

因果帰納確率モデルの情報学的考察

Informatics of a stochastic model for causal induction

上浦 基*¹ 高橋 達二*¹

Moto Kamiura Tatsuji Takahashi

*¹東京電機大学理工学部情報システムデザイン学系

Division of Information System Design, School of Science and Engineering, Tokyo Denki University

Here is abstract (max. 150 words) in English. This file is a template file for the XXth annual conference of JSAI for L^AT_EX. This is also a sample of the formatted manuscripts. And this document describes a formatting guideline for (Japanese) manuscripts of the XXth annual conference of JSAI.

1. はじめに

人間は、共起する事象を観察することにより、これらの事象のあいだに因果性を見出すことがある。因果帰納 (causal induction) と呼ばれるこの働きに関して、近年、認知科学では2つの事象の共起頻度に基づく統計モデルによる研究が進んでいる [Hattori & Oaksford 2007]。そこでの基本的な目標は、ベルヌーイ確率変数で記述される2つの事象が与えられたときに人間がこれらの事象のあいだに因果性を見出す強度をノンパラメトリックな統計モデルによって指標化することである。

認知科学とは異なる分野にありながら因果帰納が深く関わる問題として、強化学習の課題である N 本腕バンディット問題がある。これは確率的に報酬 (当たりまたは外れ) が与えられる N 本のレバー (腕) を持つスロットマシンにおいてレバーを引く試行を繰り返すことにより報酬を最大化する問題である。最もよく当たるレバーをできるだけ早く探索する問題としてこれを考えた場合、レバーの施行と当たり外れを結びつける因果帰納の問題であるとも考えられる。この N 本腕バンディット問題でも、人間の認知的バイアスに近い確率モデルがいくつかのヒューリスティクスよりも良い成績を収めており、因果帰納という行為の特性を多面的に考察することの重要性が高まっている [篠原 et al. 2007]。

最近、因果帰納の指標のひとつとして、Takahashi らにより pARIs (proportion of Assumed-to-be Rare Instances) が提案された [Takahashi et al. 2010]。これは単純な形式によって人間の因果帰納の特性をよく反映する指標であるが、人間が因果性を見出す強度と pARIs が関係付けられる理由はよくわかっていない。

本研究の目的は、pARIs における説明変数をベルヌーイ確率変数から一般の連続確率変数へ拡張しその情報学的側面を分析することにより、因果性を見出す強度と pARIs が関係づけられる理由を明らかにすることである。

2. pARIs 変数の連続化

因果帰納に関する定量的研究において、事象の共起情報は 2×2 分割表によって表現される。事象が起こる場合を A , 起こらない場合を \bar{A} のように書くことにし、原因とみなされる事象

連絡先: 上浦 基, 東京電機大学理工学部情報システムデザイン学系, 〒350-0394 埼玉県比企郡鳩山町石坂 156-102, kamiura[at]mail.dendai.ac.jp

表 1: 2×2 分割表

| Cause \ Effect | E | \bar{E} |
|----------------|-----|-----------|
| C | a | b |
| \bar{C} | c | d |

表 2: $n \times m$ 分割表

| Cause \ Effect | Y_1 | \dots | Y_j | \dots | Y_m |
|----------------|----------|---------|----------|---------|----------|
| X_1 | | | a_{1j} | | |
| \vdots | | | \vdots | | |
| X_i | a_{i1} | \dots | a_{ij} | \dots | a_{im} |
| \vdots | | | \vdots | | |
| X_n | | | a_{nj} | | |

を C , 結果とみなされる事象を E と書くことにする。 $(C \cap E)$, $(C \cap \bar{E})$, $(\bar{C} \cap E)$, $(\bar{C} \cap \bar{E})$ なる事象の生起頻度を a, b, c, d とすれば、 2×2 分割表は表 1 のようになる。

この 2×2 分割表が与えられたとき、統計的指標としての pARIs は以下のように定義される。

$$pARIs := \frac{a}{a+b+c} \quad (1)$$

ここに、 $\forall a, b, c \in \mathbb{N}$ に対して $0 \leq pARIs \leq 1$ が成り立つ。 $N := a+b+c+d$ とし、相対頻度 N および $(a+b+c)/N$ を生起確率 $p(C \cap E)$ および $p(C \cup E) = p(C \cap E) + p(C \cap \bar{E}) + p(\bar{C} \cap E)$ で置き換えることにより、因果推論の強度に関する確率モデルとして

$$pARIs = \frac{p(C \cap E)}{p(C \cup E)} \quad (2)$$

を得る。

次に、原因と結果のそれぞれについて「事象が起こる / 起こらない」ではなく、「 i 番でインデックスされた事象が起こる」という形に書き換える。すなわち、 2×2 分割表を拡張し、表 2 のような $n \times m$ 分割表を考える。ただし、 $a_{ij} (\forall i, j \in \mathbb{N} (1 \leq i \leq n, 1 \leq j \leq m))$ は事象 X_i が起こりかつ事象 Y_j が起こった回数である。分割表の拡張に伴い、(2) 式は以下のように書

き換えられる .

$$pARIs(i, j) = \frac{p(X_i \cap Y_j)}{p(X_i \cup Y_j)} \quad (3)$$

(1) 式から (2) 式を導くのと逆には、(3) 式から統計モデルをつくれれば

$$pARIs(i, j) = \frac{a_{ij}}{\sum_{j'=1}^m a_{ij'} + \sum_{i'=1}^n a_{i'j} - a_{ij}} \quad (4)$$

となる .

さらに、(4) 式の $i, j \in \mathbb{N}$ を $x_i, y_j \in \mathbb{R} (-\infty < x_i, y_j < \infty)$ に置き換え形式的に連続化することにより、以下の式を得る .

$$pARIs(x_i, y_j) = \frac{P_3}{P_1 + P_2 - P_3} \quad (5)$$

ただし、 $P_1 := \Pr(x_i - \delta_x \leq X < x_i + \delta_x) = \int_{x_i - \delta_x}^{x_i + \delta_x} \int_{-\infty}^{\infty} p(x, y) dx dy = \int_{x_i - \delta_x}^{x_i + \delta_x} p(x) dx$ 、 $P_2 := \Pr(y_j - \delta_y \leq Y < y_j + \delta_y) = \int_{-\infty}^{\infty} \int_{y_j - \delta_y}^{y_j + \delta_y} p(x, y) dx dy = \int_{y_j - \delta_y}^{y_j + \delta_y} p(y) dy$ 、 $P_3 := \Pr(x_i - \delta_x \leq X < x_i + \delta_x, y_j - \delta_y \leq Y < y_j + \delta_y) = \int_{x_i - \delta_x}^{x_i + \delta_x} \int_{y_j - \delta_y}^{y_j + \delta_y} p(x, y) dx dy$ であり、 $\delta_x, \delta_y > 0$ は x_i, y_j のまわりの微小領域を定めるパラメータである . (5) 式で定義される指標を連続型 pARIs と呼ぶことにする . P_1, P_2, P_3 の積分領域をそれぞれ $S_1, S_2, S_3 \subset \mathbb{R}^2$ とすると $S_3 = (S_1 \cap S_2) \subset (S_1 \cup S_2)$ であり、(1) 式と同様、(5) 式に関しても

$$0 \leq pARIs(x_i, y_j) \leq 1 \quad (6)$$

が一般に成り立つ .

3. 連続型 pARIs の性質

連続型 pARIs がどのような性質を持つ指標なのかを明らかにするため、(5) 式の $p(x, y)$ を 2 次元正規分布に取りその振る舞いを調べる . 相関係数 ρ の 2 次元正規分布

$$p(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \times \left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\} \quad (7)$$

に対して、平均値のまわりで連続型 pARIs を計算する . 微小な $\delta > 0$ を設定し、 $\delta_x = \delta\sigma_x, \delta_y = \delta\sigma_y$ のようにとると $P_1 \approx P_2 \approx 2\delta/\sqrt{2\pi}$ 、 $P_3 \approx 2\delta^2/\pi\sqrt{1-\rho^2}$ となるので、

$$pARIs(\mu_x, \mu_y) \approx \frac{\delta}{\sqrt{2\pi(1-\rho^2)} - \delta} \quad (8)$$

を得る .

ところで、 P_1, P_2, P_3 の定義より取り得る δ の条件を求めると、 $0 \leq P_1, P_2 < 1$ より $0 \leq \delta < \sqrt{\pi/2} \approx 1.25$ 、 $0 \leq P_3 < 1$ より $|\delta| < \sqrt{2\pi} \cdot (1-\rho^2)^{1/4}$ 、 $0 \leq P_1 + P_2 - P_3 < 1$ より $\sqrt{2\pi(1-\rho^2)} - 1 < \delta \leq \sqrt{2\pi(1-\rho^2)}$ を得る . これらより

$$\begin{cases} 0 < \delta < \sqrt{\pi/2} & (|\rho| < 3/4) \\ 0 < \delta \leq \sqrt{2\pi(1-\rho^2)} (< \sqrt{\pi/2}) & (3/4 \leq |\rho| \leq 1) \end{cases} \quad (9)$$

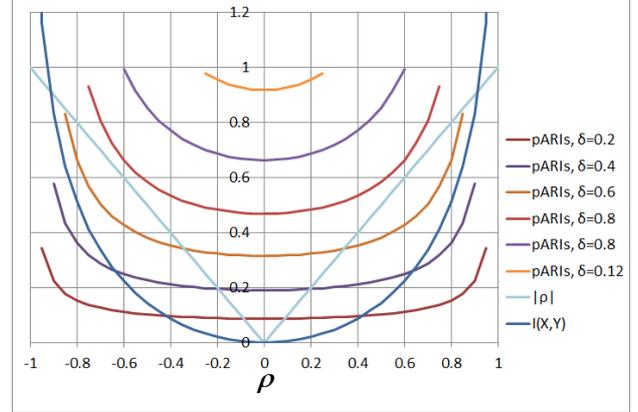


図 1: 相関 ρ に対する $pARIs(\mu_x, \mu_y)$ および相互情報量 $I(X, Y)$ の値

となる . また、(6) すなわち $0 < P_3 \leq P_1 + P_2 - P_3$ 、および $\delta > 0$ かつ $|\rho| \leq 1$ より

$$0 < \delta \leq \sqrt{\pi(1-\rho^2)/2} \leq \sqrt{\pi/2} \quad (10)$$

を得る . (10) 式は (9) 式よりも強い制約であり (10) 式を δ の条件として用いると (8) 式の特異点 $\delta = \sqrt{2\pi(1-\rho^2)}$ はこれによって避けられる .

4. 考察

1 章で述べたように、pARIs は因果帰納の強度に関する指標として導入された . 本章では連続型 pARIs が持つ特性を相関係数および相互情報量との比較において述べる . 図 1 は、相関 ρ の正規確率変数 $X \cong \text{Norm}(\mu_x, \sigma_x^2)$ 、 $Y \cong \text{Norm}(\mu_y, \sigma_y^2)$ が与えられたときの、相関 ρ に対する $pARIs(\mu_x, \mu_y)$ および相互情報量 $I(X, Y)$ の値をグラフ化したものである . ただし、 $pARIs(\mu_x, \mu_y)$ は δ の条件式 (10) によって制約されており、 $pARIs(\mu_x, \mu_y)$ が定義されない ρ の値が存在する .

指標に基づく因果推論が可能となるためには、その指標が (a) 2 つの事象の相関係数が大きくても、因果関係がなければ値が小さくなる、(b) 2 つの事象の相関係数が小さくても、因果関係がなければ値が大きくなる、という性質を持たねばならない . 図 1 は連続型 pARIs に関して性質 (a) を支持する . すなわち、帯状の積分領域 S_1, S_2 の幅 δ を小さくすることにより、広い範囲で相関係数 ρ の変化に対する $pARIs(\mu_x, \mu_y)$ の変化は小さくなる . これに対して、相互情報量 $I(X, Y)$ は、

$$I(X, Y) = -\frac{1}{2} \ln(1-\rho^2) \quad (11)$$

となり、相関係数 ρ に大きく依存する指標である .

他方、性質 (b) に関しては、以下のような点が重要である . 関数的因果性 [上浦 2013] の考え方にに基づき、事象 X, Y の間に因果関係が存在すると言明する代わりに、変数 x, y の間にある関数

$y = f(x)$ が存在すると明言することを考える。 $pARIs(x_i, y_j)$ では $S_1 = [x_i - \delta_x, x_i + \delta_x] \times \mathbb{R}$, $S_2 = \mathbb{R} \times [y_j - \delta_y, y_j + \delta_y]$, $S_3 = [x_i - \delta_x, x_i + \delta_x] \times [y_j - \delta_y, y_j + \delta_y]$ とし積分領域を 2本の帯状に、十字に取る。もし $S_1 - S_3$ の領域に多くの点が分布する場合、 x_i 近傍の点は y_j 近傍以外の点にも写像されていることを示す。すなわち、 P_3 に比して P_1 の値が大きいとき、関数 f は一価関数である度合いが低く、多価関数である度合いが高い。また、もし $S_2 - S_3$ の領域に多くの点が分布する場合、 y_j 近傍に写像される点は x_i 近傍以外にも存在することを示す。すなわち、 P_3 に比して P_2 の値が大きいとき、関数 f は単射である度合いが低い。結局、 P_1, P_2 が P_3 に比して小さくなることで、 $pARIs(x_i, y_j)$ の値を大きくする (1 に近づける) ことになり、それは関数 f の (x_i, y_j) 近傍における一価関数性と単射性が高くなることを意味している。以上の考え方を定量的な検証によって示すのは今後の課題である。

さらに、 $pARIs(x_i, y_j)$ を機械学習における指標として実用的に用いようとした場合には、 (x_i, y_j) 近傍を、データ獲得に合わせてどのように更新してゆくのがよいか考えていく必要がある。

参考文献

- [Hattori & Oaksford 2007] Hattori, M. & Oaksford, M.: Adaptive Non-Interventional Heuristics for Covariation Detection in Causal Induction: Model Comparison and Rational Analysis, *Cognitive Science*, Vol.31, pp.765-814 (2007)
- [篠原 et al. 2007] 篠原, 田口, 桂田, 新田: 因果性に基づく信念形成モデルと N 本腕バンディット問題への適用, *人工知能学会論文誌*, Vol.22, No.1, pp.58-68 (2007)
- [Takahashi et al. 2010] Takahashi, T., Kohno Y., and Oyo, K.: Causal Induction Heuristics as Proportion of Assumed-to-be Rare Instances (pARIs), In: *Proceedings of the 7th International Conference on Cognitive Science (ICCS2010)*, pp.361-362 (2010)
- [上浦 2013] 上浦 基: Implicit 相互作用: システムの局所的記述の数理, *計測自動制御学会論文集*, Vol.49, No.1, pp.190-196 (2013)