

宇野 毅明

情報・システム研究機構 国立情報学研究所  
教授

## データ粒子化による高速高精度な次世代マイニング技術の創出

### § 1. 研究成果の概要

データ研磨を中心としたモデル、アルゴリズム技術において、クラスタリングの安定化手法、階層化クラスタの抽出方法について研究した。今までの計画ではデータ研磨アルゴリズムのパラメータを調整することにより粒度の異なるクラスタを抽出するアプローチを目指してきたが、粒度が大きいクラスタにおいて芳しい結果が得られなかったため、方針転換を行い、従来の不安定なクラスタリングアルゴリズムが抽出する様々な粒度のクラスタを、そのアルゴリズムを多数回実行し、得られた大量のクラスタをクラスタリングするメタアルゴリズムを開発した。これにより従来得られなかった中小規模の階層クラスタ構造を得ることに成功し、バスケット分析などで今までは得られないような知見が得られることを確認した。

また、データ研磨とは異なる方法で粒子を得るアルゴリズムの開発にも取り組み、極大(k, l)-閉集合という新しいモデルを構築し、その効率的な発見アルゴリズムを構成した。これは、従来の閉集合パターンに対して摂動安定性の概念を導入して、安定解だけに絞込む物であり、従来の大量パターンから飛躍的少数の代表元に絞込むことを可能としている。

データの可視化と利用においては、データ研磨アルゴリズムのユーザインタフェース Nyson や可視化ツール KIZUNA を整備した。また、中堅スーパー、光洋と共同研究を行い、実際の売り上げデータから、データ研磨技術を用いたデータマイニングで顧客動向の洞察を獲得し、それに基づいて具体的な店舗レイアウト設計の施策を考案、実行し、それによる売り上げの変化を観察した。例として、スナック菓子やビールなど、一般的に不健康と思われそうな食品と、特保食品が良く一緒に買われることを発見し、それらを買う客が最後に特保食品に目が行くよう、レジのそばに特保陳列棚を置いた。売り上げは大きく伸びた。データ解析結果を現場と研究者が共同で議論をし、店舗での施策を行った例として極めて希有で価値の高い研究である。

さらに、データの意味解釈からそれにつづく施策の設計までを高質化、効率化するための複眼的ブラウジングを実現するデータインタラクション環境について、複数端末から得られた時系列デ

ータを統合し、ブラウジングを行って、意味的なアラインメントを行うツールの開発を行った。意味解  
積だけではない、因果の獲得や施策による将来の変化に対する予測を行う基礎が構築されつつ  
ある。

**【代表的な原著論文】**

Takako Hashimoto, Takeaki Uno, Tetsuji Kuboyama, Kilho Shin, Dave Shepard, “Time Series  
Topic Transition Based on Micro-Clustering”, BigComp 2019: 1-8

Alessio Conte, Gaspare Ferraro, Roberto Grossi, Andrea Marino, Kunihiko Sadakane, Takeaki  
Uno, “Node Similarity with  $q$ -Grams for Real-World Labeled Networks”, KDD '18 Proceedings  
of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining  
Pages 1282-1291.

Yasuhiro Yamamoto, Kumiyo Nakakoji, Visual Interactivity to Make Sense of Heterogeneous  
Streams of Design Activity Data, Proceedings of Design Computing and Cognition DCC'18, J.S.  
Gero (Ed.), Springer, pp.745-763, Lecco, Italy, July 2018.

## § 2. 研究実施体制

### (1)「計算技術とモデル化」グループ

- ① 研究代表者:宇野 毅明 (情報・システム研究機構 国立情報学研究所 教授)
- ② 研究項目
  - ・様々なマイニングタスクに対するデータ粒子化モデルと研磨アルゴリズムの開発
  - ・粒子化したデータの利用における補助モデルとアルゴリズムの開発
  - ・匿名化など他の情報処理技術への応用技術開発

### (2)「意味構造解析」グループ

- ① 主たる共同研究者:山本 章博 (京都大学情報学研究科 教授)
- ② 研究項目
  - ・データの粒子化の意味論
  - ・属性選択の高速化と意味論の構成
  - ・構造データに対する意味論

### (3)「実応用」グループ

- ① 主たる共同研究者:羽室 行信 (関西学院大学経営戦略研究科 准教授)
- ② 研究項目
  - ・データ整備
  - ・実データでの効果・効率の検証
  - ・実データへの適用に関わる手法開発
  - ・現実のビッグデータと手法の性質・特性の解明
  - ・ハーデングメカニズムに関する理論構築

### (4)「データインタラクション」グループ

- ① 主たる共同研究者:中小路 久美代 (京都大学学際融合教育研究推進センター 特定教授)
- ② 研究項目
  - ・洞察を誘導し着目点や思考の変化に柔軟に対応する効果的なビジュアルインタラクティブの解明
  - ・ユーザの着目点の抽出と連携および複合化のためのフォーカシング表現技術の確立
  - ・粒子化されたデータ空間の複眼的ブラウジングを実現するデータインタラクション環境の構築