

山西健司

東京大学 大学院情報理工学系研究科
教授

複雑データからのディープナレッジの発見と価値化

§ 1. 研究実施体制

(1)「山西」グループ

- ① 研究代表者:山西 健司(東京大学 大学院情報理工学系研究科、教授)
- ② 研究項目
 - ・ディープナレッジのモデル論、推定論の構築

(2)「増田」グループ

- ① 主たる共同研究者:増田 直紀 (ブリストル大学 Department of Engineering Mathematics, Senior Lecturer)
- ② 研究項目
 - ・ディープナレッジとしてのテンポラル・ネットワークの解析理論の構築推進

(3)「IBM」グループ

- ① 主たる共同研究者:恐神 貴行 (日本アイ・ビー・エム株式会社東京基礎研究所、リサーチスタッフメンバー)
- ② 研究項目
 - ・ディープナレッジを価値につなげるための意思決定最適化技術

(4)「大澤」グループ

- ①主たる共同研究者:大澤 幸生(東京大学 大学院工学系研究科、教授)
- ②研究項目
 - ・ディープナレッジの利用価値を創造するデータ市場の構築手法

§ 2. 研究実施の概要

従来の BigData 研究はデータの大量性に関心が集中してきた。しかし、本研究では、BigData の複雑さ、多様性、変動性に注目し、巨大なデータの背後に眠る潜在知識(これを「ディープナレッジ」とよぶ)を発見し、価値を与えるための方法論を開発することを目的に研究している。

本研究チームは、4つのグループ(山西 G、増田 G、IBMG、大澤 G)に分かれて研究している。

山西 G では、基礎理論として、1)潜在的ダイナミクス と 2)関係データ統合予測 の研究を行った。1)潜在的ダイナミクスとは、観測時系列データの背後にあるディープナレッジの変化を検知することにより、重大イベントの発見を行う研究である。今年度は、多変量時系列の背後にあるディープナレッジとして「木構造」と「独立成分」を考え、記述長最小原理に基づく動的モデル選択により、それらの構造変化を検知する新しいアルゴリズムを開発した。これにより観測データから直接変化検知するよりも、イベント検知の精度を最大 30%強高めることを確認した。これらはメールのやりとりのデータからの不正検知、複数企業の株価時系列データからのイベント検知等に活用できる。

2)の関係データ統合予測とは、属性、関係、画像、音声等の多様なデータが存在するとき、これらに関係づけるディープナレッジを抽出し、欠損したデータの補完や予測に活用する研究である。今年度はディープナレッジ表現として「制限付きボルツマンマシン」を用いて、これを効率的に学習し、データ補完するための新しいアルゴリズムを開発した。これにより統計的マッチング等の従来手法より最大 20%高い補完率を実現した。本手法はマーケットの潜在需要の推定等に適用できる。

実験研究では、東大医学部眼科教室から緑内障患者の視野測定データの提供を受け、緑内障進行予測の問題に取り組んだ。従来手法では、個別患者のデータのみを用いて線形回帰しており、そのデータの少なさが問題であった。今年度は、「時空間パタンクラスタリング予測手法」を開発した。これは大規模患者データの時空間進行パタンをクラスタリングし、似ているパタンの患者データを用いて予測する方法である。本手法により、従来に比べ、検査回数2回にて平均 60%以上の予測精度を達成した。今後は、疾患典型パタンの抽出や疾患因子の特定等に拡大していく。

増田 G では、時間的に構造変化するネットワークであるテンポラルネットワーク(TN)の研究を行っている。ネットワークにはノードの中心性(重要性)という概念があり、静的なネットワークではランダムウォークの訪問回数で定められる中心性がページランクなどに活用されてきた。今年度は、動的に変化し得る TN に対して、ランダムウォークに基づく中心性を世界で初めて定式化した。本指標は、今後、時間的に変動する社会や脳内のネットワークの構造推定に適用して有効性を検証する予定である。

IBMG では、行動データを対象とするディープナレッジの抽出と活用の研究に取り組んでいる。今年度は、A)人の選択行動 と、B)逐次的意思決定 の基本モデルを開発した。特に、A)では、人が与えられた選択肢の中から何かを選ぶ過程には、魅力・類似効果・妥協等の潜在因子が関係することが知られている。これを「制限付きボルツマンマシン」と呼ばれる潜在変数モデルを用いてモデル化することで、従来のロジットモデルと比べて選択確率の予測誤差を最大40%程度削減できることを実証した。本手法はサービスの選択等の複雑で非合理的な選択の予測に有効である。

大澤 G では、データが価値を持つにいたるシナリオの研究を行っている。データが価値をもつためには、他のデータと一緒に、ある目的をもって解析される必要がある。そこで、どのようなデータと

結合して、どのような解析を行ったらよいかというシナリオを議論するための方法論として、IMDJ (Innovators Marketplace on Data Jacket)を提案している。そこでは、データは生のままではなく、ダイジェスト化されたDJ(Data Jacket)として扱われる。今年度は、IMDJを通じて、1)ディープナレッジを抽出するための要素技術と、2)それが抽出されたかどうかを定量的に評価する方法を開発した。1)には、DJ収集法(約200件の実績)、DJ関係の可視化ツール、オントロジーに基づく異種データ結合方式が含まれる。2)には、IMDJにおける要求頻度に基づく定量化があげられる。IMDJは、自動車産業などの製造業において、顧客のニーズをもとに新機能を生み出すシナリオ作成などに用いられる。

§ 3. 成果発表等

(3-1) 原著論文発表

論文詳細情報(国内)

なし

論文詳細情報(国際)

A-1. Shoichi Sato and Kenji Yamanishi, “Graph partitioning change detection using tree-based clustering,” Proceedings of IEEE International Conference on Data Mining (ICDM2013), pp:1169–1174, 2013 (DOI: 10.1109/ICDM.2013.75).

A-2. Zenghan Liang, Ryota Tomioka, Hiorshi Murata, Ryo Asaoka, Kenji Yamanishi, “Quantitative prediction of glaucomatous visual field loss from few measurements,” Proceedings of IEEE International Conference on Data Mining (ICDM2013), pp:1121–1126, 2013 (DOI:10.1109/ICDM.2013.93).

A-3. Sho-ichi Sato and Kenji Yamanishi, “Graph partitioning change detection using tree-based clustering,” ACM Transactions on Knowledge Discovery from Data (投稿中).

A-4. Yoshiki Sakai and Kenji Yamanishi, “Data fusion using restricted Boltzmann machines,” Proceedings of ACM Conference on Knowledge Discovery and Data Mining (KDD2014) (投稿中).

A-5. Takuya Akasaka and Kenji Yamanishi, “Enhancing change-point detection using independent component analysis,” Proceedings of ACM Conference on Knowledge Discovery and Data Mining (KDD2014) (投稿中).

A-6. Yoshiki Sakai and Kenji Yamanishi, “An NML-based model selection criterion for general relational data modeling,” IEEE Transactions on Knowledge and Data Engineering (投稿中).

A-7. Hideki Nakayama, “Stacked Local Autocorrelation Features: Revisiting HLAC with Deep Approach”, IPSJ Transactions on Computer Vision and Applications (CVA) Express Paper (投稿中).

A-8. Shin Matsushima, Hyokun Yun, S.V.N. Vishwanathan, “Doubly Separable Stochastic Optimization for Regularized Empirical Risk Minimization”, International Conference on Very Large Data Bases (VLDB2014) (投稿中)

A-9. Ryo Asaoka, Aiko Iwase, Tae Tsutsumi, Hitomi Saito, Shinichiro Otani, Kazunori Miyata, Hiroshi Murata, Chihiro Mayama, Makoto Araie, “Combining multiple HRT parameters using the ‘Random Forests’ method improves the diagnostic accuracy of glaucoma, in emmetropic and highly myopic eyes,” *Invest Ophthalmol Vis Sci.* 2014 Mar 7. pii: iovs.14-14009v1. (DOI: 10.1167/iov.14-14009).

A-10. Hiroyo Hirasawa, Hiroshi Mutara, Chihiro Mayama, Makoto Araie, Ryo Asaoka, “Evaluation of various machine learning methods to predict vision related quality of life from visual field data and visual acuity in patients with glaucoma,” *British Journal of Ophthalmology* (in press).

B-1. Naoki Masuda, “Evolution via Imitation among Like-Minded Individuals”, *Journal of Theoretical Biology*, vol. 349, pp. 100-108, 2014 (DOI:10.1016/j.jtbi.2014.02.003)

B-2. Koji Oishi, Manuel Cebrian, Andres Abeliuk, Naoki Masuda, “Iterated Crowdsourcing Dilemma Game”, *Scientific Reports*, vol. 4, article No. 4100, 2014 (DOI:10.1038/srep04100).

B-3. Takamitsu Watanabe, Satoshi Hirose, Hiroyuki Wada, Yoshio Imai, Toru Machida, Ichiro Shirouzu, Seiki Konishi, Yasushi Miyashita, Naoki Masuda, “Energy Landscapes of Resting-State Brain Networks”, *Frontiers in Neuroinformatics*, vol. 8, article No. 12, 2014 (DOI:10.3389/fninf.2014.00012).

B-4. Yasuhiro Mochizuki, Shigeru Shinomoto, “Analog and Digital Codes in the Brain”, *Physical Review E*, vol. 89, article No. 022705, 2014 (DOI:10.1103/PhysRevE.89.022705)

C-1. Takayuki Osogami and Takayuki Katsuki, “A Bayesian hierarchical choice model with visibility,” in *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR 2014)*, August 2014 (accepted)

C-2. Takayuki Osogami and Makoto Otsuka, “Restricted Boltzmann machines modeling human choice,” *The 28th Conference on Artificial Intelligence (AAAI-14)* (投稿中)

C-3. Takayuki Osogami, “Tailoring sequential decision making by learning personality traits,” *The 28th Conference on Artificial Intelligence (AAAI-14)* (投稿中)

C-4. Tetsuro Morimura, Takayuki Osogami, and Tomoyuki Shirai, “Mixing-time regularized policy gradient,” *The 28th Conference on Artificial Intelligence (AAAI-14)* (投稿中)

C-5. Satoshi Hara, Tetsuro Morimura, Toshihiko Takahashi, Hiroki Yanagisawa, and Taiji Suzuki, “Anomaly localization via the sparsest subgraph problem,” The 31st International Conference on Machine Learning (ICML 2014) (投稿中)

C-6. Takayuki Osogami and Rudy Raymond, “An efficient policy iteration for cyclic Markov decision processes,” The 21st European Conference on Artificial Intelligence (ECAI 2014) (投稿中)

C-7. Rudy Raymond, Teruo Koyanagi, and Takayuki Osogami, “An Approximate Counting for Big Textual Data Stream,” The 21st European Conference on Artificial Intelligence (ECAI 2014) (投稿中)

C-8. Makoto Otsuka and Takayuki Osogami, “Dynamic Boltzmann machines with homeostatic spike-timing dependent plasticity,” Science (再投稿準備中)

D-1. Teruaki Hayashi, Yukio Ohsawa, “Processing Combinatorial Thinking: Innovators Marketplace as Role-based Game plus Action Planning,” International Journal of Knowledge and Systems Science, Vol.4(3), pp.14-38, 2013. (DOI: 10.4018/ijkss.2013070102).

D-2. Yukio Ohsawa, Hiroyuki Kido, Teruaki Hayashi, Masahiro Akimoto, Masanori Fujimoto, Masaki Tamada, “Strategies for Creative Argumentation: Learned from Logs of Innovators Market Game,” Brain & Health Informatics (BHI13), Lecture Notes in Computer Science Volume 8211, pp 369-378 (2013) (DOI : 10.1007/978-3-319-02753-1_37).

D-3. Hiroyuki Kushi, “A Method for Generating Ontologies in Requirements Domain for Searching Data Sets in Marketplace,” Proceedings of IEEE International Conference on Data mining, Market of Data workshop (MoDAT2013), pp. 688-693, Dec.2013 (DOI:10.1109/ICDMW.2013.54).

D-4. Chang Liu, Yukio Ohsawa, and Yoshitaka Suda, “Valuation of Data through Use Scenarios in Innovators’ Marketplace on Data Jackets,” Proceedings of IEEE International Conference on Data mining, Market of Data workshop (MoDAT2013), pp. 694 - 701 (DOI: 10.1109/ICDMW.2013.120).

D-5. Jun Nakamura, Masahiko Teramoto, “Frame as a Clue to Intention of Data: Toward New Product Ideas with Framed Components,” Proceedings of IEEE International Conference on Data mining, Market of Data workshop (MoDAT2013), pp. 720-725, Dec.2013 (DOI:

10.1109/ICDMW.2013.28).

D-6. Masahiko Teramoto, Jun Nakamura, “e-Trucks Realize Four Zeros Expectations, The Challenge by Market of Data,” Proceedings of IEEE International Conference on Data mining, Market of Data workshop (MoDAT2013), pp. 726-732, 2013 (DOI: 10.1109/ICDMW.2013.140).

D-7. Hiroyuki Kido, Yukio Ohsawa, “Justifying Underlying Desires for Argument-based Reconciliation,” in Post Proceedings of the 2nd International Workshop on Theory and Applications of Formal Argumentation (TAFA2013), vol. 8306 of LNAI, pp. 143-157, Springer (2014) (DOI: 10.1007/978-3-642-54373-9_10).

D-8. Yukio Ohsawa, Chang Liu, Yoshitaka Suda, and Hiroyuki Kido, “Innovators Marketplace on Data Jackets for Externalizing the Value of Data via Stakeholders’ Requirement Communication,” in Proceedings of AAI 2014 Spring Symposium on “Big data becomes personal: Knowledge into Meaning, March 2014, Stanford (2014) (in Press)