

個人情報保護に配慮した組織協調型の安心・安全な コンテンツ推薦モデル

Safe, Secure and Reliable Recommendation Model Considering Cross-organization Privacy Preservation

2171023



研究代表者 大阪府立大学大学院 工学研究科 教授 本多克宏

[研究の目的]

クラウドをはじめ無尽蔵な情報共有が進む高度情報化社会においては、情報をトリガーとした人間と機械の調和が不可欠であり、ビッグデータからの情報推薦に基づくユーザ支援が最も重要な基盤技術である。一方で、組織横断的に分散収集・蓄積されたデータを相互活用する上では、プライバシー保護が阻害要因となり、情報が持つ価値を十分に享受できていない。

本研究では、関係性データの暗号化・匿名化と情報縮約による推薦モデル構築の手法の確立を研究目標とする。Web データ解析をはじめとするビッグデータ解析においては、文書×単語の共起関係データや顧客×購入商品の購買履歴データなどのような、2項間の関連性をもとにした関係性データ解析が重要性を増している。そこで、暗号化・匿名化によるプライバシー保護を考慮しながら、ネットワークに組織横断的に分散蓄積されている関係性データを共同活用することで、中規模の企業群などの連合組織で Amazon.com のような IT 巨人に匹敵する推薦モデルを実現することを目標に設定した。

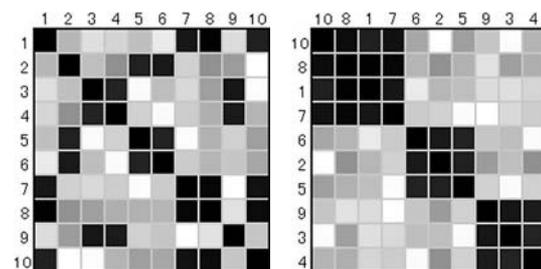
[研究の内容, 成果]

1. 共クラスター構造の直感的な把握手法の開発

1.1 視覚的な構造把握法の概要

まず、文書×単語の共起関係データや顧客×購入商品の購買履歴データに内在する関連性の共クラスター構造の存在性を確認する手法として、新たに視覚的・直感的な共クラスター把握方法を開発した [1]。共起関係行列の行と列を適切に並べ替えることにより、関連の強い個体と項目の対が対角ブロックに強調され、要素の共起度による色彩的な表示法により、関連性の存在性が容易に把握できる手法となっている。

例えば、10 個体間の類似度がグレースケール値を要素とする 10×10 行列として図 1(a) のように与えられた場合、図 1(b) のように並べ替えることで三つのクラスターからなる構造が対角に見いだされる。本研究では、同じ考え



(a) 並べ替え前

(b) 並べ替え後

図 1 視覚的なクラスター構造の把握の例

方を個体×項目の共起関係データに拡張した。

1.2 計算アルゴリズム

n 人のユーザが m 種類の商品を買ったか否かを示すような、 n 個の個体と m 種類の項目の間の $n \times m$ 共起関係データ R が与えられたとする。いま、行列 R を正方行列に拡張した類似度行列

$$S = \begin{pmatrix} O & R \\ R^T & O \end{pmatrix} \quad (1)$$

において、個体と項目の区別なく図1のように類似性の高い個体・項目が近くに配置される要素 i の順番の並べ替え $i \rightarrow \pi_i$ を、以下の目的関数の最大化により実現する。

$$J = \frac{1}{2} \sum_{i=1}^{n+m} \sum_{j=1}^{n+m} (\pi_i - \pi_j)^2 s_{ij} \quad (2)$$

本研究では、この最大化問題の緩和が固有値問題から求められることを明らかにし、さらに、計算効率を改善した計算アルゴリズムを提案した [1]。

1.3 クラスタ数 の推定

また、並べ替えた図に内在するクラスタ数を推定する指標として、Cluster Crossing を提案した [2,3]。並べ替え後の図において、各要素の近傍でどの程度、類似度の大きな要素が凝集しているかを測るため、帯域幅を変化させながら Cluster Crossing を算出することで、3次元での図示を通して、視覚的にクラスタ構造を山と谷で表現することを可能とした。

1.4 数値実験の結果

有効性の数値検証においては、都道府県の統計データへの適用として、人口移動データに提案法を適用した。47都道府県の転出を個体、転入を項目とみなし、転出入の両方の特性を同時並べ替えした結果が、図2となった。また、四つの地域特性を白地図に記載したものを図3に示す。同じ地域内の都道府県へ転居する傾向が強い一方、九州北部のように九州地域内よりも中四国地域とのつながりが示されている県もあるなど、地域特性が把握できることが示された。

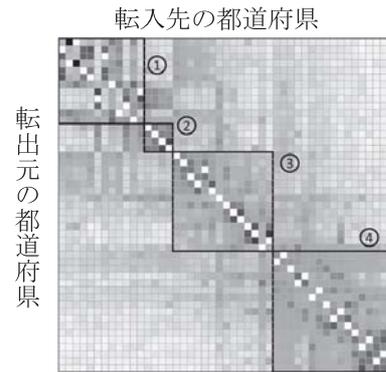


図2 47都道府県の転出入による並べ替え

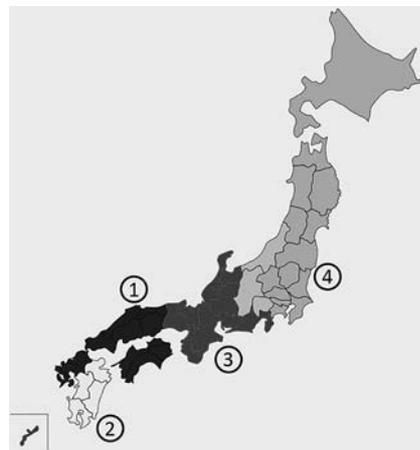


図3 図2における地域特性

2. 組織間協調のための3モード共クラスタリング

2.1 組織間協調モデルの概要

つぎに、組織間協調モデルへの発展として、個体×項目の間の共起関係に加えて、項目×要素の間の共起関係も含んだ3モード共起関係データからの共クラスタ抽出のための新たな分析手法を開発した [4]。研究代表者は先行研究において、化粧品による肌トラブル予測モデルのシステム化を進め、ユーザと化粧品との相性診断に際して、化粧品と材料との共起関係も加味した警告・推薦システムの高精度化に携わった。

ユーザと化粧品との相性には、ユーザ×化粧品の間の共起情報のみならず、潜在的にはユーザ×材料の間の隠れた共クラスタ構造があると想定される。ここで、ユーザ×化粧品の相性

情報と化粧品×材料の成分情報は異なるリソースであり、異なる組織で独立に収集・保持されている場合には、必ずしも協調利用が可能とは限らない。そこで、両データを保持する組織間で協調的にデータを活用するモデルとして、個々のユーザと化粧品の相性情報や化粧品と材料の成分情報は秘匿するものの、多数のユーザや化粧品を群化して表現した共クラスター構造という匿名化情報のみを共有する分析手法を実現した。図4にデータ共有モデルのイメージを示す。

2.2 計算アルゴリズム

n 個体と m 項目に関する共起関係データ R を所有する組織 A と、 m 項目と p 要素に関する共起関係データ S を所有する組織 B で、協調して個体×項目×要素のクラスター構造を分析することを目的に、データ要素を互いに公開

することなしに、個体 i 、項目 j および要素 k のクラスター c への分類指標となるメンバシップ u_{ci} 、 w_{cj} および z_{ck} を求める。

組織ごとに保持する情報と共有される情報の流れを図5のようにおくと、3種類のメンバシップの更新則は、それぞれ以下のように求められる [4]。

$$u_{ci} = \frac{\exp\left(\frac{1}{\lambda_u} \sum_{j=1}^m w_{cj} r_{ij} \beta_{cj}\right)}{\sum_{i=1}^n \exp\left(\frac{1}{\lambda_u} \sum_{j=1}^m w_{ij} r_{ij} \beta_{ij}\right)} \quad (3)$$

$$w_{cj} = \frac{\exp\left(\frac{1}{\lambda_w} \alpha_{ci} \beta_{cj}\right)}{\sum_{i=1}^n \exp\left(\frac{1}{\lambda_w} \alpha_{ci} \beta_{ci}\right)} \quad (4)$$

$$z_{ck} = \frac{\exp\left(\frac{1}{\lambda_z} \sum_{j=1}^m w_{cj} s_{jk} \alpha_{cj}\right)}{\sum_{i=1}^p \exp\left(\frac{1}{\lambda_z} \sum_{j=1}^m w_{ij} s_{ji} \alpha_{ci}\right)} \quad (5)$$



図4 組織間協調のイメージ

ここで、二つの組織で共有される行列 A と B には、元の共起関係行列の要素が隠されただけでなく、組織 A が対象とする個体の数 n や、組織 B が考慮する要素の数 p も秘匿されており、安心・安全な分析方法が実現されている。

2.3 数値実験の結果

組織 A が 40 個体×50 項目の、組織 B が 50

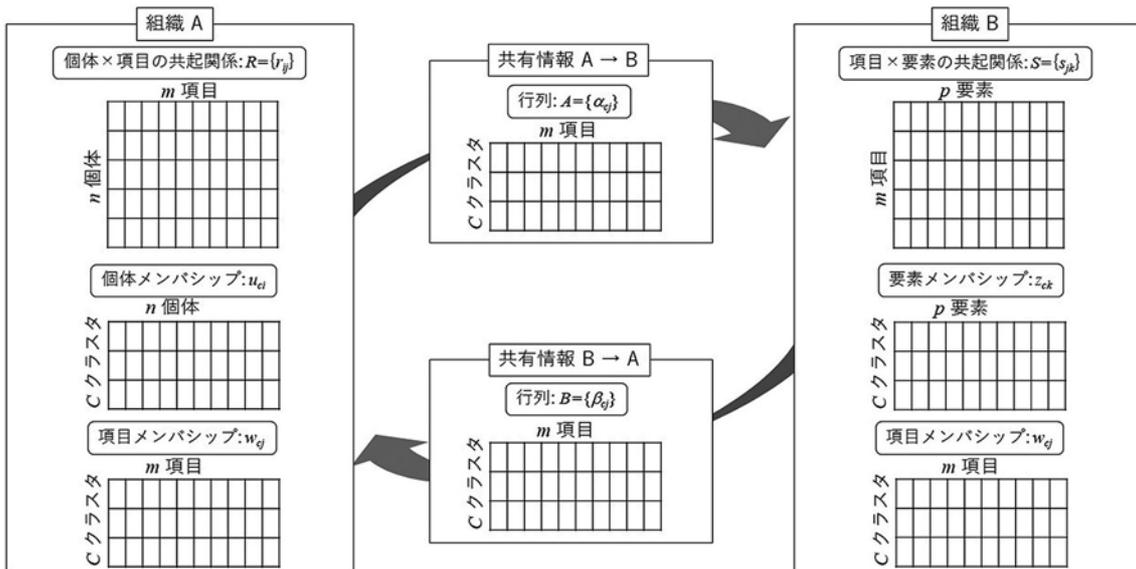


図5 組織ごとに保持する情報と共有される情報の流れ

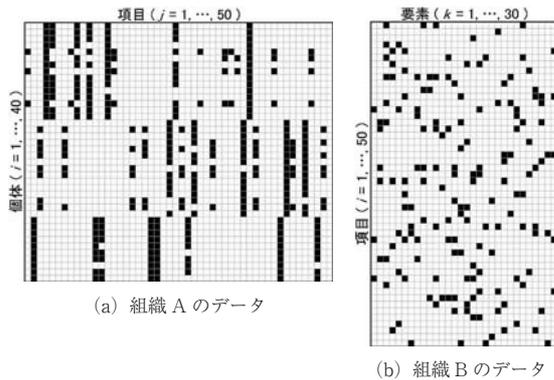


図6 実験用データ

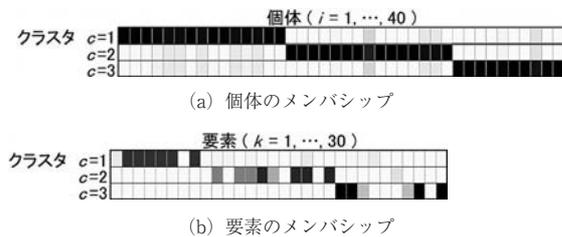


図7 提案法で得られたクラスター構造

項目×30 項目の共起情報をそれぞれ図6のように保持する状況を想定し、行列要素の秘匿を考慮した提案モデルにより、潜在的なクラスター構造を抽出する数値実験を行った。

図6を一見しただけでは、個体と要素の間の関連性をとらえることは困難であるが、提案法により、図7に示したメンバシップがそれぞれ得られた。40個の個体は、(個体1~15)(個体16~30)(個体31~40)の三つのクラスターに分類され、それぞれの個体群を特徴づける潜在因子として、(要素1~10)(要素11~20)(要素21~30)が示された。

これは、二つの組織で共有される行列AとBが、協調的な分析の中で等質化された結果であり、組織間協調による高度情報処理[5]が実現された証左である。

[今後の研究の方向, 課題]

本研究では、個体や項目の間の共起関係情報のクラスター構造の抽出を通して、個人の嗜好

を効率的に要約するデータ分析手法を開発することで、情報推薦システムの基盤技術を確立した。また、複数の組織が個人情報の保護を考慮しながら協調的にデータ活用するための情報共有モデルを開発し、組織間協調により高度な情報処理が可能となることを示した。

今後の方向としては、実応用の際に、組織間で共有される情報の質を評価して取捨選択するモデルの開発が望まれる。ビッグデータ解析では、多様な観点からの観測データが混在する状況があり、それらを融合利用する上では、マルチビューの質の評価が課題となっている。提案手法の利点である秘匿性を保持しながら質を評価するアプローチの開発が必要となる。

また、クラウド上での分散実装などを通して、安価で活用しやすいシステムとして提供できるプラットフォームを実現することが、実用化に向けた次のステップとなる。

[成果の発表, 論文等]

- [1] 佐古拓也, 本多克宏, 生方誠希, 野津 亮: Spectral Orderingに基づく共クラスター構造の視覚化とその特徴, システム制御情報学会論文誌, 31, 5, 177-183, 2018.
- [2] K. Honda, T. Sako, S. Ubukata, A. Notsu: Visual Co-cluster Assessment with Intuitive Cluster Validation through Cooccurrence-Sensitive Ordering, Journal of Advanced Computational Intelligence and Intelligent Informatics, 22, 5, 585-592, 2018.
- [3] T. Sako, K. Honda, S. Ubukata, A. Notsu: Cluster Crossing Curve for Intuitive Validation in Visual Assessment of Co-cluster Tendencies, Proc. of 2018 International Conference on Fuzzy Theory and Its Applications, #T2a-1, 75-79, 2018.
- [4] K. Honda, S. Matsuzaki, S. Ubukata, A. Notsu: Privacy Preserving Collaborative Fuzzy Co-clustering of Three-Mode Cooccurrence Data, Proc. of 15th International Conference on Modeling Decisions for Artificial Intelligence, Lecture Notes in Artificial Intelligence 11144, 232-242, Springer, 2018.
- [5] T. Chen, K. Honda: Fuzzy Collaborative Forecasting and Clustering, Springer Nature, 2019.