計算可能な予測の収束速度

宮部賢志(明治大学)

2025年1月28日(火) 冬のLAシンポジウム2024 RIMS共同研究(公開型) 「計算理論の基礎と新潮流」

自己紹介・概要

自己紹介

- 宮部賢志(みやべけんし)
- アルゴリズム的ランダムネス(algorithmic randomness)
  - 計算論(computability theory, 数理論理学の1分野)および確率論
  - ゲーム論的確率論
  - 機械学習への応用
- 所属学会：日本数学会(数学基礎論)，人工知能学会，Association for Symbolic Logic, Association Computability in Europe

目次

概要

- データからモデルを予測する問題を考察する
- その収束速度が計算論的な複雑さから定まることを示す
    - 確率や位相から定まる限界ではない
- 計算の複雑性による学習限界の理論的基盤の応用可能性を議論する

TuringとSolomonoff

Turing machine

1936年，Alan Turingによる**Turing機械**の提案

- 再帰関数やラムダ計算との等価性 → Church-Turingのテーゼ
- 自然な問題の解決不可能性の議論が可能に
    - 停止問題の計算不可能性
    - Gödel の不完全性定理
    - ヒルベルトの第10問題(ディオファントス方程式の可解性問題)
    - ポストの対応問題
    - Riceの定理
    - P vs NP問題

Solomonoff

1964年，Ray Solomonoff(1926-2009)による**万能帰納推論**の提案

- universal induction, the theory of inductive inference
- 計算概念に基づく確率・予測の提案
- 最適な予測の存在の証明
- 計算近似可能だが，計算可能ではないため，実装は不可能

汎用的予測

筆者による**汎用的予測**(general prediction)の提案

- 計算可能な予測に対して，万能帰納推論のような議論を展開する
- 特に予測の収束速度が議論できるようになる
  - その解析にはアルゴリズム的ランダムネスの理論の概念が使われる
- 汎用的予測の枠組みの拡張，特に計算複雑性理論への応用を議論したい

万能予測と汎用予測

設定

$\{0,1\}^{\mathbb{N}}$: カントール空間

μ: カントール空間上の計算可能測度, 未知

X: μに対してランダムな元, $X_{\leq n}$: Xの先頭nビット

$X_{\leq n}$が与えられたときの，X_{n+1}の真の条件付き確率

$$\mu(k|X_{\leq n}) = \frac{\mu(X_{\leq n}k)}{\mu(X_{\leq n})}$$

予測測度ξによる条件付き確率

$$\xi(k|X_{\leq n}) = \frac{\xi(X_{\leq n}k)}{\xi(X_{\leq n})}$$

最適な予測

ξがνを**優越する**(dominate)とは，

$$\exists c \forall \sigma \nu(\sigma) \leq c\xi(\sigma)$$

計算可能な測度の中ですべてを優越するものは存在しない．

c.e.半測度の中であればすべてを優越するものが存在する．
このような測度を**最適な測度**(optimal measure)と呼ぶ．
証明は万能Turing機械(universal Turing machine)の存在とほぼ同じ．

半測度(semi-measure): 測度から加法性を緩めたもの
c.e.=computably enumerable
c.e.半測度: 下側から計算近似可能な半測度

定理 (Solomonoff 1964)
任意の最適な測度ξは以下を満たす．任意の計算可能測度μに対し，

$$\xi(\cdot|X_{\leq n}) - \mu(\cdot|X_{\leq n}) \to 0, (n \to \infty)$$

が，μ確率1で成立する．

つまり，最適な測度から作られる予測は，任意の計算可能測度の規則を見つけることができる．

欠点: $\xi(\cdot|\cdot)$は計算不可能，収束速度も解析不可能

汎用的予測

十分大きな自然数nに対し，性質$P(n)$が成り立つとは，

$$\exists N \forall n > N . P(n)$$

が成り立つことを言う．

十分汎用的な(sufficiently general)計算可能測度ξに対し，性質Pが成り立つとは，ある計算可能測度νが存在して，νを優越するすべての計算可能測度ξに対し，性質Pが成り立つことを言う．

注意：「優越する」測度は，より多くのモデル測度μの規則を見つけることができる．

主定理

定理 (M.)

任意の計算可能測度μに対し、十分汎用的な測度ξをとると，

$$E_{X\sim\mu}\sum_n D(\mu(\cdot|X_{\leq n})||\xi(\cdot|X_{\leq n})) < \infty$$

となり，かつこの値は 左c.e. Martin-Löfランダム実数

ここで，Dはカルバック・ライブラー情報量．

解釈

重要な点3つ.

- Solomonoffの結果の計算可能測度版になっている
- 誤差の和が有限であることから「収束が速い」ことが分かる
- Martin-Löfランダム性から「収束が遅い」ことが分かる

特に収束の速さが計算論的性質から得られているものであることに注意

今後の課題

- 万能推論ではKolmogorov複雑性との関連も知られている．
  確率論に基づかない複雑性による学習理論として展開する
- 「計算可能な測度の族」に対し，計算量による制約を導入できないか？もしくは代数概念による制約を導入することは可能か？その場合，収束速度はどうなるか？
  - 大まかには，考察する族を小さくすれば，収束も速くなると予想される．
- 連続値に拡張し，他の学習理論と比較する

終わり