

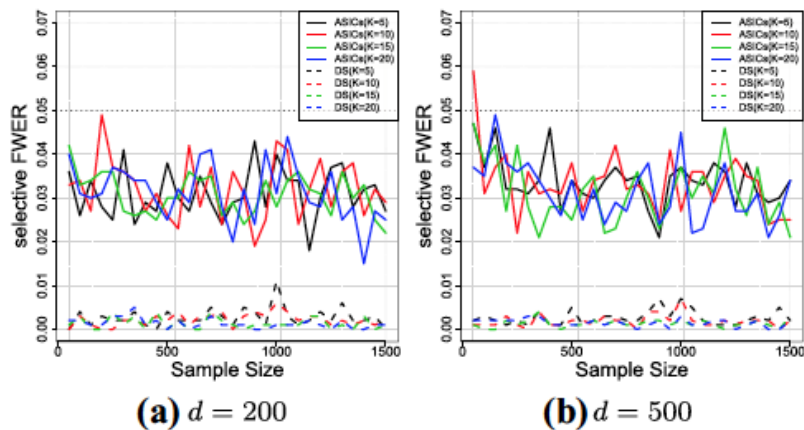
津田 宏治

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. 研究成果の概要

本研究では、選択的推論や、多重検定といった多次元データ解析に不可欠な統計検定手法の開発と、その癌科学への応用を目的としている。最近では、癌科学だけでなく、医療全般への応用にも取り組んでいる。2019年度の理論的な進展としては、論文1に発表された二値分類問題における特徴選択に関する選択的推論手法の開発があげられる。これまでは、LASSOなど線形回帰に関する選択的推論手法は知られていたが、これらの手法では、ノイズの正規性を仮定しており、分類問題には適用できなかった。論文1では、Marginal screeningと呼ばれる特徴選択手法をロジスティック回帰に適用した場合に関して、新たな選択的推論手法 ASIC (Asymptotic Selective Inference for Classification)を提案し、漸近論を用いて正当性を証明した。データ分割(Data Splitting, DS)による手法との比較を下図に示す。ここでは、Family-wise error rate (FWER)を0.05を超えない範囲で近づけるのが目標となるが、特徴空間の次元数 d 、選択された特徴数 K によらず、ASIC(実線)のほうがDS(破線)よりも、FWERを適切にコントロールできていることがわかる。



また、国立循環器病研究センターの北風政文教授との共同研究の一環として、急性心筋梗塞を主なターゲットとし、約 3 万人の定期検診のデータ、および病院の受診歴や診断データの解析を進めた。これまでに、定期検診後 5 年間の急性心筋梗塞に統計的に有意に関連している検診項目として、年齢・血圧・中性脂肪の組み合わせなど、多数の予測因子の候補が検出されている。現在、得られた結果をもとに、今後の解析の方向性や論文化だけではなく、定期検診や人間ドックに結果を組み込むような実用化も視野に入れ、議論を進めている。

【代表的な原著論文】

1. Umezu Y., Takeuchi I. Selective inference via marginal screening for high dimensional classification. *Japanese Journal of Statistics and Data Science*, vol. 2, pp. 559-589 (2019)
2. Mahito Sugiyama, Hiroyuki Nakahara, Koji Tsuda. Legendre decomposition for tensors. *Journal of Statistical Mechanics*, 124017, 2019.
3. Sakuma T. et al., Efficient learning algorithm for sparse subsequence pattern-based classification and applications to comparative animal trajectory data analysis. *Advanced Robotics*, vol. 33, pp. 134-152 (2019)

§ 2. 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 ((株)ヒューマノーム研究所 代表取締役社長)
- 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