Solomonoffの万能推論
・アルゴリズム的確率

宮部賢志　明治大学数学科

2017年9月15日 (金) 人工知能学会 AGI 研究会

はじめに

自己紹介

- 宮部賢志
- 明治大学数学科
- 専門：アルゴリズム的ランダムネス

アルゴリズム的情報理論は，確率論の数学的定式化を目標に出発し，確率・統計・予測などの概念を，計算の側面から理解することを目標としている．
数理論理学（計算論）を背景に持つ研究者が多いが，確率論や力学系や，情報科学を背景に持つ研究者もいる．

アルゴリズム的ランダムネス

初期の主な貢献者

- von Mises 1881-1973
- Andrey Kolmogorov 1903-1987
- Ray Solomonoff 1926-2009

基本は Solomonoff の考え方に従う.
説明の都合上 von Mises や Kolmogorov の言葉も引用する.

小史

1900　国際数学者会議で Hilbert により物理学の公理化（確率論の数学的定式化）が問題として提起される
1919-　von Mises により頻度説に基づく確率論の提案
"First collective, then probability."
1933　Kolmogorov による『確率概念の基礎』
1940s-50s　論議
1966　Martin-Löf によるランダム列の定義の提案
1970s　Solomonoff による複雑性を使った万能推論の提案
2000-　計算論の研究者が興味を持ち始める
2009-　考察空間の一般化

確率とは何か

Kolmogorov の公理的確率論では，
「確率とは，(ルベーグ) 測度である」(伊藤清，確率論の基礎)
確率測度を定めて，その標本のランダム性を調べる．

Solomonoff のアルゴリズム的確率では，
データから出発し，計算の概念を使ってランダム性を定義し，万能推論を得る．確率はそこから導かれる．

Kolmogorov 自身は頻度論者で von Mises に共感していた．
Solomonoff の万能推論はベイジアンなのか？頻度説？傾向説？そういう枠組みでは説明しにくい．計算の概念に基づくという意味で，アルゴリズム的確率と表現している．

アルゴリズム的確率の考え方

設定

2 進無限列 $X \in \{0,1\}^{\mathbb{N}}$ を予測する．正確には，最初の n 桁が与えられて，次の $n+1$ 桁目が 0 か 1 かを予測する．ベイズ予測の設定に従って，$\{0,1\}^{\mathbb{N}}$ 上の初期確率 M を定めれば，$\sigma \in \{0,1\}^n$ が与えられた時，次のビットが $i \in \{0,1\}$ となる確率は，

$$M(i|\sigma) = \frac{M(\sigma i)}{M(\sigma)}$$

で与えられる．
では，M はどう選ぶか？

計算可能性という制約

$M : \{0,1\}^* \to [0,1]$ は計算可能であるものの中から選ぶ.
計算可能=適当なプログラミング言語で表現可能

- 計算可能な関数は十分に表現力が豊か
- 測度全体は非可算だが，計算可能な測度全体は可算

$$\text{データ} \Rightarrow \text{モデル} \Rightarrow \text{予測}$$

ではなく,

$$\text{データ} \Rightarrow \text{予測}$$

モデルとプログラムを同一視している

万能機械の存在

定義 1 (Church-Turing のテーゼ)**.**
$f :\subseteq \mathbb{N} \to \mathbb{N}$ が計算可能 $\iff$ f は Turing 機械で計算可能.

定理 2 (万能機械の存在)**.** 以下を満たす万能機械
$U :\subseteq \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ が存在する：任意の計算可能関数
$f :\subseteq \mathbb{N} \to \mathbb{N}$ に対して，ある $k \in \omega$ が存在し，すべての
$n \in \omega$ について，

$$U(k, n) \simeq f(n)$$

を満たす.

ここで，$\simeq$ は共に定義されていて等しいか，共に定義されていないことを表す.

下側半計算可能な半測度

λ で空列 (長さ 0 の文字列) を表す.

定義 3. 関数 $m : \{0,1\}^* \to [0,1]$ が半測度 (semimeasure) であるとは，$m(\lambda) \leq 1$ で，すべての $\sigma \in \{0,1\}^*$ に対して，

$$m(\sigma) \geq m(\sigma 0) + m(\sigma 1)$$

を満たすことをいう.

定義 4. 実数 $x \in \mathbb{R}$ が下側半計算可能 (lower semicomputable) であるとは，計算可能な単調増加の有理数列 $\{a_n\}$ が存在して，$x = \lim_n a_n$ となることをいう.

下側から計算可能に近似できる数のこと.

万能予測の存在

定理 5. 以下の性質を満たす下側半計算可能な半測度 $M : \{0,1\}^* \to [0,1]$ が存在する：すべての下側半計算可能な半測度 m に対して，ある定数 $c \in \mathbb{N}$ が存在し，すべての $\sigma \in \{0,1\}^*$ に対して，

$$cM(\sigma) \geq m(\sigma)$$

を満たす．

$M(\sigma)$ は X が σ から始まると思う M の信念の度合いを表している．
c を含むのでとても弱い主張のように聞こえる．しかし，私達が真に興味があるのは，$M(i|\sigma) = \frac{M(\sigma i)}{M(\sigma)}$ なので，定数倍の違いは問題にならない．

アルゴリズム的確率

上記の万能予測 M を 1 つ固定し，

$$M(i|\sigma) = \frac{M(\sigma i)}{M(\sigma)}$$

をアルゴリズム的確率 (algorithmic probability) と呼ぶ．
万能予測は複数考えられる．以下では，万能予測 M の選び方に関わらず成立する事実に注目する．

複雑性による表現

良い理論とは

アルゴリズム的確率は,「モデルとプログラムを同一視した上で，モデル選択をしている」と見ることもできることを説明する.

良い理論とは何か？

Epikuros によると,

それらが現象に矛盾しない限りすべて保持せよ

Occam の剃刀によると,

最も単純なものを採用せよ

⇒ 情報量規範

モデルの足し合わせ

モデル=プログラム
モデルに合致する=プログラムのこれまでの出力が正しい
プログラムの長さが短い=可能性が高い
すべての予測を足し合わせる

$$M(\sigma) = \sum\{2^{-|p|} \,:\, U(p) = \sigma *\}$$

定理 6. M は下側半計算可能な半測度で，万能な予測となる．

圧縮可能性と予測可能性の同値性は数学的に示すことができる．哲学的には一体何を意味しているのだろうか？

アルゴリズム的確率の性質

勝利の確率

M から作られる予測は本当に良い予測なのか？
ゲームで n 連勝しているとして，次に勝つ確率は？
$n \to \infty$ で確率は 1 に近づくはずだが，その収束の速さは？
調べるべきは

$$M(1|1^n) = \frac{M(1^{n+1})}{M(1^n)}$$

計算可能な列の予測

Kolmogorov 複雑性

$$K(\sigma) = \min\{\tau \ : \ U(\tau) = \sigma\}$$

定理 7. 計算可能な $X \in \{0,1\}^{\mathbb{N}}$ について，ある定数 C が存在して，

$$2^{-K(n)} \leq 1 - M(X_n | X_{<n}) \leq C \cdot 2^{-K(n)}$$

ほとんどの複雑な n に対しては，$K(n) \geq \log n + \log\log n$ なので，誤差は $\frac{1}{n}$ より速く収束する．(log の底は 2)
単純な n に対しては，$K(n)$ は $\log n$ よりも小さくなるので，誤差は $\frac{1}{n}$ よりかなり大きい．
ただし，$K(n) \to \infty$ なので，$\lim_n M(X_n | X_{<n}) = 1$ である．

注意

すべての万能な予測が持つ性質である．

M は何も情報がないところから，X の規則性を見つけることができる．

収束は単調ではない．このような振動は計算可能性 (または圧縮可能性) から出て来る．

人類最後の日の日付は特別な日付かもしれない？

ランダム列の予測

例えば「偶数番目は1/3で，奇数番目は2/3の確率で1が出る」というモデルを考える．
この標本を見て，この規則を見つけるのは人間には難しい．
この計算可能な確率測度を m とする．
すると，m の意味での確率1で，

$$\lim_n |M(X_n|X_{<n}) - m(X_n|X_{<n})| = 1$$

が成り立つ．

注意

やはり M は事前情報無しで規則を見つけることができる．確率 1 で正しい値に収束していると思うことができる．
アルゴリズム的確率の立場からすると，この確率モデルの確率とは一体何なのか？
データから始めて M で収束する測度とは何かをどう理解すべきか？

課題

万能な予測の計算不可能性

アルゴリズム的確率において中心的な概念である M は計算不可能である．
そのため，そのまま応用することはできない．
アルゴリズム的確率における最大の欠点と言われる．
しかし，アルゴリズム的確率の考え方まで放棄するのはもったいないと思う．

空間の一般化

標準的な考察空間は 2 進無限列の空間である．
自然に有限種類のアルファベットの無限列に拡張できる．
更なる一般化には計算可能解析などの手法が必要である．
設定を一般化することで，他の手法との比較が容易になる．

予測できる確率

M の計算可能性と予測できる確率には深い関係がある.
数学的には Levy の 0-1 法則の計算可能性について考察することになる.
最近では，Lebesgue の微分定理や Radon-Nikodym の定理の計算可能性などの関連研究がある.
計算量の観点から対応する階層を研究することも重要だと思われる.

確率概念について

アルゴリズム的確率の概念は一体何を表しているのか.
数学的にも，哲学的にも，(少なくとも私は) ハッキリ答えることはできない.

終わり

ありがとうございました