

統計的機械学習研究室
Statistical Learning助教 中村 知繁
Assistant Professor
Tomoshige Nakamura

研究概要 / Research overview

機械学習を用いた因果推論による**個人や個別の状況に応じた最適な介入を見出すための手法の開発**を行っています。機械学習による因果推論には、**3つの大きな壁**が存在します。

- ① **データの高次元性への対処**：無数の要因が考えられる中からいかにして正確な因果効果を推定するか？
- ② **ブラックボックス性への対処**：高度なモデルが生み出す予測の根拠をどのように示すか？
- ③ **モデル選択**：多数の推定手法の中から、信頼性の高いものをどのように選択するか？

私たちの研究室は、これら3つの課題に対し、統計的学習理論と実践的アルゴリズムの両面から統合的にアプローチすることで、機械学習による信頼できる統計的因果推論の手法の開発を行います。

最新の研究成果 / Latest research outputs

Shrinkage Bayesian Causal Forests (Shrinkage BCF) の事後分布収束レートの導出

モデル

観測データ Y_i を、ベースライン $\mu(X_i)$ と因果効果 $\tau(X_i)$ を用いて以下のようにモデル化を行う

$$Y_i = \mu(X_i) + \tau(X_i) \cdot Z_i + \varepsilon_i \quad (Z_i \in \{0,1\})$$

各関数 $\mu(\cdot), \tau(\cdot)$ に対して BART 事前分布を置く。さらに、BART事前分布に対して木を成長させる際の変数選択確率 s_μ, s_τ に対して、ディリクレ事前分布を置く

$$s_k \sim \text{Dirichlet} \left(\frac{\alpha}{p^\xi}, \dots, \frac{\alpha}{p^\xi} \right), \quad (k \in \{\mu, \tau\}, \xi > 1)$$

この事前分布は観測された p 個の変数から、予測に有効な変数にのみ高い選択確率を割り当てることで、冗長な変数の影響を自動的に抑制します (Shrinkage 効果)

主結果

Shrinkage BCF モデルは、真の関数のペア (μ_0, τ_0) に対して以下の縮退レートを持つ

$$\varepsilon_n^2 = \sum_{j \in \{\mu, \tau\}} n^{-\frac{2\alpha_j}{2\alpha_j + d_j}} \log n + \frac{d_j}{n} \log p$$

結果の解釈

収束率が理論的に達成可能な最適レート（ミニマックス最適レート）に（対数因子を除き）一致。Shrinkage BCFは、事前に未知である関数の滑らかさ α_j と真の変数選択の難しさ（ p 個から d_j を選ぶ）の両方に適応しながら、理論的に最も効率の良い学習を行っていることが示された。

メッセージ / Message

私たちの研究は、医学や経済学などさまざまな応用分野で利用される統計・機械学習手法を作り、その取扱説明書を作ることです。例えば、Sh-BCFの理論解析の結果は、複雑で高次元なデータが絡む因果推論の問題に対し、安心してShrinkage BCFが適用できることを示しています。このような結果が、データ分析から得られる結果の基盤を支えています。

現在、私たちは木構造モデル（一般化ランダムフォレスト、ベイズ加法回帰木）と呼ばれる機械学習手法を中心に、新たな推定手法の理論開発、モデルの解釈可能性の向上のための手法の提案に取り組んでいます。確率論や数理統計学への深い理解が必要で大変なことも多いですが、じっくりと理論に向き合いたい、この分野にチャレンジしたいという方からのコンタクトをお待ちしています。