

「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」
平成 26 年度採択研究代表者

H29 年度
実績報告書

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 (国立情報学研究所 教授)

② 研究項目

- ・クラスタリングアルゴリズムの改良・精度向上
- ・バイクラスタリングアルゴリズムの開発、精度向上、検証
- ・アルゴリズムの並列化

(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

- ・2部グラフ研磨とバイクラスタリングアルゴリズムの併用の効果検証
- ・バイクラスタリングアルゴリズムの分類と体系化
- ・人工データの作成
- ・教師無し学習における特徴(座標)選択法の開発

(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)「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デザイン学ユニット 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ティの解明
- ・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

§ 2. 研究実施の概要

計算技術とモデル化グループでは、基本となるデータ研磨アルゴリズムに関する論文が採択された。基本アルゴリズムの性質を理論的に解析し、いくつかの場合における挙動を明らかにし、その挙動に対するある程度の精度の保証を行い、また他の多くのアルゴリズムと性能比較を行うことで、データ研磨手法の優位性を立証した。それにより、トップレベルの国際会議に採択されることとなった。データマイニングの基礎アルゴリズムについても研究を行った。先進的なアルゴリズム技術を適用し、コミュニティのモデルとなるようなクリークの列挙アルゴリズム、現実データの疎性を表す縮退数が低数以下であるような部分構造を列挙するアルゴリズムを開発した。また、データマイニングでは従来考えられてこなかった、有向グラフの概念を導入したマイニングモデルの研究も行い、有向の意味での連結性を担保したクリークを定式化し、その探索アルゴリズムを開発した。また、議論の中で、新たにクラスタリングの安定性が実応用で大変重要な概念であることがわかった。それを実現するため、メタクラスタリングを用いたクラスタリング安定化のアルゴリズムを開発した。これにより、既存アルゴリズムでは、実行のたびに異なる解が得られ、大きな不安定性を持っているところを、既存のアルゴリズムの特徴をいかしつつ、安定した解を出力するアルゴリズムを開発した。

意味構造グループではデータ研磨手法で得られるデータのまとまりは、密な構造を持ち、自然に包含関係による束構造をなし、形式概念解析(FCA)という形で数理的に理論化される。これまで形式概念解析の自然言語処理への応用として、シソーラスに登録されていない単語を新たに登録する方法について、論文としてまとめた。一つの新単語をシソーラスに登録することは、その単語に複数の意味に関係づけることになるため一種のマルチラベル・クラスタリングとなる。さらに、その単語をどの意味に関係づけるかはシソーラスに依存する。そこで、近年劇的に改善された自然言語構文解析技術を利用して、自然言語文中の構文的情報だけを用いて単語の集合を形式概念解析し、多数得られる密構造を意味の候補とし、シソーラスに依存した評価関数を利用して、新単語に関連づけるのに単語な意味を高速に探索する。

実応用グループでは、ファイナンス、及びマーケティング分野においてハーディング現象の解明についての研究を実施した。株式市場におけるハーディング現象のモデル化について、個別株の値動きの関係性変化が市場全体にどのような波及するかについて、ディープラーニングの手法を用いることでモデル構築を行った。また、また、企業間の取引関係データから、価格の急変動イベントを伴った銘柄の情報がどのように伝わるかについてのモデル化を行った。また、大規模データを効率的に処理するためのデータ解析ツールである NYSOL の新たなツールを追加公開した。ヘアサロンの購買行動データの類似度グラフをグラフ研磨し、コミュニティ抽出を行って解析を行う手法を開発した。この成果は、データ解析コンペティション関西予選で最優秀賞、全国大会にて研究奨励賞を受賞した。

インタラクショングループでは、これまでに解明、構築してきた粒子化データに対するビジュアルインタラクティブ性をベースとして、同一データに対して多様なビジュアルインタラクティブ性を適用すること、同一のインタラクティブ性で多様な粒子化データを表現すること、という双方を並行して研究を進めた。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性を解明する研究項目としては、インタラクティブな可視化表現にアニメーション表現を

重畳することによる、動的な表現と静的な表現の複合化を行った。また、線形や円形、矩形といった多様な2次元・3次元表現、さらには地図空間やテキスト空間といった外部的に意味を有する空間上にデータ空間を重畳し、概観と複数フォーカシングのシームレスなスイッチと、視点移動を可能とするようなインタラクションのデザインオプションを探った。また、テキストデータ、数値データ、画像、動画、音声といった個別のマルチメディアデータを時系列で連携させ、表示順序や表示範囲、粒度等を自在に変更できるような環境を構築した。さらに、粒子化されたデータとのユーザインタラクションを記録、分析することで、データ体験から重層的にデータを生成する環境の構築に着手した。

本年度の代表的な論文

Takeaki Uno, Hiroki Maegawa, Takanobu Nakahara, Yukinobu Hamuro, Ryo Yoshinaka, Makoto Tatsuta, "Micro-clustering by data polishing", 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, pp. 1012-1018, Dec., 2017.

Natsuki Kiyota, Sho Shimamura, Kouichi Hirata, "Extacting Mutually Dependent Multisets", Proc. 20th International Conference on Discovery Science DS 2017), Lecture Notes in Artificial Intelligence 10558, pp.267-280, Oct., 2017

Yusuke Kita, Nozomu Kutomi, Tomohiro Sakaguchi, Kumiyo Nakakoji, Design Process Depiction for Richer Process Experience, Proceedings of IASDR 2017 Conference, University of Cincinnati, USA, pp.1-8, November, 2017.