

「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 27 年度採択研究代表者

H29 年度  
実績報告書

津田 宏治

東京大学大学院新領域創成科学研究科  
教授

離散構造統計学の創出と癌科学への展開

## § 1. 研究実施体制

### (1) 津田グループ

① 研究代表者: 津田 宏治 (東京大学大学院新領域創成科学研究科 教授)

② 研究項目

・離散構造統計学の創出・普及

### (2) 門松グループ

① 研究代表者: 門松 健治 (名古屋大学大学院医学系研究科 教授)

② 研究項目

・癌検体の収集、実験データの取得および介入実験

### (3) 瀬々グループ

① 研究代表者: 瀬々 潤 (産業技術総合研究所人工知能研究センター 研究チーム長)

② 研究項目

・統計的検定手法構築、高速化、大規模化及び手法の普及

### (4) 竹内グループ

① 研究代表者: 竹内 一郎 (名古屋工業大学大学院工学研究科 教授)

② 研究項目

・網羅的遺伝情報の複合要因探索アルゴリズム構築・ソフトウェア実装・癌科学における実証

### (5) 山田グループ

① 研究代表者: 山田 亮 (京都大学大学院医学研究科 教授)

② 研究項目

・離散構造統計学の遺伝疫学・コホートスタディへの展開

## § 2. 研究実施の概要

今年度は、新たな統計的仮説検定法として近年注目を集めている Selective Inference に関連する研究を実施した。Selective Inference とは、データに基づいて選択された仮説を同一のデータで評価する際に、選択バイアスを除去するため、仮説を選択するイベントの条件付分布により統計的仮説検定を行うものである。これまで、このような条件付分布の計算は困難であるとされてきたが、近年、仮説選択のアルゴリズムがあるクラスの演算で記述できるときには、正確に条件付分布の計算が可能となることが示された。本研究では、この考え方を発展させ、高次交互作用モデルのための Selective Inference 法を開発した。

本年度は dbGaP 等の遺伝情報データを含むデータベースに対して、LAMP/FastWY を適用し、順次結果の整理を進めた。この過程で、がん情報の遺伝情報に関する実験結果が蓄積されている TCGA では、LAMP/FastWY で扱っていた 疾患有無のバイナリの値ではなく、生存時間および最終観測時間で表されるサバイバルデータで、疾患の状態が表されていることがわかり、TCGA 等に LAMP/FastWY を適用するには、サバイバルデータを解析できるようにする必要があった。このため、LAMP をサバイバルデータの検定ができるように拡張した Survival LAMP を開発し、<https://rtrelator.github.io/SurvivalLAMP/> から公開を行った。

組み合わせ要因発見には、非常に大きな計算量を要するため、アルゴリズムの並列化は避けては通れない課題である。並列化には、共有メモリ型の並列化と、分散メモリ型の並列化があるが、Amazon Web Service などのクラウド上での利用や、京コンピュータなどのスーパーコンピュータでの利用を考えると、分散メモリ型の並列化が必要である。MP-LAMP では、動的 load balancing の技術を採用し、動作中に計算機間でタスクのやりとりを行って、計算機の空転を防ぐ。TSUBAME 上で、1200 ノードを使用した際、GWAS データで 1174 倍を記録するなど、ほぼ理想的な結果を得た。Amazon Web Service、京コンピュータ、Oakforest-PACS でも同様の結果を得ている。

代表的な原著論文

S. Suzumura, K. Nakagawa, Y. Umezumi, K. Tsuda, I. Takeuchi, Selective Inference for Sparse High-Order Interaction Models, Proceedings of the 34th International Conference on Machine Learning (ICML 2017), pp. 3338-3347, 2017.

R.T. Relator, A. Terada and J. Sese, Identifying statistically significant combinatorial markers for survival analysis, BMC Medical Genomics, 11(Suppl 2), 31, 2018.

K. Yoshizoe, A. Terada and K. Tsuda, MP-LAMP: Parallel Detection of Statistically Significant Multi-Loci Markers on Cloud Platforms, Bioinformatics, published online, 2018.