調フィルタリングにおける種々の次元削減手法の活用
3. 学会等名 計測自動制御学会 システム情報部門 学術講演会2022
4. 発表年 2022年

1 . 発表者名 S. Ubukata, H. Hatanaka, A. Notsu, K. Honda
2 . 発表標題 Collaborative Filtering Based on Rough C-Means Clustering Using Dimensionality Reduction by Random Projection
3 . 学会等名 The SICE Annual Conference 2021 (国際学会)
4 . 発表年 2021年

1 . 発表者名 H. Kim, S. Ubukata, A. Notsu, K. Honda
2 . 発表標題 Two Types of Collaborative Filtering Based on Rough Membership C-Means Clustering
3 . 学会等名 22nd International Symposium on Advanced Intelligent Systems (国際学会)
4 . 発表年 2021年

1 . 発表者名 S. Ubukata, Y. Murakami, A. Notsu, K. Honda
2 . 発表標題 Basic Consideration of Collaborative Filtering Based on Rough Set C-Means Clustering
3 . 学会等名 22nd International Symposium on Advanced Intelligent Systems (国際学会)
4 . 発表年 2021年

1 . 発表者名 高橋 秀, 生方 誠希, 野津 亮, 本多 克宏
2 . 発表標題 Online Rough C-Meansクラスタリングとその協調フィルタリングへの応用に関する一検討
3 . 学会等名 第37回ファジィシステムシンポジウム
4 . 発表年 2021年

1. 発表者名 杉本 篤史, 生方 誠希, 野津 亮, 本多 克宏
2. 発表標題 Rough Set C-Medoidsクラスタリングに関する一検討
3. 学会等名 インテリジェント・システム・シンポジウム2021
4. 発表年 2021年

1. 発表者名 野岳 就拓, 生方 誠希, 野津 亮, 本多 克宏
2. 発表標題 ラフ集合理論に基づく粒状性を考慮した共クラスタリングに関する一検討
3. 学会等名 インテリジェント・システム・シンポジウム2021
4. 発表年 2021年

1. 発表者名 S. Ubukata, N. Nodake, A. Notsu, K. Honda
2. 発表標題 Basic Consideration of Co-Clustering Based on Rough Set Theory
3. 学会等名 8th International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making (国際学会)
4. 発表年 2020年

1. 発表者名 S. Ubukata, A. Sugimoto, A. Notsu, K. Honda
2. 発表標題 Basic Consideration of Rough C-Medoids Clustering with Minkowski Distance
3. 学会等名 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (国際学会)
4. 発表年 2020年

1. 発表者名 S. Ubukata, S. Takahashi, A. Notsu, K. Honda
2. 発表標題 Basic Consideration of Collaborative Filtering Based on Rough C-Means Clustering
3. 学会等名 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (国際学会)
4. 発表年 2020年

1. 発表者名 キム ヘラン, 生方 誠希, 野津 亮, 本多 克宏
2. 発表標題 Rough Membership C-Means法に基づく協調フィルタリングに関する一考察
3. 学会等名 2020年度計測自動制御学会関西支部・システム制御情報学会シンポジウム
4. 発表年 2021年

1. 発表者名 Ubukata Seiki, Futakuchi Kazushi
2. 発表標題 Collaborative Filtering Based on Rough C-Means Clustering with Missing Value Processing
3. 学会等名 10th International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making (国際学会)
4. 発表年 2023年

1. 発表者名 Ubukata Seiki, Ehara Kazuma
2. 発表標題 Collaborative Filtering Based on Probabilistic Rough Set C-Means Clustering
3. 学会等名 10th International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making (国際学会)
4. 発表年 2023年

1. 発表者名 Ubukata Seiki、Hayashi Ryoya、Honda Katsuhiro
2. 発表標題 Collaborative Filtering Based on Rough Set C-Means Clustering with Noise Rejection Mechanism
3. 学会等名 24th International Symposium on Advanced Intelligent Systems (国際学会)
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------