

戦略的創造研究推進事業 AIP 加速課題
研究課題「信頼できるデータ駆動科学のための統計学の深化」

研究終了報告書

研究期間 2021年04月～2024年03月

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)

§ 1 研究実施の概要

(1)実施概要

選択的推論法の開発においては、竹内 G を中心に、説明可能な機械学習分野で有力な手法である高次交互作用モデルの選択的推論法の開発を行った。また、従来手法に比べて、検出力の高いホモトピー法に基づく手法を開発した。さらに、選択的推論を、深層学習による画像セグメンテーションに適用し、医用画像において有効な結果を得た。これらは、選択的推論の適用範囲を広げ、検出力を高める成果である。一方、津田 G を中心に行なった量子計算機の利用においては、量子アニーラを統計検定に必要なサンプリングに利用するための技術 Fast Stein Correction を開発した。これにより、ターゲット分布からの正確なサンプリングが可能になり、統計検定のみならず、量子アニーラのサンプリング目的での利用が促進される。

生命科学への応用においては、前身の CREST 時代から継続して研究している神経芽腫自然退縮機構の解明に関して、退縮に関与する神経芽腫サブタイプの同定などの成果が得られた(門松 G)。また、多重検定手法 LAMP の応用により、健康診断データの解析、および、食品タンパク質のアレルギー反応予測を行なった。また、(株)ヒューマノーム研究所において、LAMP をユーザのデータに対して簡単に利用できる有償サービスを開始する (瀬々 G)。

(2)顕著な成果

<優れた基礎研究としての成果>

1. 高次交互作用モデルの選択的推論法開発

概要:

本成果は選択的推論の枠組を用いて高次交互作用モデルの統計的推測を行うための方法を確立したものである。高次交互作用モデルとは特徴の高次交互作用項を持つような線形モデルで、解釈性を保ちつつ複雑なモデリングが可能のため、さまざまな分野で活用されている。高次交互作用モデルのスパース推定においては、分子限定法に基づく枝刈り操作が必要なため、従来の選択的推論法をそのまま適用するのは困難であったが、分子限定法とパラメトリック計画法を効果的に組み合わせたアルゴリズムを開発したことでこの問題を克服し、妥当性を持ちつつ検出力の高い統計的推測が初めて可能となった。この成果は人工知能分野のトップ会議の一つである AAI2021 に採択された。

2. ホモトピー法に基づく選択的推論

概要:

本成果は従来の選択的推論において実用の障壁となっていた検出力を向上させる方法を確立したものである。従来の選択的推論では仮説選択イベントを線形不等式、もしくは、二次不等式の積集合として特徴づける必要があったが、本論文で提案した方法により線形/二次不等式の積集合の和集合として表される選択イベントに対しても選択的推論が可能となった。この拡張により、従来の選択的推論法の検出力が大幅に向上しただけでなく、これまでは適用が困難であったアルゴリズムにも選択的推論を適用できるようになった。この成果は機械学習分野のトップ会議の一つである ICML2021 に採択され、国内外から注目を集めている。

3. 深層学習モデルに関する選択的推論

概要:

本成果は選択的推論法を深層学習モデルに基づく画像セグメンテーションの信頼性評価に関するものである。深層学習モデルでは複雑な演算が行われるため、従来の選択的推論を適用することは困難であったが、本研究では、幅広いクラスの畳み込みニューラルネットワークが区分線形関数として表現できることに着目し、これを実現した。提案法を脳腫瘍領域同定問

題に適用し、従来は困難であった正確な P 値を与えることができるようになった。この成果は機械学習分野のトップ会議の一つである Neurips2022 に採択された。

< 科学技術イノベーションに大きく寄与する成果 >

1. 神経芽腫自然退縮機構の解明

概要:

神経芽腫マウスモデル及びヒト神経芽腫検体を対象にシングルセル遺伝子発現解析を行い、今まで謎に包まれていた神経芽腫の自然退縮機構について一定の成果を得ることができた。特にマウスモデルにおいて自然退縮様現象を見出し、自然退縮に寄与する新しい神経芽腫細胞のサブタイプや遺伝子群の同定、癌化に寄与する遺伝子群の同定を行った。この成果は、これら遺伝子群をターゲットにした新たな治療戦略となる可能性を秘めている。

2. LAMP による食品タンパク質のアレルギー反応予測

概要:

本成果は食品タンパク質のアレルギー反応予測の問題に、瀬々グループ、津田グループらで開発した LAMP を基盤とした技術を活用したものである。ゲノム編集などの技術発展により食品の人工的な「編集」が可能となりつつあるが、編集された食品に対するアレルギー反応のリスクを適切に管理する必要がある。本論文では、アミノ酸配列に基づくタンパク質性能予測モデルに LAMP を基盤とする技術を活用したもので、従来のアレルギー反応予測よりも高精度で解釈可能なモデルを開発することができた。本論文は関連分野のトップジャーナルの1つである、Journal of Biological Chemistry で出版された。

3. 量子アニーラによるサンプリング技術の開発

概要:

本成果では、Fast Stein Correction という分布補正技術を開発し、量子アニーリングで生成されたサンプルへの応用を行なった。統計検定や、機械学習において、与えられた分布からサンプリングする技術は重要であるが、量子アニーリングのサンプリング分布は解析的に表すことができない。我々の手法は、分布が解析的に表せない場合にも補正することが可能であり、統計検定も含めた様々な応用が可能である。

< 代表的な論文 >

1. Diptesh Das, Vo Nguyen Le Duy, Hiroyuki Hanada, Koji Tsuda, Ichiro Takeuchi. Fast and More Powerful Selective Inference for Sparse High-Order Interaction Model. 36th AAAI Conference on Artificial Intelligence, 2022.

概要:この論文は、優れた基礎研究の成果1に対応しており、高次交互作用モデルでの選択的推論を実現したものである。

2. Kazuya Sugiyama, Vo Nguyen Le Duy, Ichiro Takeuchi. More Powerful and General Selective Inference for Stepwise Feature Selection using Homotopy Method. Proceedings of International Conference on Machine Learning 2021 (ICML2021), 2021.

概要:

この論文は、優れた基礎研究の成果 2 に対応しており、ホモトピー法による選択的推論によって、検出力を向上したものである。

3. 3. Vo Nguyen Le Duy, Shogo Iwazaki, Ichiro Takeuchi. Quantifying Statistical Significance

概要:

この論文は、優れた基礎研究の成果 3 に対応しており、深層学習による画像分割への選択的推論を実現したものである。

§ 2 研究実施体制

(1)研究チームの体制について

津田グループ

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)

研究項目

・アルゴリズム開発と量子計算機応用

門松グループ

主たる共同研究者:門松 健治(名古屋大学 糖鎖生命コア研究所 所長)

研究項目

・癌科学への展開

瀬々グループ

主たる共同研究者:瀬々 潤 (株式会社ヒューマノーム研究所 代表取締役社長)

研究項目

・健康科学への展開

竹内グループ

主たる共同研究者:竹内 一郎 (名古屋大学 大学院工学研究科 教授)

研究項目

・汎用的な選択的推論手法の構築

当初は、この他に山田グループが存在したが、山田亮教授の退職に伴い解消された。

(2)国内外の研究者や産業界等との連携によるネットワーク形成の状況について

門松グループは、シングルセル遺伝子発現解析のために東京大学大学院新領域創成科学研究科メディカル情報生命専攻の鈴木穰研究室、名古屋大学大学院医学系研究科システム生物学の島村徹平研究室、埼玉県立がんセンター臨床腫瘍研究所の上條岳彦研究室、Children's Cancer Institute (オーストラリア)の Daniel Carter 先生など、国内外の共同研究者と協力し解析を行った。