

## 科学研究費助成事業 研究成果報告書

平成 26 年 5 月 14 日現在

機関番号：24403

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500283

研究課題名(和文) 不完全データのファジィ共クラスター抽出に関する研究

研究課題名(英文) Study on Fuzzy Co-clustering of Incomplete Data

研究代表者

本多 克宏 (Honda, Katsuhiko)

大阪府立大学・工学(系)研究科(研究院)・教授

研究者番号：80332964

交付決定額(研究期間全体)：(直接経費) 3,900,000円、(間接経費) 1,170,000円

研究成果の概要(和文)：購買履歴データや文書-キーワード頻度データのような共起関係データの分析においては、関連の強い個体と項目の組からなる共クラスター抽出が必要である。しかし、協調フィルタリングによる推薦モデルやテキスト情報の自動要約などへの展開をはかる上で、データの不完全性(データの欠測や汚濁、非ユークリッド性など)に対応した効率的なアルゴリズムの開発が不可欠となっている。本研究では、共起関係や不完全性、非ユークリッド性などのデータに内在する特異な特性を許容するデータ分類手法を開発し、協調フィルタリングによる情報選別、テキスト分析などに応用することで、ヒトに優しい知的情報処理技術を開発した。

研究成果の概要(英文)：Cooccurrence information analysis is an important technique for handling such cooccurrence information as purchase history data or document-keyword frequency data with the goal of extracting co-clusters of mutually familiar pairs of objects and items. In applying co-clustering algorithms to collaborative filtering and document summarization, it is needed to handle such data incompleteness as missing elements, noise and non-Euclidean nature. In this research, several co-clustering algorithms, which can handle incomplete cooccurrence information having intrinsic singular features, were studied and applied to personalized recommendation and document analysis for developing human-friendly intelligent information processing techniques.

研究分野：総合領域

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 共クラスタリング 協調フィルタリング 意思決定支援 ファジィクラスタリング

1. 研究開始当初の背景

(1) 大規模なデータベースから意思決定に有益な情報を探し出すデータマイニングでは、重要な要素として「四つのS：測定、視覚化、相関、層別」がある。層別で用いられるクラスター分析は教師なし分類の代表的な手法であり、ヒトが直感的に行う分類行動を知能化技術により再現する。

(2) 購買履歴データや文書 - キーワード頻度データのような共起関係データの分析においては、関連の強い個体と項目の組からなる共クラスター抽出が必要であり、協調フィルタリングによる推薦モデルやテキスト情報の自動要約などへの展開をはかる上で、データの不完全性（データの欠測や汚濁、非ユークリッド性など）に対応した効率的なアルゴリズムの開発が不可欠となっている。

2. 研究の目的

(1) 教師なし分類の代表的な手法であるクラスター分析は、ヒトが直感的に行う分類行動を知能化技術により再現するものであり、データベースの情報要約において有効である。

(2) 本研究では、第1の目的として、共起関係や不完全性、非ユークリッド性などのデータに内在する特異な特性を許容するデータ分類手法を開発する。

(3) また、第2の目的として、開発した手法を協調フィルタリングなどの情報選別、テキスト分析などに応用し、ヒトに優しい知的情報処理技術を開発する。

3. 研究の方法

(1) 情報要約においてクラスター間の相違を明確化するための方策として、相互の変量の依存性の排除を取り入れたファジィ共クラスタリングアルゴリズムの開発を目指した。個体と項目の両方の分割において排他性を導入することで、協調フィルタリング問題における嗜好のあいまい性を明確化した情報要約を実現し、推薦性能の向上を目指した。

(2) クラスター数などのモデルパラメータの設定を自動化するための方策として、共クラスタリングのための妥当性尺度の開発を目指した。種々に設定したモデルパラメータで得られる解の中から最も妥当なものを選択するための尺度を開発することで、分析プロセスを自動化する機構の確立を目指した。

(3) データに内在する非ユークリッド性を処理するための方策として、非ユークリッド関係データの分析手法の開発を目指した。従来のクラスタリングに基づくデータの局所的な線形構造へのデータ要約は、多次元観測値ベクトルを基としていたが、データ間の類

似度行列のような非ユークリッド性を持つ関係性データからの局所的な線形構造の抽出法を開発することで、直感的にとらえやすいデータ視覚化法の確立を目指した。

(4) 購買履歴情報や医療診断データなどの個人情報や安心・安全に処理するための方策として、データのk匿名化クラスタリング手法の開発を目指した。少なくともk人の観測値を互いに区別不可能とすることでプライバシーを定量的に保証するk匿名化の機構を共クラスタリングモデルに導入することで、安心・安全な共起関係データの分析を可能とする技術の確立を目指した。

(5) 複数の組織で分散的に収集・蓄積されている大規模な共起関係データを効率的に活用するための方策として、分散型データベースのファジィ共クラスタリング手法の開発を目指した。組織ごとに収集・蓄積されたデータを他の組織に公開して使用することは個人情報の保護の観点から好ましくないことから、個々のデータを公開することなしに共クラスタリングを施す機構の確立を目指した。

4. 研究成果

(1) 相互の変量の依存性を排除しながらファジィ共クラスタリングを実装するために、項目の排他性を考慮したアルゴリズムを開発した。従来の共クラスタリングモデルでは、個体については唯一のクラスターに帰属するように排他的な制約が課されていたのに対して、項目についてはクラスター内での相対的な重要度を表す指標となっていたために、複数のクラスターで重要とみなされる項目については、クラスターの特徴の比較において解釈が困難となる場面があった。開発したアルゴリズムでは、指定された項目についてなるべく唯一のクラスターに属するように制約することで、クラスター間の相違が明確化され、クラスターごとの特徴づけが容易となった。また、共クラスタリング構造に基づく個別推薦を行う協調フィルタリングへの適用においては、クラスター間での共有度の強い項目について排他性の制約を付加することで、推薦性能が向上する場合があることを明らかにした。

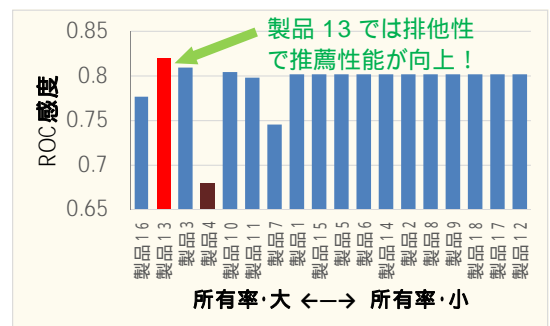


図1: 購買履歴データでのROC感度の比較

(2) 共クラスタリングで最も妥当なパラメータ設定の解を自動的に選択する基準として、多次元データのファジィクラスタリングでしばしば用いられる Xie-Beni 型の妥当性尺度の改良版を開発した。Xie-Beni 尺度ではクラスター内誤差とクラスター間相違の比で分割の妥当性を評価しているが、開発した尺度では共起情報からクラスター内のコンパクトさとクラスター間のセパレートさを評価している。協調フィルタリングへの応用実験では、妥当性尺度の値が最も大きい分割において推薦性能が最大となっており、応用可能性の高い解を適切に選択することができる妥当性尺度であることが示された。

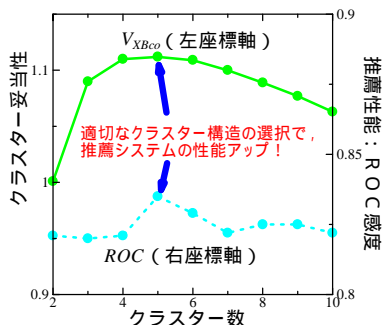


図 2：購買履歴データでの妥当性尺度と推薦性能（ROC 感度）の比較

(3) 非ユークリッドな関係性データに内在する局所的な線形依存関係を抽出する線形クラスタリング法を開発し、ノイズや不完全データを処理する手法を提案したほか、意思決定支援ツールとしてデータ構造を視覚化する手順とその有効性を示した。電子メールなどのテキストデータに代表される非構造データの自動分類などの基盤技術を構築する成果を得た。

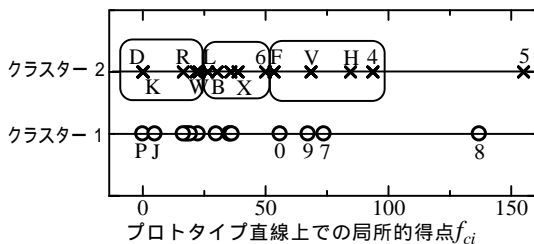


図 3：モースル信号の類似度データでの視覚化実験（従来法よりも、信号間の類似性を適切に表現した視覚化が可能となった）

(4) ファジィクラスター分析に基づくデータのコード化の新たな応用展開として、データの匿名化によるプライバシー保護データマイニングにおける有効なアルゴリズムを開発した。ファジィ分割を用いた匿名化により、従来法よりも少ない情報損失でプライバシーが保護できることを示した。また、共起関係データにおいても k 匿名性を保持したファジィ分割を可能とし、個人のプライバシーを保護したデータ活用ができることを示した。

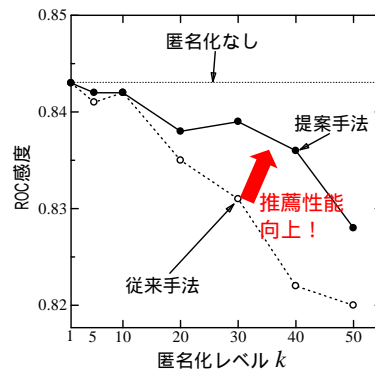


図 4：購買履歴データでの匿名化手法と推薦性能（ROC 感度）の比較

(5) 分散型データベースの複合的な活用を可能にする FCM 型の共クラスタリングモデルとして、ビッグデータの活用を目指した計算効率化のアルゴリズムを提案し、協調フィルタリングへの適用可能性を確認した。また、複数の組織で個別に収集・蓄積された共起関係データをプライバシーの侵害なしに相互活用するために、暗号化機構を取り入れた分散計算アルゴリズムを開発した。

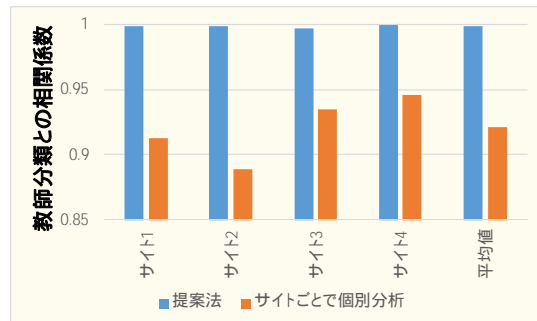


図 5：4 サイトに分散保存されたデータの分析精度の比較（教師分類との相関係数。提案法による相互活用で精度向上）

(6) その他の展開においては、ファジィクラスタリングの理論的基盤の考察として、正規化法とノイズ感度の関連や不完全データの主成分分析との関連についての検証研究を行った。また、太陽光発電によるスマートハウス実現のための発電量分析・予測モデルへ

月	回帰式
5月	$y = 1.020 - 0.040x_1 + 1.404x_2$
6月	$y = 1.248 - 0.052x_1 + 1.391x_2$
7月	$y = 1.238 - 0.059x_1 + 1.358x_2$
8月	$y = 1.272 - 0.048x_1 + 1.125x_2$
9月	$y = 1.397 - 0.083x_1 + 1.467x_2$
10月	$y = 0.017 - 0.008x_1 + 1.221x_2$
11月	$y = 0.692 - 0.040x_1 + 1.514x_2$
12月	$y = 0.489 - 0.019x_1 + 1.384x_2$
1月	$y = 0.981 - 0.044x_1 + 1.334x_2$
2月	$y = 0.683 - 0.029x_1 + 1.660x_2$
3月	$y = 1.152 - 0.072x_1 + 1.720x_2$

図 6：月ごとの発電量予測回帰式の比較（y：発電量，x1：時刻，x2：日照時間，冬季の極端な落ち込みを表現している）

の拡張なども視野に入れ、向きに依存した不  
確実性やノイズにロバストなクラスタリ  
ングモデルを提案し、回帰モデルや局所的低次  
元マップの構築法への展開に関する研究を  
行った。

## 5. 主な発表論文等

### 〔雑誌論文〕(計15件)

本多克宏, ロバストk-Means クラスタリ  
ングとファジィ主成分分析, 日本知能  
情報ファジィ学会誌, 査読無, Vol. 25,  
No. 3, pp.74-80 (2013)

K. Honda, A. Kawano, H. Kasugai, A.  
Notsu, Procedia Computer Science, 査  
読有, Vol. 22, pp. 477-484 (2013)  
DOI: 10.1016/j.procs.2013.09.126

K. Honda, T. Yamamoto, A. Notsu, H.  
Ichihashi: Visualization of  
Non-Euclidean Relational Data by  
Robust Linear Fuzzy Clustering Based  
on FCMdd Framework, Journal of  
Advanced Computational Intelligence  
and Intelligent Informatics, 査読有,  
Vol. 17, No. 2, pp. 312-317 (2013)

本多克宏, クラスタリングの概念と意  
思決定支援への応用, オペレーション  
ズ・リサーチ, 査読無, Vol. 57, No. 5,  
pp. 520-254 (2012)

K. Honda, C.-H. Oh, Y. Matsumoto, A.  
Notsu, H. Ichihashi, Exclusive  
Partition in FCM-type Co-clustering  
and Its Application to Collaborative  
Filtering, International Journal of  
Computer Science and Network Security,  
査読有, Vol. 12, No. 12, pp. 52-58  
(2012)

URL: [http://paper.ijcsns.org/07\\_book/201212/20121209.pdf](http://paper.ijcsns.org/07_book/201212/20121209.pdf)

K. Honda, S. Nakao, A. Notsu, H.  
Ichihashi, Alternative Fuzzy c-Lines  
and Local Principal Component  
Extraction, International Journal of  
Knowledge Engineering and Soft Data  
Paradigms, 査読有, Vol. 3, No. 2, pp.  
188-200 (2011)

DOI: 10.1504/IJKESDP.2011.045728

### 〔学会発表〕(計71件)

本多克宏, S. Tong, 野津 亮, k-Medoids  
法の概念に基づくファジィ共クラスタ  
リング, 第5回横幹連合コンファレンス,  
2013年12月21日, 高松

K. Honda, A. Notsu, C.-H. Oh, 14th  
International Symposium on Advanced  
Intelligent Systems, 2013年11月16  
日, Daejeon, Korea

K. Honda, C.-H. Oh, A. Notsu, H.  
Ichihashi, FCM-type Co-clustering  
with Exclusive Partition of Selected

Items and Application to  
Collaborative Filtering, 6th  
International Conference on Soft  
Computing and Intelligent Systems and  
13th International Symposium on  
Advanced Intelligent Systems, 2012  
年11月23日, Kobe, Japan

K. Honda, A. Kawano, A. Notsu, H.  
Ichihashi, A Fuzzy Variant of k-Member  
Clustering for Collaborative  
Filtering With Data Anonymization,  
2012 IEEE International Conference on  
Fuzzy Systems, 2012年06月11日,  
Brisbane, Australia

K. Honda, A. Notsu, H. Ichihashi,  
Collaborative Filtering by User-Item  
Fuzzy Co-cluster Extraction with  
Partially Exclusive Condition, 12th  
International Symposium on Advanced  
Intelligent Systems, 2011年10月1  
日, Suwon, Korea

K. Honda, A. Notsu, H. Ichihashi,  
Partially Exclusive Condition for  
Sequential Fuzzy Co-cluster  
Extraction, 2011 IEEE International  
Conference on Fuzzy Systems, 2011年  
6月29日, Taipei, Taiwan

## 6. 研究組織

### (1)研究代表者

本多 克宏 (HONDA Katsuhiko)  
大阪府立大学・大学院工学研究科・教授  
研究者番号: 80332964

### (2)研究分担者

市橋 秀友 (ICHIHASHI Hidetomo)  
大阪府立大学・大学院工学研究科・教授  
研究者番号: 30151476

野津 亮 (NOTSU Akira)  
大阪府立大学・大学院工学研究科・准教授  
研究者番号: 40405345