

「統計的機械学習」の中核としての 統計数理 開催挨拶とオリエンテーション

情報・システム研究機構 統計数理研究所
椿 広計

御礼

Preferred Networks 代表取締役

役 最高研究責任者 岡野原大

輔先生

ご参加いただいた全ての方々へ

通常の統計数理研究所産学連携シンポジウムでの

樁の役割：枕としての歴史オリエンテーション

Chat GPTに聴いててみました

司会の北村さん: 「chat GPTをパネルのテーマの一つにしては?」

椿: 「時間もない中でそれはどうかな」

オリエンテーションの枕にしよう?!

優等生的回答! とても「chat椿」ではかないません

統計学、数理科学が統計的機械学習の中核をなす理論であると考えられますか？

- 統計学と数理科学は、統計的機械学習の中核をなす理論と考えられます。
 - 統計学は、データの解析や推測における基本的な理論や手法を提供し、データの収集・分析・解釈における確率的なアプローチを基盤としています。
統計学は、データのモデリング、推定、仮説検定、信頼区間などの概念や手法を提供することで、データに関する情報を把握し、意思決定や予測に役立てるための基礎を提供します。
 - 一方、数理科学は数学的な手法や理論を応用して、様々な現象や問題をモデル化し解析する学問領域です。数理科学は、確率論、最適化理論、線形代数、微分方程式などの数学的な枠組みを提供し、複雑なデータや問題に対する数学的なモデリングや解析を可能にします。
- 統計的機械学習は、統計学や数理科学の理論と手法を応用して、データからパターンや関係性を学習し、予測や意思決定を行うためのモデルを構築します。
- **統計学や数理科学の基礎的な概念や手法（例えば確率論、最適化、線形代数など）を統合し、
機械学習の枠組みに適用することで、統計的機械学習の理論や手法が形成されます。**
- したがって、**統計学と数理科学は統計的機械学習の基盤となる理論であり、
統計的機械学習の中核を形成していると考えられます。**これらの理論と手法を活用することで、データに基づく問題の解決や予測モデルの構築をより効果的に行うことができます。

統計学や数理学はどのようにして 統計的機械学習の理論と応用に貢献するのですか？

1. データ解析の基礎となる理論の提供:

統計学はデータの収集、整理、要約、分析における基本的な理論や手法を提供します。統計学の概念やモデル（確率分布、統計的推定、仮説検定など）を用いることで、データの特性や分布、パターンを把握し、データ解析の基礎を築くことができます。

2. データのモデリングと統計的推定:

統計学はデータのモデリングにおいて重要な役割を果たします。統計的モデルの構築やパラメータの推定に基づいて、データからの情報や関係性を抽出することが可能です。統計学の手法を応用することで、データの背後にある確率的なプロセスやパターンを推定し、データの特徴を捉えることができます。

3. 機械学習アルゴリズムの開発と解析:

数理学は、機械学習アルゴリズムの開発と解析において重要な役割を果たします。数学的な枠組みや最適化手法、確率論、線形代数などの数理的な手法を用いて、機械学習アルゴリズムの理論的な特性や収束性、最適性を解析します。数理学の応用により、機械学習アルゴリズムの開発やパフォーマンスの評価が進められます。

4. モデルの解釈と説明可能性:

統計学や数理学は、モデルの解釈と説明可能性に関しても貢献します。統計的な手法やモデルを使用することで、データのパターンや関係性を理解し、結果を解釈することが可能です。また、数理的な手法やモデルの特性を活用して、モデルの予測結果や重要な変数の特定、影響の評価などを行うことができます。

- 統計学や数理学の貢献により、統計的機械学習はより信頼性の高いモデルの構築やデータ解析が可能となります
- **データの理解や予測の精度向上、特徴の選択や次元削減、異常検出などの応用において、統計学や数理学は重要な役割**を果たします。

悪戯質問：特異値分解を提案したのは誰で何時でしょう か？

- 特異値分解 (Singular Value Decomposition, SVD) を提案したのは、1965年にGene H. GolubとWilliam Kahanの2人の数学者です。彼らはそれぞれ独立にSVDを研究し、その後共同で論文「Singular Value Decomposition and Least Squares Solutions」を發表しました。この論文は、1965年に出版されたLinear Algebra誌に掲載されました。GolubとKahanの貢献により、特異値分解は線形代数や数値解析の重要なテクニックとして広く利用されています。
- **Chat 椿**→特異値分解と数理的に同値な行列データの乗法近似を提案したのはFisherです。
 - Fisher and Mackenzie(1923) Studies in crop variation. II. The manurial response of different potato varieties, *The Journal of Agricultural Science*:
 - 要因効果の加法的分解 (分散分析) と乗法的分解 (特異値分解) :後者の行列べき近似解法
 - $Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$:同一論文の分散分析が有名になりすぎた
 - $Y_{ij} = \mu \times \alpha_i \times \beta_j + \varepsilon_{ij}$: **対数線形モデルとは異なる乗法モデル**
 - Tukey(1962) Future of Data Analysis, Annals of Mathematical Stat. で意義強調

統計理論の古典的貢献

Tukey以前の主としてモデル論的貢献

- Gauss(1809) 誤差論
 - 最小二乗推定量の**逐次推定** (One Rank Modification)
- 特異値分解の原型：主成分分析(Pearson, 1901)
- ニューラルネットの原型：因子分析(Spearman, 1903)
 - Goodfellow et al.(2016) 「深層学習」で今後の研究テーマとされる
- グラフィカルモデルの原型：偏相関分析 (Pearson, 1904),
- Fisher(1921) On the Mathematical Foundation of Statistics,
 - 尤度原理の確立：最小二乗推定→最尤推定 (KL Distance)
 - 先ほどの加法近似と乗法近似も重要
- Wald (1950) 統計的決定理論: 統計家とNatureのゲーム
 - 逐次推論：逐次尤度比も1940年代提唱
- Gower(1956) Metric MDS:
 - $n \times n$ 距離行列から $p \times p$ 共分散行列の主成分分析を再現

EDA: Exploratory Data Analysis登場

Tukeyの探索的データ解析後の貢献

- **Tukey (1962) Future of Data Analysis: 2012年はデータサイエンス50周年と**
いった動きも
 - 1958: ジャックナイフ提唱 (n -foldクロスバリデーション) も提唱
 - Efron 1979: Bootstrap 経験分布に基づく推論
 - **Tukey→Breiman, Friedman, Hastie, Tibshirani**
- CART(Tree Based Model)の源流: AID(Automatic Interaction Detector: 自動層別)
- Morgan, J. N. and Sonquist, J. A.(1963)
 - 第2世代AIの旗手: **Breiman et al. (1984) Classification and Regression Trees**
 - Breiman,1996:**Bagging** Predictors (アンサンブル学習) も提唱→第3世代のRandom ForestはFriedman
- **Bayes決定**の許容性(正則化の正当なアプローチ) : Lindley(1965)
- 甘利俊一(1978)神経回路網の数理
- 甘利俊一: 情報幾何学の構想と発展 (Efron, 1975の統計モデルの曲率への注目)
 - 深層学習の多様体構造へと発展: 曲率の小さいモデルと大域的予測誤差との関係性
 - Calin (2020) Deep Learning Architectures, Springer
- Friedman (1981) Projection Pursuit Regression→GAMとの結合
 - Hastie and Tibshirani(1984)GAM,一般化加法モデル: Boostingも彼らの提案
 - ニューラルネットワークの活性化関数をデータから学習可能にする
 - 統計学者らしい機械学習のテキスト(Bayes決定との比較): The Elements of Statistical Learning

データを効果的に取得する 実験計画法と強化学習

実験計画法による平均特性最適化

- 実験計画法 (Fisher) : **情報効率最適化データ収集**
 - 与えられたデータセットから出力を最適化する入力決定
 - 最適決定を最小実験サイズで実現する入力データセット設計
- **逐次実験計画** : **Black Box最適化**
 - ICI:G.E.P Box : 統計的山登りで品質・生産性向上
 - Box and Draper(1969) Evolutionary operation: A statistical method for process improvement, Wiley
 - 全入力セットを事前に決めず
 - 結果を見ながら効果的に山登り

強化学習による予測最適化

- 強化学習
 - 価値関数最適化入力の効率的学習
 - 価値関数最適Input予測
 - どの種のInputを与えればより迅速に最適Inputを予測できるか?
 - **数理的には逐次実験計画の進化系??**
 - AI研究者もかつては意識
Kaelbling et al. (1996)
Reinforcement Learning: A Survey, J. Artificial Intelligence Research, 4, 237-285
 - Kiefer(1959) 最適実験計画 (データ採取の次
の一手)
 - AIでも近年KieferのD-optimalityは利用??

ロバスト・パラメータ設計と敵対的学習

田口のロバストパラメータ設計

- 損失関数のロバスト最適化
 - 制御因子による損失関数最小化
 - 誤差因子：人間が誤差を与える
 - 統計的実験計画法からの決別(1980年代後半)
- 制御因子と誤差因子の直積実験
 - **1953:伊奈製陶タイル実験**
 - 誤差因子に負けない制御因子
 - 負けなさ加減の尺度；1980年代以降SN比
- 1990年代：誤差因子の調合（最悪化）
 - もっとも意地悪な誤差因子の水準組み合わせを技術者が考える

Goodfellow et al.(2014)

Generative adversarial nets, NIPS, 2014.

- 価値関数のロバスト最適化
 - 価値関数最適化問題をゲーム化
 - 価値関数最大化プレイヤー
 - 価値関数最小化プレイヤー
 - 両者が強化学習を行って自身の最適化を図る
- 逐次ロバストパラメータ設計と位置づけ
 - 制御因子の最適水準候補を効率よく攻撃するノイズ因子のWaldの決定関数論的なMin-Max調合

というわけで

『統計数理』は統計的機械学習の中核!
と古典統計家は信じてますが、

本論で現代における統計的機械学習とその中核たる数理の
社会的・理論的可能性を様々な立場で
議論いただければ幸いです

シンポジウム本論を楽しんでchat頂ければ幸いです