様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２９　年　　５　月　　９　日現在

機関番号：　２４４０３
研究種目：　基盤研究(C)（一般）
研究期間：　2014～2016
課題番号：　２６３３０２８１

研究課題名（和文）大規模共起関係データから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 from Large Scale Co-occurrence Data

研究代表者
本多　克宏（Honda, Katsuhiro）

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）　　3,700,000円

研究成果の概要（和文）：大規模な共起関係データに対する共クラスタリングによる情報縮約を通して，文書データやwebデータの効率的な解析技術を確立し，ヒトに優しい知的情報処理技術を実現することを目的に研究を実施した．
理論的側面では，統計的共クラスタリングモデルに内在する分割ファジィネスの調整による分割性能の向上や，ノイズ除去機構との融合による外れ値へのロバスト性の向上を実現した．応用的側面では，Twitter文書解析や固有顔識別などへの展開において，半教師情報の活用や個人情報匿名化によるプライバシー保護との融合を実現した．

研究成果の概要（英文）：In this study, we tried to realize human-friendly intelligent information processing technologies through development of effective analysis techniques for documents and web data with co-clustering-based summarization of large scale co-occurrence data.
In theoretical aspects, we achieved improvement of partition quality by adjusting intrinsic fuzziness of statistical co-clustering models and robustification of co-clustering models against outliers by introducing a noise rejection mechanism. In practical aspects, we achieved utilization of semi-supervision in Twitter document analysis and privacy preservation in eigen-face authentication with anonymization of personal information.

研究分野：　ソフトコンピューティング

キーワード：　ファジィクラスタリング　共クラスタリング　意思決定支援　文書解析　Webデータ解析

１．研究開始当初の背景

(1) データマイニングでは，従来の統計手法では得られない隠れた知識の発見が目的となる．しかし，実世界の問題に対峙する際には，個別のデータに特化した障害があり，それらを許容する分析手法の開発が不可欠となっていた．

(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，大規模データに内在するノイズの除去やセンシティブな個人情報の取り扱いなどに配慮したアルゴリズムの開発が強く望まれていた．

２．研究の目的

(1) 大規模な共起関係データに対する共クラスタリングによる情報縮約を通して，文書データや web データの効率的な解析技術を確立し，ヒトに優しい知的情報処理技術を実現することを目的に研究を実施した．

(2) 本研究では，第１の目的として，大規模な共起関係データの分析で問題となるデータの分散配置やセンシティブ情報の混入，ノイズによる汚濁といったデータ固有の特異性に対応した分析アルゴリズムの理論の確立を掲げた．

(3) また，第２の目的として，実応用において可用性を高める改善を目指し，モデルパラメータの設定を容易とする理論モデルの確立を掲げた．実世界の Twitter データをはじめとする文書データ解析での有効性を明らかにすることで，ヒトに優しい知的情報処理技術の開発が実現される．

３．研究の方法

(1) 個体と項目の対の親近性に基づく同時クラスタリングにおいては，個体と項目の両方の分割ファジィ度の最適化が必要となり，従来のファジィ共クラスタリング手法ではパラメータ設定の困難性が問題となっていた．そこで，統計的な共クラスタリングモデルである混合多項分布や確率的潜在意味解析が潜在的に含む分割ファジィ度に着目し，確率混合分布のファジィ度との対比によりモデルパラメータを容易に調整可能なファジィ共クラスタリングモデルを新たに開発した．

(2) 付随的な参考情報を活用することで，分割性能を向上するアプローチとして，一部の個体に関するクラス情報が与えられた場合の半教師付き共クラスタリング機構を開発し，IT 関連企業からの協力を得て実施した Twitter データにおける性別判定課題における有効性を検証した．

(3) 大規模な共起関係データが複数のサイトに分散保存されている場合に，それらを一箇所に集約することなしに，効率的でかつプライバシーの侵害なしに共クラスタリングを適用するアルゴリズムを開発した．分散するデータベースの個々に内在する共クラスター構造のみを共有することで，計算効率の良いクラスター構造分析が可能になると同時に，他のサイトに対する情報漏えいを防いだ安心・安全なデータ活用を可能としている．

(4) 個人情報を含む実データへの応用展開を進める工夫として，固有顔特徴量や共起情報の k 匿名化に有効なファジィクラスタリングアプローチを開発した．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 匿名化の機構に，ファジィ分割の概念を導入し，区別不可能な k 個体の群の境界をあいまいとして，群の間の重なりを許容することで，情報損失の少ない匿名化アルゴリズムを目指した．

(5) ノイズを含む実データへの拡張性を向上させるべく，ノイズ除去機能を導入したファジィ共クラスタリングモデルを開発した．FCM 型のノイズクラスタリングにおいてノイズクラスターが一様分布に従うプロトタイプを持つことに着想を得て，混合多項分布に一様な生起確率を持つノイズクラスターを導入した．ノイズとみなされる個体の寄与度をすべてのクラスターに対して低減させることで，ロバストな共クラスター構造の抽出を可能とした．

４．研究成果

(1) 統計的な共クラスタリングモデルである混合多項分布や確率的潜在意味解析の内在的な分割ファジィ度を可変なファジィ化重みで調整することで，i）統計モデルに勝る解釈容易性や分割性能を持つファジィ共クラスタリングモデルの開発と，ii）決定論的アニーリングとの融合による局所解からの脱却を実現した．高い分割性能を維持しながらクラスターの特性の解釈を容易にすることで，共クラスター構造解析の可用性が高まった．

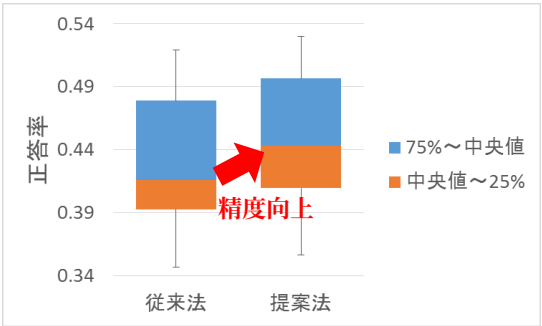

図１：決定論的アニーリングによる精度向上

(2) 実世界のデータ解析では，多数の個体に教師ラベルを人手で付与することは高いコストを要する一方，完全な教師なし学習の分割性能には限界がある．そこで一部の個体にのみ付与された半教師ラベル情報を活用するアプローチとして，i) 初期プロトタイプのロバストな推定と，ii) 共クラスター寄与度の精度向上の 2 ステップでの活用モデルを導入した．Twitter データ解析における性別識別実験においては，i) 統計的な共クラスタリングに比して分割ファジィ度を大きく設定することで分割性能が向上することと，ii) 全個体の教師ラベルが利用可能な場合よりも高い識別性能が半教師学習により実現可能なことが明らかとなった．

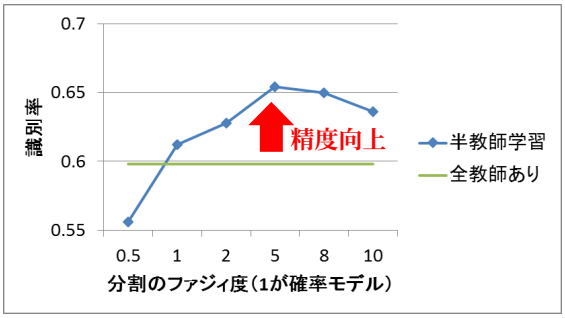

図 2：半教師付き学習でのファジィ度と分割性能との関連の比較

(3) 複数の組織で分散的に保存されている共起関係データからの共クラスター構造分析法として，暗号鍵を付加した共クラスター情報の集約法を開発した．ユーザによる商品の購買履歴データへの応用においては，個々のユーザによる商品の購買の有無だけでなく，販売組織におけるユーザの購買傾向をも秘匿した情報集約を可能とした．実世界の購買履歴データを用いた協調フィルタリング推薦モデルの数値実験において，水平分散型データの場合に 2.9%，垂直分散型データの場合に 12.4%の推薦性能の向上を確認した．

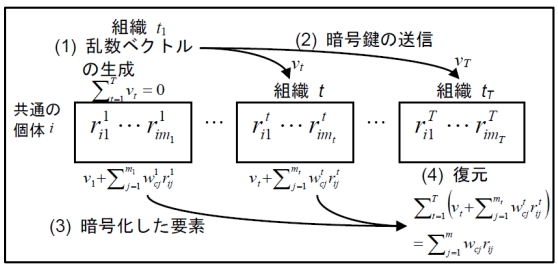

図 3：垂直分散型データの共クラスター情報集約機構のイメージ（組織ごとの購買履歴情報 r と共クラスター情報 w に暗号鍵 v を付加して集約することで，プライバシーを保護しながら共クラスター構造情報を集約）

(4) 駅などの大規模複合施設を想定し，複数の出入口で撮像される顔画像から各個人の入口-出口の対応を推定し出入口の間の移動人数を計数する課題において，固有顔特徴量を出入口ごとに k 匿名化して集約した後に顔識別するプライバシー保護群集行動分析の数値実験を行った．ファジィ帰属度を用いるアプローチにより，匿名化における情報損失を減らして推定性能を向上できることが示された．k 匿名化レベルを k=9 程度まで高めても行動比率ベクトルをコサイン相関 0.97 以上に推定できた．

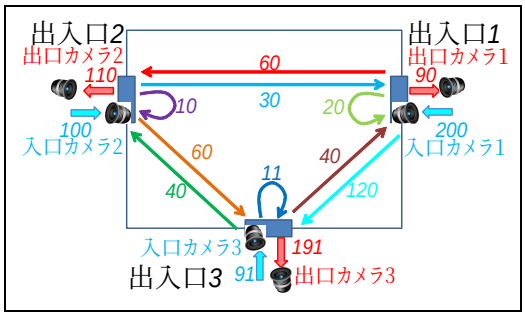

図 4：出入口における撮像モデル設定（移動人数の比率を再現することが目標）

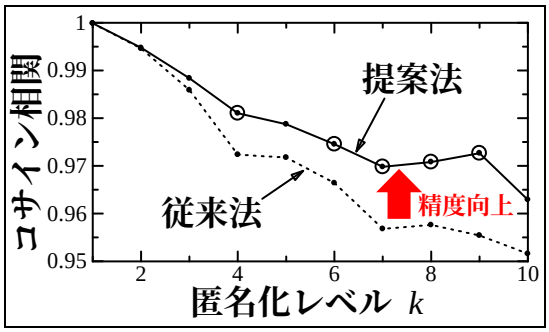

図 5：移動比率ベクトルの推定精度についての匿名化レベルごとの比較（丸囲みのマーカーは統計的に有意な差を示す）

(5) FCM 型ノイズクラスタリングにおけるノイズクラスターの一様分布性を取り入れ，全項目が一様に生起するノイズ共クラスター機構を開発した．ノイズ個体を除去しながら共クラスター構造を推定することで，クラスター純度の向上を確認した．さらに，クラスター数を順次変化させた実装により，最適なクラスター数を推定する手順を開発し，実応用における可用性の向上を実現した．本成果に対して，ソフトコンピューティングに関する国際会議（SCIS＆ISIS2016）において，論文賞を受賞した．

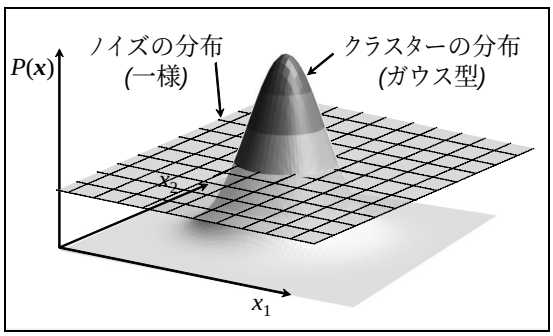

図 6：FCM 型ノイズクラスタリングにおけるクラスター分布のイメージ

５．主な発表論文等

〔雑誌論文〕（計　10　件）

① K. Honda, M. Omori, S. Ubukata, A. Notsu, Fuzzy Clustering-based k-anonymization of Eigen-face Features for Crowd Movement Analysis with Privacy Consideration (プライバシーに配慮した群集行動分析のための固有顔特徴量のファジィクラスタリングに基づいた k 匿名化), International Journal of Innovative Computing, Information and Control, 査読有, Vol. 12, No. 4, pp. 1375-1384 (2016).
http://www.ijicic.org/ijicic-si2015-08.pdf

② 本多克宏, 大森正博, 生方誠希, 野津亮,ファジィ k-member クラスタリングによる顔画像匿名化を伴うプライバシー保護群集行動分析, システム制御情報学会論文誌, 査読有, Vol. 29, No. 3, pp. 130-135 (2016).
DOI: 10.5687/iscie.29.130

③ K. Honda, T. Nakano, C.-H. Oh, S. Ubukata, A. Notsu, Partially Exclusive Item Partition in MMMs-induced Fuzzy Co-clustering and Its Effects in Collaborative Filtering (MMMs の概念に基づくファジィ共クラスタリングにおける部分的な項目の排他的分割と協調フィルタリングにおける効果), Journal of Advanced Computational Intelligence and Intelligent Informatics, 査読有, Vol. 19, No. 6, pp. 810-817 (2015).
DOI: 10.20965/jaciii.2015.p0810

④ K. Honda, S. Oshio, A. Notsu, Fuzzy Co-clustering Induced by Multinomial Mixture Models (混合多項分布からの派生によるファジィ共クラスタリング), Journal of Advanced Computational Intelligence and Intelligent Informatics, 査読有, Vol. 19, No. 6, pp. 717-726 (2015).
DOI: 10.20965/jaciii.2015.p0717

⑤ K. Honda, T. Oda, D. Tanaka, A. Notsu, A Collaborative Framework for Privacy Preserving Fuzzy Co-clustering of Vertically Distributed Cooccurrence Matrices (垂直分散型の共起関係行列のための協調的なプライバシー保護ファジィ共クラスタリング), Advances in Fuzzy Systems, 査読有, Vol. 2015, No. 729072, pp. 1-8 (2015).
DOI: 10.1155/2015/729072

〔学会発表〕（計　43　件）

① K. Honda, Fuzzy Co-clustering and Application to Collaborative Filtering, (ファジィ共クラスタリングと協調フィルタリングへの応用), Integrated Uncertainty in Knowledge Modelling and Decision Making 2016, 2016 年 12 月 1 日, ダナン（ベトナム）

② K. Honda, N. Yamamoto, S. Ubukata, A. Notsu, A Noise Fuzzy Co-Clustering Scheme in MMMs-Induced Clustering (混合多項分布に基づくクラスタリングにおけるノイズファジィ共クラスタリングの枠組み), Joint 8th International Conference on Soft Computing and Intelligent Systems and 17th International Symposium on Advanced Intelligent Systems, 2016 年 8 月 27 日, 北海学園大学（北海道・札幌）

③ K. Honda, M. Omori, S. Ubukata, A. Notsu, A Study on Fuzzy Clustering-based k-anonymization for Privacy Preserving Crowd Movement Analysis with Face Recognition (顔画像認識を伴うプライバシー保護群集行動分析のためのファジィクラスタリングに基づく k 匿名化に関する研究), 7th International Conference of Soft Computing and Pattern Recognition, 2015 年 11 月 14 日, 九州大学（福岡）

④ K. Honda, S. Ubukata, A. Notsu, N. Takahashi, Y. Ishikawa, A Semi-supervised Fuzzy Co-clustering Framework and Application to Twitter Data Analysis (半教師ありファジィ共クラスタリングの一手法とツイッターデータ解析への応用), 4th International Conference on Informatics, Electronics & Vision, 2015 年 6 月 16 日, 北九州国際会議場（福岡・北九州）

⑤ K. Honda, S. Oshio, A. Notsu, Item Membership Fuzzification in Fuzzy Co-clustering Based on Multinomial Mixture Concept (混合多項モデルの概念に基づくファジィ共クラスタリングでのアイテムメンバシップのファジィ化), 2014 IEEE International Conference on Granular Computing, 2014 年 10 月 23 日, 登別グランドホテル（北海道・登別）

〔図書〕（計　1　件）

① K. Honda, Fuzzy Clustering/ Co-clustering and Probabilistic Mixture Models-induced Algorithms (ファジィクラスタリング・共クラスタリングと確率混合モデルに基づくアルゴリズム), Springer, V. Torra, A. Dahlbom, Y. Narukawa(編), Fuzzy Sets, Rough Sets,

Multisets and Clustering, Studies in Computational Intelligence 671, pp. 29-43 (2017).

６．研究組織

(1)研究代表者
本多　克宏（HONDA Katsuhiro）
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
野津　亮（NOTSU Akira）
大阪府立大学・人間社会システム科学研究科・准教授
研究者番号：　４０４０５３４５

生方　誠希（UBUKATA Seiki）
大阪府立大学・大学院工学研究科・助教
研究者番号：　１０７５５６９８