アルゴリズム的情報理論

Algorithmic Information Theory

宮部賢志 (明治大学理工学部数学科)

2025年3月11日(火)

自己紹介

- 宮部賢志(みやべけんし)
- 岐阜市出身
- 京都大学理学部・理学研究科（学部・修士・博士・ポスドク）
- 現在：明治大学理工学部数学科
- 専門：計算論・確率論
- 興味：確率の哲学，予測の限界

アルゴリズム的情報理論

- アルゴリズム的情報理論(Algorithmic Information Theory, AIT)
- Algorithmic randomnessやKolmogorov complexityとも呼ばれる
- ランダム性や情報量，予測限界などの概念を計算の概念を用いて定式化し，その性質を調べる

歴史概略

- 1900年：Hilbertの23の問題のうち第6問題『物理学の諸公理の数学的扱い』の中で，確率の公理的基礎を求める
- 1919〜: von Misesによる頻度主義による確率論の提案，collectiveと呼ばれるランダムな列の概念の提案，Villeの構成などにより不十分であることが認識される
- 1970年代：Kolmogorov, Solomonoff, Chaitinらにより，「文字列の複雑性」の概念が提案され，ランダム性について数学的に議論できるようになった
- 2010年前後〜：計算可能解析学(computable analysis)の手法により，通常の数学的対象のランダム性を議論できるようになった

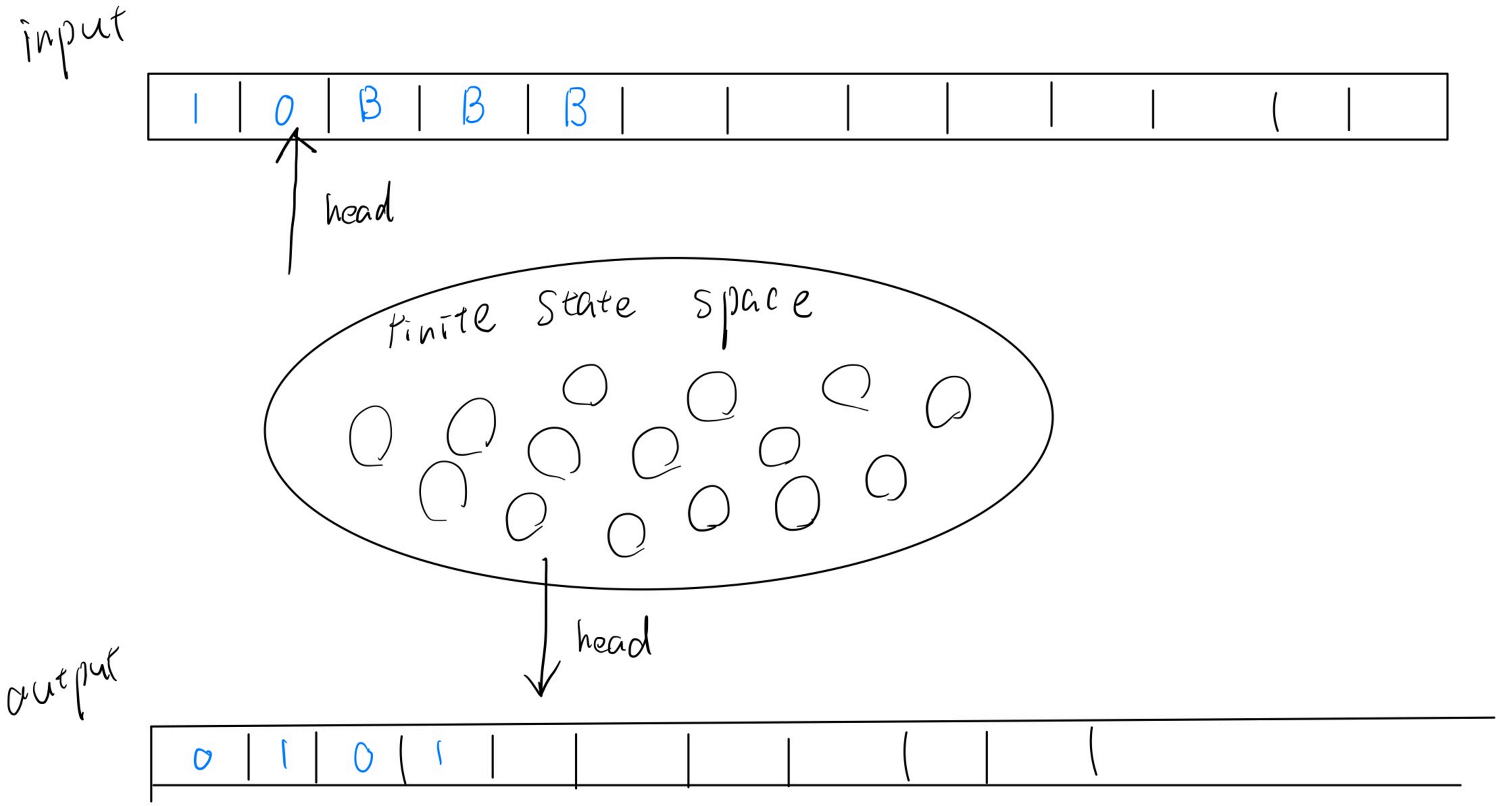
計算論入門

Church-Turingのテーゼ

計算論の基本的な知識の確認

- Hilbertの第2問題「算術の公理間の整合性」や第10問題「ディオファントス方程式の可解性の決定問題」などの提起
- 1930年代に，ラムダ計算，再帰関数，Turing機械などが提案され，その計算可能性が等価であることが示される
- Church-Turingのテーゼ：「計算可能な関数とは，Turing機械により計算できる関数のこと」

Turing機械



計算可能関数

Turing機械の性質

入力と出力は有限アルファベットで，読み書きできる

各ステップは有限状態に従って決定的に定まる

機械の動きは有限情報で記述できる

定理

計算可能関数は可算である

万能機械

プログラムが文法的に正しいかどうかは計算可能に判定可能.
計算が停止しないときには未定義とする．一般には部分関数となる.

定義

$\Phi_e :\subseteq 2^{<\omega} \to 2^{<\omega}$を$e$番目の機械とする.

定理

以下を満たす**万能機械**(universal machine)Uが存在する．任意の機械Mに対し，ある$\sigma \in 2^{<\omega}$が存在して，すべての$\tau \in 2^{<\omega}$で$U(\sigma\tau) \simeq M(\tau)$.

停止問題

定理 (停止問題の計算不可能性)

$H = \{\langle e, k\rangle \mid \Phi_e(k)\downarrow\}$を**停止問題**(Halting problem)と呼ぶ．停止問題Hは計算不可能

系

部分計算可能関数は計算可能に数え上げられるが，全域計算可能関数は計算可能に数え上げることができない．

万能推論入門

万能推論概要

- 主な貢献者はRay Solomonoff(1926-2009)
- 1964年に発表された"A Formal Theory of Inductive Inference"のPart I, IIがよく引用される
- 2000年代にはMarcus HutterらによるAIXIへの拡張もある
- 万能推論(Universal induction)と呼ばれる.
- Solomonoff自身は，確率概念の計算論的な再定義という意味で，Algorithmic Probabilityと呼んでいる.

設定1

カントール空間: 2進無限列$\{0,1\}^{\mathbb{N}}$

μ: カントール空間上の計算可能測度(未知のモデルを表す)

X: μに対してランダムな列

$X \in \{0,1\}^{\mathbb{N}}$が順に与えられるとする．つまり，

$$X = X_1 X_2 X_3 \cdots$$

はμの情報を持っており，

$$X_{\leq n} = X_1 X_2 \cdots X_n$$

として，$X_{\leq n}$からX_{n+1}の分布を予測する問題を考える．

設定2

真の分布は

$$\mu(k \mid X_{\le n}) = \frac{\mu(X_{\le n}k)}{\mu(X_{\le n})}$$

である． 予測もまた測度ξにより，

$$\xi(k \mid X_{\le n}) = \frac{\xi(X_{\le n}k)}{\xi(X_{\le n})}$$

と表される． すなわち，ξはベイズ予測における事前分布で，$\xi(\cdot \mid \cdot)$は事後分布である．

注意

モデル測度μは何らかの積測度になっているとは一般には仮定しない.
つまり，独立同分布性は仮定しない.

そのかわり，モデル測度μは計算論の意味で計算可能であると仮定している.
また，予測測度ξにも弱い計算可能性を仮定する.

ゴール

多くの未知測度μに対して，$\xi(\cdot\mid\cdot)$が$\mu(\cdot\mid\cdot)$が近いような事前分布ξを作る.

未知測度μは様々なタスク(教師あり学習)に対応する． 例えば，分類，回帰，画像認識，音声認識，自然言語処理，金融予測などなど.

予測測度ξはデータから学習するモデルに対応する.

優越

Solomonoffの主定理を述べるのに必要な言葉を準備する.

測度ξが測度νを**優越する**(dominate)とは，ある定数$c \in \mathbb{N}$があってすべての$\sigma \in \{0,1\}^*$で $\nu(\sigma) \le c \cdot \xi(\sigma)$ となることをいう．ここで，$[\sigma] = \{X \in \{0,1\}^{\mathbb{N}} \mid \sigma \prec X\}$とし，$\nu(\sigma)$は$\nu([\sigma])$を意味する.

定理

任意のc.e.半測度を優越するようなc.e.半測度は**最適**(optimal)と呼ばれる．最適なc.e.半測度は存在する.

このことは，万能Turing機械が存在することと同様に示すことができる.

主定理

定理 (Solomonoff 1964)

μを計算可能な測度，ξを最適なc.e.半測度とする．　このとき，μ-確率1で,

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

この予測$\xi(\cdot \mid \cdot)$は万能であることを意味する．

Solomonoffはこの$\xi(\cdot \mid \cdot)$を**アルゴリズム的確率**(algorithmic probability)と呼んでいる．

短所

- $\xi(\cdot \mid \cdot)$は計算可能ではなく，近似可能だがその困難度も高い.
- $\xi(\cdot \mid \cdot)$の$\mu(\cdot \mid \cdot)$への収束は計算可能に抑えることができない.
- "completely useless"と表現する人もある...

注意: No Free Lunch定理とは矛盾しない.

Kolmogorov complexity 入門

概要

- ここまではSolomonoffの結果を最短距離で紹介した.
- 万能な予測の哲学的議論を行う前に，複雑性の概念を紹介して，その関係も議論できるようにしておきたい.
- モデルの複雑さを定量的に表現することで，「単純なモデルほど可能性が高い」と考えるモデルが良いモデルになることを主張する.

モデル選択

モデル選択：観察データに合致する複数のモデルからどのモデルを選ぶか？

2つの大きな異なる考え方がある.

- Epiculusの多説明原理

> If more than one theory is consistent with the observations, keep all theories.
> (Principle of Multiple Explanations)

- オッカムの剃刀(Occam's razor)

  『オッカムのかみそり: 最節約性と統計学の哲学』エリオット・ソーバー(著), 森元 良太 (翻訳), 勁草書房(2021)

Solomonoffの万能推論

アイディア

単純なモデルは可能性が高く，複雑なモデルは可能性が低いとして，確率を足し合わせせよう．

この意味で，Epiculusの多説明原理とオッカムの剃刀のいいとこ取りをする．

ではモデルの複雑性をどう定量化するか？

複雑性

定義 (Kolmogorov complexity)

文字列$\sigma \in 2^{<\omega}$のplain Kolmogorov complexity $C(\sigma)$を，

$$C(\sigma) = \min\{|p| \mid U(p) = \sigma\}$$

と定義する．ただし，Uは万能機械である．

ここで，$|p|$はプログラム$p \in 2^{<\omega}$の長さ． 単純なほど小さく，複雑なほど大きくなる． 例えば，ほとんどのランダムな列$\sigma \in 2^n$に対して，$C(\sigma) \approx n$となる．

定義 (Solomonoff's prior)

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

ここで，Uは万能単調機械(universal monotone machine)で，$U(p) = \sigma*$は出力がσから始まることを意味する.

$M(\sigma)$はσが単純であれば大きくなる.

主定理

定理

上記で定義したMは，最適なc.e.半測度となる．

複雑性から作られた事前分布が最適な予測を与えることが分かる．

確率の哲学

疑問

疑問

アルゴリズム的確率(Algorithmic Probability, AP)は確率の哲学の文脈ではどう評価されるのか？

疑問(詳細)

- APは主観説か？デュエム-クワイン問題を解決しているか？クロムウェルの規則に従っているか？
- APは頻度説や傾向説と解釈できるか？
- APはaleatoryかepistemicか？
- No Free Lunch定理との関係は？(議論済み)
- Nicod's criterionに従うか？(議論済み)
- Pure inductive logicでの公理はどうか？
- 機械学習における更新とベイズ更新との関係は？

主観説と客観説

APは主観説と客観説の両面を併せ持つ.

- 枠組みは主観説であり，観察が少ないときは，確率の値が万能機械の取り方に依存する.
- 一方，観察が多いときは，万能機械の取り方に依らずに，真の分布に近づく.
- すべての計算可能な測度に対して収束するので，「主観的すぎる」という批判は当たらず，クロムウェルの規則に弱い意味で従う.

その他の問題

- 事前分布に主観的な要素は入らないので，デュエム-クワイン問題についても自動的に解決される．
- Grueのパラドックスについても，複雑性の観点から説明できる．
- ただし，考察する空間や言語にどう依存するかについては，議論が進んでいない．
- 観察から学習するという意味でepistemicな確率だと思うが，複雑さを考察する中で構造やパターンを見つけ，aleatory的な要素が加味される．ここについても議論が必要．

参考文献

- Rathmanner, S., & Hutter, M. (2011). A Philosophical Treatise of Universal Induction. Entropy, 13 (6), 1076–1136. https://doi.org/10.3390/e13061076