

## 2014年度特別研究期間 研究成果概要

所属・職・氏名：経営戦略研究科・准教授・羽室行信

研究課題：ビッグデータ解析システムの並列化に関する研究・開発

研究期間：2014年4月1日～2015年3月31日

研究成果概要（日本文（全角）の場合は2,000字程度）

本研究の目的は、ビッグデータ解析基盤であるKGMOD(トータルシステム名:NYSOL)の並列化を行うと共に、そのシステムを利用した応用研究を実施することでその有効性を示すことに有る。研究内容は、1)基盤技術の開発、および2)基盤技術を利用した応用研究に大別することができる。以下で、それぞれの内容について報告する。

### 1) 基盤技術の開発

上述のビッグデータ解析システムであるNYSOLの並列処理システムの開発を行った。本研究においては、複数ノードを想定した分散処理の実現に先立ち、一台のPC内での並列処理の汎用的システムの構築を行った。この実装により、CPUの数にスケラブルな処理が可能となり、大規模なビジネスデータを、その基本処理単位での(例えば、株価データにおける銘柄、小売店行動記録データにおける顧客や店舗)高速演算可能となった。このシステムは既にwebページ上でオープンソース・ソフトウェアとして公開済みである(<http://nysol.jp/>)。また内容については、国際会議で報告した[5]。またNYSOLについては論文として発表した[2]。

### 2) 応用研究

スーパーマーケット(全国にチェーン展開する食品販売店)から提供された顧客ID付きPOSデータの解析を行った。商品の同時購買情報を使って商品間の類似度を定義し、顧客別に商品類似度グラフを構成した。そしてその商品類似度グラフからグラフ特徴量を説明変数として抽出し、来店回数の多寡を目的変数とする罰則付きロジスティック回帰モデルを構築した。ここでは顧客単位でのモデル化が必要となるため、上述の並列計算システムを有効に活用することができた。本研究内容は、データ解析コンペティション関西予選で報告し、最優秀賞を獲得した[6]。

また、アパレル関連のECサイトの購買/アクセスログデータを用いた、レコメンデーションシステムの研究も行った。データ提供元であるZozo Townが主催するコンテストの課題に沿う形で研究を進めた。与えられたデータを使って、95の顧客グループ別にPV(page view)率とCV(conversion)率を予測する事が課題である。ここでも上述の並列処理システムを利用することにより、顧客グループ別のレコメンデーションモデルを高速に構築することができ、仮説探索型の分析プロセスに有効に機能した[1]。

また、ファイナンスの分野においては、これまでに大規模なテキストデータ(Bloomberg社が提供する電子ニュースなど)を使ったセンチメント解析を実施してきており、株価予測への有効性を示してきた。しかし、テキストの形態素解析および構文解析に数週間の間を要するという問題があった。そこで、上述の分散処理システムとAmazonが提供するHadoopを組み合

わせることで、20 台のマシンを利用し、数時間で完了できることを確認した。ノード間分散処理については、あくまでも Hadoop の機能を利用したため、その部分を今後独自に実装し、より高速で利用しやすいシステム開発を目指したい。

さらに、経済/ファイナンスの分野における世界の最先端の研究動向に関する調査のため、EEAE-ESEM(European Economic Association & Econometric Society)に出席した。大規模なデータに基づいた研究(特に、行動経済学)を中心に聴講し、経済/ファイナンス理論の動向のみならず、研究者のデータの利用動向について調査を行った。この分野における、今後のビッグデータ研究および処理技術の開発についての有用な知見を得ることができた。そこで得られた発想を元にしたファイナンスデータの解析手法についての報告を[4]にて行った。

#### 研究業績

- [1] 羽室行信、宇野毅明、北島聡, グラフ研磨手法を応用したレコメンデーションシステムの構築,ZOZO Town Data Science Festival Recommend Contest, ホテル椿山荘東京, 2014 年 11 月 8 日.
- [2] 羽室行信、植野剛、鷺尾隆,”極大クリーク列挙技術のビジネス応用とソフトウェアツール”, 電子情報通信学会, 97(12), pp.1103-1109, 2014/12.
- [3] 中原孝信、羽室行信、宇野毅明, “グラフ研磨手法を用いた顧客の店舗選択モデルの構築”, オペレーションズ・リサーチ, 60(2), pp.89-95, 2015/02.
- [4] 羽室行信、グラフ研磨手法のテキスト解析への応用、計算理論とビッグデータワークショップ、東北大学、2015 年 2 月 14 日.
- [5] S.Cheung, M. Nakamoto, Y.Hamuro, “NYSOL: A User-Centric Framework for Knowledge Discovery in Big Data”, ICIT2015, Singapore, 2015/2/3.
- [6] 中原孝信、中元政一、Stephane Cheung, 羽室行信, “グラフ特徴量にもとづく優良顧客分析”, データ解析コンペ関西予選最終報告会、大阪難波 I-site なんば、2015/3/3.

#### 受賞

平成 26 年度データ解析コンペティション関西予選最優秀賞、中原孝信、中元政一、Stephane Cheung、宇野毅明、羽室行信、2015 年 3 月 3 日.