

記述性と合理性は両立できるか？無向グラフィカルモデルに基づく因果帰納への構造的アプローチ

Can descriptiveness and rationality coexist? A structural approach to causal induction based on undirected graphical models

樋口 滉規[†], 市野 弘人[‡], 高橋 達二[‡]

Kohki Higuchi, Hiroto Ichino, Tatsuji Takahashi

[†] 中部大学, [‡] 東京電機大学

Chubu University, Tokyo Denki University

kohki.higuchi@gmail.com

概要

本研究では、無向グラフの枠組みに対して「因果の対称性」や「事象の稀少性」といった認知バイアスを組み込むことで、人間の因果帰納推論を記述するモデルを導出した。先行研究のデータを用いたメタ分析の結果は、提案モデルが高い記述性能を有することを示し、「因果の対称性」に焦点をあてた認知実験の結果は、人間の因果判断が無向モデル型と有向モデル型に二分されることを示した。この結果は、人間の因果推論の認知モデリングにおいて「規範モデル」と「記述モデル」が相互排他的な仮説ではなく、両者が共存している可能性を強く示唆している。
キーワード：因果帰納, 帰納推論, 二重過程理論, 認知モデル

1. 背景と目的

因果関係に関する認知機能は、人類が不確実性に満ちた環境に適応し、今日の繁栄に至る上で中心的な役割を果たしてきたと考えられる。後述する単純因果帰納という枠組みを用いて、これまで多くの心理学者や認知科学者が因果関係を検出する人間の認知システムのモデリングを試みてきた。それらは、(1)「数学的規範との整合性」と「認知バイアスとの整合性」のどちらを重視するか、(2)「因果関係の強度」と「因果関係の構造」のどちらに着目するか、という二つの軸で分類できる。しかし、「記述&構造モデル」に位置するモデルはこれまでに提案されておらず、この不在により網羅的な比較が困難であった。この空白を埋めるため、我々は無向グラフィカルモデルに基づく新たなモデル Undirected Causal Support (UCS) を提案し、メタ分析と認知実験の実施に基づいてその妥当性を示す。

2. 因果帰納

2.1 単純因果帰納

単純因果帰納は、単一の原因候補事象 (C) が単一の結果事象 (E) を引き起こす度合いを、それらの共起情報に基づいて評価する枠組みである。このとき共起情報は、表 1 に示す 2×2 分割表で表現される。

表 1 2×2 分割表。 a, b, c, d は原因候補事象 C と結果事象 E の各組み合わせが生起した回数を指す。

	Effect (E)	No Effect ($\neg E$)
Cause (C)	a	b
No Cause ($\neg C$)	c	d

2.2 規範的因果モデル

単純因果帰納の代表的なモデルには、例えば ΔP [6]、*PowerPC* [1] などがあるが、これらのモデルは規範性を持つものの、人間の推論を十分に説明できないことが分かっている [3, 4]。

Griffiths & Tenenbaum (2005) は単純因果帰納の問題を「因果構造」の文脈から説明するモデル、*Causal Support* (CS) を提案した。CS は、図 1 に示すような背景原因のみが結果事象に影響を及ぼす構造 (G_0) と背景原因と原因候補事象が結果事象に影響する構造 (G_1) の対数尤度比として定式化される。

$$CS = \log \frac{P(D|G_1)}{P(D|G_0)} \tag{1}$$

また、彼らは CS によって、 ΔP や *Power PC* では説明することのできない人間の推論傾向を説明できることを示した [3]。

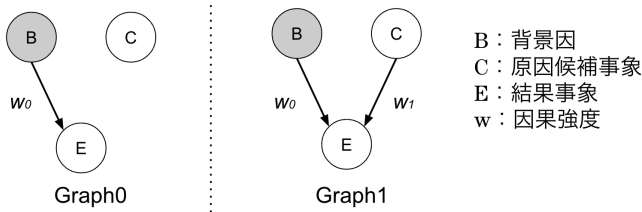


図1 有向グラフにおいて、 C と E が独立である構造 (G'_0) および、従属である構造 (G'_1)

2.3 記述的因果モデル

人間の因果帰納が因果の規範的定義に則ったものである必然性はなく、実際には共変動の推定を行なっているとする立場の研究も存在する。そのようなモデルには Dual-Factor Heuristics (DFH) [4] や proportion of Assumed-to-be Rare Instances (pARIs) [9] がある。これらは共に原因と結果の双方向性と稀少性を仮定したモデルであり、人間の推論と高度に適合することが実証されている [5]。ここで因果の双方向性とは、任意の変数 X, Y に関して、「 X が Y に与える影響」と「 Y が X に与える影響」を実質的に区別せず、その両方が大きい場合に強い因果性があると判断する推論の性質を指す。また、稀少性仮定とは人間が因果推論に際して着目する事象の生起確率は一般に小さいとする仮定であり、人間が事象の稀少性を仮定していることは、様々な認知課題において実証されている (e.g. [7])。

3. 無向グラフィカルモデル

無向グラフィカルモデルと有向グラフィカルモデルはともに、確率変数間の条件付き独立性をグラフを用いて表現する枠組みである。例えば、あるグラフにおいて変数 A から変数 B へのすべての経路が変数 C を経由する場合、そのグラフは A と B が C を所与として独立 ($A \perp B | C$) であることを意味する。ある変数集合が完全グラフであるときその変数集合をクリークと呼び、特に、一つでも変数を増やすとクリークでなくなるときには極大クリークと呼ぶ。無向グラフィカルモデルにおいて変数の同時分布は、ポテンシャル関数と呼ばれる、極大クリーク上の任意の関数 ψ の積として表現される。

4. 提案モデルの導出

CS と同様に図2に示す C と E が独立であるグラフ (G'_0) と C と E が従属であるグラフ (G'_1) の二つを想定する。なお、我々の提案するモデルにおいては C の生起確率がモデルの値に影響するため、その

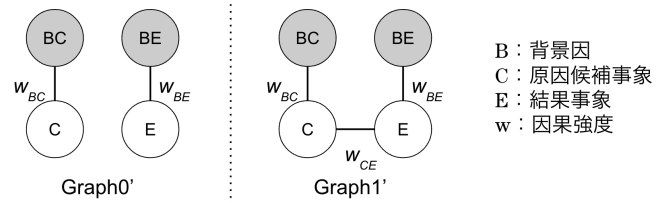


図2 無向グラフにおいて、 C と E が独立である構造 (G'_0) および、従属である構造 (G'_1)

背景因である BC を明示的に考慮する必要があるのに対し、CS において C の生起確率はモデルの値に影響を与えないため省略されている。

我々は CS と同様に G'_1 と G'_0 の対数尤度比を Undirected Causal Support (UCS) として定式化する。

$$\text{UCS} = \log \frac{P(D|G'_1)}{P(D|G'_0)} \quad (2)$$

なお、各グラフの尤度は以下のように計算できる。

$$\begin{aligned} P(D|G'_0) &= \int \int P(D|w_{BC}, w_{BE}, G'_0) \\ &\quad P(w_{BC}, w_{BE}|G'_0) dw_{BC} dw_{BE} \end{aligned} \quad (3)$$

$$\begin{aligned} P(D|G'_1) &= \int \int \int P(D|w_{BC}, w_{BE}, w_{CE}, G'_1) \\ &\quad P(w_{BC}, w_{BE}, w_{CE}|G'_1) dw_{BC} dw_{BE} dw_{CE} \end{aligned} \quad (4)$$

ただし、

$$\begin{aligned} P(D|w_{BC}, w_{BE}, G'_0) &= \prod_{e,c} \frac{1}{Z} (\psi(bc, c) \psi(be, e))^{N(e,c)} \end{aligned} \quad (5)$$

$$\begin{aligned} P(D|w_{BC}, w_{BE}, w_{CE}, G'_1) &= \prod_{e,c} \frac{1}{Z} (\psi(bc, c) \psi(be, e) \psi(c, e))^{N(e,c)} \end{aligned} \quad (6)$$

ここで、 Z は規格化定数であり、ポテンシャル関数 ψ は極大クリーク上の任意の関数である。我々は ψ を、極大クリークに含まれる各変数について背景因とその他の変数を所与とする条件付き確率を求めたうえで、それらの幾何平均を取ったものとして定義した。

$$\psi(b, c) = \sqrt{P(b|c; w_{BC}) P(c|b; w_{BC})} \quad (7)$$

$$\psi(b, e) = \sqrt{P(b|e; w_{BE}) P(e|b; w_{BE})} \quad (8)$$

$$\psi(c, e) = \sqrt{P(c|b, e; w_{BE}, w_{CE}) P(e|b, c; w_{BE}, w_{CE})} \quad (9)$$

実質的に $P(C)$, $P(E)$ を意味する w_{BC} , w_{BE} がとりうる値の範囲を制限することで、各グラフにおける稀少性を表現した。 w_{BC} , w_{BE} , w_{CE} はそれぞれ独立に $\text{Unif}(0, x)$, $\text{Unif}(0, x)$, $\text{Unif}(0, 1)$ で表される一様分布に従い、これに対応する確率密度関数を $f_{w_{BC}}$, $f_{w_{BE}}$, $f_{w_{CE}}$ とする。このとき、同時確率密度関数は以下のように表される。

$$P(w_{BC}, w_{BE} | G'_0) = f_{w_{BC}}(w_{BC}) \cdot f_{w_{BE}}(w_{BE}) \tag{10}$$

$$\begin{aligned} P(w_{BC}, w_{BE}, w_{CE} | G'_1) \\ = f_{w_{BC}}(w_{BC}) \cdot f_{w_{BE}}(w_{BE}) \cdot f_{w_{CE}}(w_{CE}) \end{aligned} \tag{11}$$

人間が事象の生起確率について抱いている事前信念はドメインごとに異なる可能性が高いため、また、人間の因果推論が稀少性の仮定に基づくことを確認するために、今回は対象となるデータと最も適合する x を選択した。一方、 w_{CE} は人間の因果強度に関する事前信念にしたがうものと解釈できるが、ここでは無情報事前分布である $[0, 1]$ の一様分布を仮定した。

4.1 強度モデルとの関連

極端な稀少性を仮定し、 w_{BC} , w_{BE} を 0 に近づける極限操作を行なった場合に、式 6 で表現される尤度関数を最大化する w_{CE} は式 12 と一致する。

$$w_{CE}^{\hat{}} = \frac{2}{\frac{1}{P(E|C)} \frac{1}{P(C|E)}} = \frac{2a}{2a + b + c} \tag{12}$$

これは $P(E|C)$ と $P(C|E)$ の調和平均であり、集合間の類似性の指標である Dice 係数に等しい。ここで、pARIs と DFH はともに、それぞれ Jaccard 係数、落合係数として知られる集合間の類似性の指標であるが、これは興味深い一致点である。このことは、因果構造を考慮した UCS は、そのパラメータの最尤推定量の近似として、従来の強度に関する記述モデルを位置づけることができることを意味する。

5. メタ分析

我々は、Hattori & Oaksford (2007) と Perales & Shanks (2007) によるメタ分析用実験データを用いて、提案モデルを評価した [4, 8]。なお、前者には 5 つの独立した研究の計 8 個の実験結果が、後者には 9 つの独立した研究の計 19 個の実験結果が含まれた。比較モデルは、 ΔP , Power PC, pARIs, DFH, CS, UCS とし、実験 $\times$ モデルごとに、人間の判断の平均値とモデル値の相関係数を算出した。そのうえで、同一データセットに含まれる実験と対応する全ての相関係数

が「真の相関係数」の推定値であると仮定する固定効果モデルを用いて、真の相関係数の点推定値を算出した。なお、全てのモデルの値は Griffiths & Tenenbaum (2005) が用いた変換 $f(x) = \text{sign}(x) \cdot \text{abs}(x)^\gamma$ によって正規化を行った。ここで、 x は人間の判断の平均値、 γ はスケーリング・パラメータであり、 γ は変換された予測値とデータ間の線形相関を最大化するものが選択された。

5.1 結果と考察

メタ分析の結果を表 2 に示す。UCS の稀少性のパラメータである x は、各実験データに対して $[0, 1]$ の範囲を 0.1 刻みで探索した結果、相関係数が最大となる $x = 0.3$ を採用した。得られた結果は、(1) 稀少性を仮定した UCS は、記述モデルと同程度に人間の平均推定値と高度に適合（各データセットにおいて $\hat{\rho} = 960$, $\hat{\rho} = 947$ ）すること、(2) 人間が因果強度／構造の推定のどちらを行っているという立場に立った場合にも、因果の方向を考慮しない推論を行っていることを示唆するものである。

表 2 Hattori & Oaksford (2007) のデータを用いたメタ分析結果（左）と Perales & Shanks (2007) のデータを用いたメタ分析結果（右）。ここで、 $\hat{\rho}$ はモデル値と人間の回答値の相関係数の推定値、 ρ_L , ρ_U はその下側信頼限界と上側信頼限界である。

	$\hat{\rho}$	ρ_L	ρ_U		$\hat{\rho}$	ρ_L	ρ_U
Dice	.969	.944	.983	pARIs	.948	.921	.966
DFH	.969	.944	.983	Dice	.947	.920	.965
pARIs	.968	.943	.983	UCS	.942	.913	.962
UCS	.960	.927	.978	DFH	.858	.789	.905
CS	.883	.796	.935	ΔP	.842	.768	.894
ΔP	.729	.554	.842	CS	.841	.766	.893

6. 実験

メタ分析の結果は人間の推論における平均的な傾向を明らかにしたものの、異なる実験参加者が、因果の対称／非対称性に関して、異なる立場からの推論を行っている可能性を検証するための実験を実施した。表 1 における a , d を固定し、 b , c の比率を制御することで、有向モデル (CS, ΔP 等) の値は変化するが無向モデル (UCS, pARIs 等) の値は変化しないような刺激を設計した (表 3)。クラウドソーシングサービスで募集した 200 人を対象に、Web アプリケーションを用いた実験を行い、Instructional Manipulation Check (IMC) を通過した 132 人を分析の対象とした。得ら

表 3 刺激設計とモデルの予測値。表 1 に示した分割表を 6 通りに変化させた。

#	刺激設計					モデルの予測値			
	a	b	c	d	合計	ΔP	CS	UCS	Dice
1	6	6	0	6	18	.500	1.55	21.8	.667
2	6	3	3	6	18	.333	0.64	21.7	.667
3	6	0	6	6	18	.500	2.17	21.8	.667
4	7	6	0	7	20	.538	2.25	26.0	.700
5	7	3	3	7	20	.400	1.18	25.9	.700
6	7	0	6	7	20	.538	2.82	26.0	.700

れた結果に対してクラスタリングを行い、全体及び各クラスタの平均回答値についてモデル値との相関係数を算出した。

6.1 結果と考察

得られたデータに対してシルエット法を適用した結果、最適なクラスタ数が 2 であることを特定した。その後、k-means 法によるクラスタリングを行った結果、計 132 名の参加者は、65 人からなるクラスタ 1 と 67 人からなるクラスタ 2 に二分された。全参加者および各クラスタの平均回答値を表 4 に示し、モデルと参加者の回答値の相関係数を表 5、表 6 に示す。ここで、値に変化がないクラスタ 2 の回答値と

表 4 各クラスタの平均回答値

#	参加者の回答平均		
	All	Cluster 1	Cluster 2
1	51.5	53.9	48.9
2	57.1	63.6	50.4
3	70.0	91.3	48.2
4	55.5	57.5	53.4
5	63.4	69.7	57.0
6	72.2	88.2	55.7

表 5 各クラスタの平均回答値と、因果の方向性を区別するモデルによる予測値の相関係数。

対象の参加者	$P(E C)$	$P(C E)$	ΔP	CS
All	0.972	-0.891	0.213	0.536
Cluster1	0.995	-0.911	0.221	0.520
Cluster2	0.090	-0.087	0.009	0.177

変化があるクラスタ 1 の回答値の平均は、必然的に後者の値と単調関係にあるため、今回の刺激設計において、参加者全体に対する相関係数を評価することは殆

表 6 各クラスタの平均回答値と、因果の方向性を区別しないモデルによる予測値の相関係数。

対象の参加者	UCS	pARIs	DFH	Dice
All	0.276	0.273	0.277	0.273
Cluster1	0.082	0.077	0.209	0.077
Cluster2	0.921	0.929	0.359	0.929

ど無意味であることに注意されたい。結果は、参加者が $P(E|C)$ と高く適合 ($r = .995$) するクラスタ 1 と UCS, pARIs, Dice と高く適合 ($r \approx .92$) するクラスタ 2 に明確に二分されることを示している。規範的なモデルである ΔP , CS の適合は各クラスタのいずれにおいても顕著に高くはないものの、CS はクラスタ 1 に対して中程度の適合 ($r = .520$) を示している。これらの結果は、異なる参加者が、因果の方向性に関して異なる立場からの推論を行った可能性を示唆するとともに、人間の推論や意思決定が反射的なヒューリスティック処理である「システム 1」と熟慮的かつ分析的な処理である「システム 2」によって行われるとする二重過程理論の可能性を、単純因果帰納の文脈において示唆している [2]。

文献

[1] Patricia W. Cheng. From Covariation to Causation: A Causal Power Theory. *Psychological Review*, 104(2):367–405, 1997.

[2] Jonathan Evans. Dual-processing accounts of reasoning, judgment, and social cognition. *Annual review of psychology*, 59:255–78, 2008.

[3] Thomas L. Griffiths and Joshua B. Tenenbaum. Structure and strength in causal induction. *Cognitive Psychology*, 51(4):334–384, 2005.

[4] Masasi Hattori and Mike Oaksford. Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. *Cognitive Science*, 31(5):765–814, 2007.

[5] Kohki Higuchi, Kuratomo Oyo, and Tatsuji Takahashi. Causal intuition in the indefinite world: Meta-analysis and simulations. *Biosystems*, 225:104842, 2023.

[6] H. M. Jenkins and W. C. Ward. Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied*, 79(1):1–17, 1965.

[7] Mike Oaksford and Nick Chater. A Rational Analysis of the Selection Task as Optimal Data Selection. *Psychological Review*, 101(4):608–631, 1994.

[8] José C. Perales and David R. Shanks. Models of covariation-based causal judgment: A review and synthesis. *Psychonomic Bulletin and Review*, 14(4):577–596, 2007.

[9] Tatsuji Takahashi, Yu Kohno, and Kuratomo Oyo. Causal induction heuristics as proportion of Assumed-to-be Rare Instances (pARIs). In *Proceedings of the 7th International Conference on Cognitive Science (ICCS2010)*, pages 361–362, 2010.