

「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

H27 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 (国立大学法人 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 (国立大学法人 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 (国立研究開発法人 産業技術総合研究所・創薬基盤研究部門・主任研究員)
- 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 (国立大学法人 名古屋工業大学・大学院工学研究科・准教授)
- 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 (国立大学法人 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

平成27年度は、生命科学と統計理論の両面で研究を行った。門松 G においては、SMG 組織から単細胞懸濁液を得るための実験手法を確立し、一細胞レベルで細胞を単離するなど、神経芽腫の原因解明に向け、一細胞 RNA-seq 実験の準備を着実に進めることができた。山田 G においては、長浜コホートなどの集団疫学データに LAMP などの離散統計学手法を適用した。一方、他のグループにおいては、理論的な研究が大幅に進展した。主な成果は、以下の3つである。

Selective Inference による組み合わせ要因検定手法の開発(竹内・津田 G)

Selective inference (選択的推論)とは、Stanford 大学の Taylor, Tibshirani らによって提唱されている新たな統計的検定の方法である。ビッグデータ解析において、多数の仮説があるときに、予め仮説数を絞る操作(事前選択)が行われるが、選択的推論では、選択イベントに対して条件づけた P 値を計算することによって、事前選択によるバイアスを補正することができる。我々は、この方法を、パターンマイニングと組み合わせることにより、組み合わせ要因発見・検定を行うアルゴリズムを開発し、HIV の薬剤耐性データ・遺伝子発現データなどに適用した。論文は現在投稿中である。

遺伝学データ解析ツール LAMPLINK の開発・公開(瀬々・津田 G)

本プロジェクトの課題の一つは、ヒトのゲノム上の多型(polymorphism)と疾病との関係を明らかにするアルゴリズム及びソフトウェアの開発である。今年度、ヒトの多型データに、組み合わせ検定手法 LAMP (無限次数多重検定法)を簡単に適用できるソフトウェア LAMPLINK を開発し、公開した。このソフトウェアを用いれば、既存の解析パイプラインの一部を変更するだけで LAMP が適用できるため、LAMP の普及が促進されることが期待できる。本ソフトウェアを用いて、日本人に特異的に存在する多型の組み合わせや、アルツハイマー病に関連する多型を発見することができた。

情報幾何に基づく組み合わせ要因検定手法の開発(津田 G)

本成果は、当領域のさきがけ研究員である杉山磨人助教(大阪大学)との共同研究の結果である。情報理論を用いて、複数要因の相乗作用を計測するためのアルゴリズムは、これまでに、いくつか発表されている。その中でも、甘利によって2001年に発表された、確率分布の情報幾何的分解の方法は、非常に有名である。本研究では、これまで、仮説のべき集合だけに適用可能であった甘利理論を拡張し、半順序関係を持つ仮説集合に拡張した。その際、これまでに知られていなかった、順序理論と情報幾何の間関係を初めて明らかにした。この理論は、がん科学・脳科学・空間疫学などに広く適用可能である。論文は、情報理論分野のトップ会議である、International Symposium on Information Theory (ISIT) に採録された。

代表的な原著論文

A. Terada, R. Yamada, K. Tsuda and J. Sese, LAMPLINK: detection of statistically significant SNP combinations from GWAS data, Bioinformatics, 2016, in press.

M. Sugiyama, H. Nakahara, K. Tsuda, Information Decomposition on Structured Space, 2016 IEEE

International Symposium on Information Theory, 2016 (to appear)