マーケティング・リサーチにおける
統計的因果探索を用いた因果仮説構築に関する研究

データサイエンス研究科, 株式会社マクロミル
小西 伶児

2020 年 11 月 26 日

概要

本研究では,

目次

1　序論

1.1　はじめに

2　モデルと識別可能性

本章ではまず、本論文で用いる数学記号を導入し、非巡回有向グラフ (Directed Acyclic Graph, DAG) モデルを定義する。その後、既存研究として、2 次分散関数 (Quadratic Variance Function, QVF)DAG モデル [2] と、混合因果モデル (Mixed Causal Model)[4] について概説し、本論文で提案するモデルの詳細について述べる。

2.1　数学的準備

グラフは頂点 (node) の集合 $V = \{1, 2, \ldots, p\}$ と、頂点同士をつなぐ辺 (edge) の集合 $E \subset V \times V$ によって、$G = (V, E)$ と表現される。グラフの辺は有向辺 (矢線) と無向辺 (双方向矢線) に分けることができ、2 つの頂点 $j, k \in V$ において、$(j, k) \in E$ かつ $(k, j) \notin E$ のとき、j から k への矢線があるという。これを $j \to k$ と表現することもある。一方で、$(j, k) \in E$ かつ $(k, j) \in E$ のとき、j と k の間に双方向矢線があるという。すべての辺が有向辺であるグラフを有向グラフ (directed graph) という。本論文では、特に断りのない限り、頂点 j から k への矢線がある場合、j が k の原因であるといった因果関係があることを表すとする。つまり、本論文で扱うグラフにおける矢線の有無は因果関係の有無を表しており、矢線の始点が原因で、矢線の終点が結果である。このような定性的な因果関係を表すグラフを因果グラフ (causal graph) という。また、グラフ G からすべての矢印を取り除くことによって得られるグラフを G のスケルトンという。

頂点の系列 $\alpha_1, \alpha_2, \ldots, \alpha_{n+1}$ について、すべての $i = 1, 2, \ldots, n$ で、$\alpha_i \to \alpha_{i+1}$、または $\alpha_{i+1} \to \alpha_i$ となる矢線がある時、長さ n の道 (path) という。特に、すべての $i = 1, 2, \ldots, n$ で、$\alpha_i \to \alpha_{i+1}$ となる矢線がある時、長さ n の有向道 (directed path) という。また、長さ n の有向道で、$\alpha_1 = \alpha_{n+1}$ となるものを巡回閉路 (cycle) という。一方で、巡回閉路のない有向グラフは非巡回的 (acyclic) であるという。本論文では、非巡回有向グラフ (Directed Acyclic Graph; DAG) のみを扱う。

頂点 j から k への矢線がある時、j を k の親 (parent) といい、k を j の子 (child) という。また、$(j, k) \in E$ であるすべての頂点 j からなる集合を $Pa(k)$ と表記する。頂点 j から k への有向道がある時、j を k の祖先 (ancestor)、k を j の子孫 (descendant) という。頂点 k のすべての祖先からなる集合を $An(k)$、すべての子孫からなる集合を $De(k)$ と表記する。また、すべての頂点から k と k の子孫を除いたものを、k の非子孫 (non-descendant) といい、その集合を $Nd(k) \equiv V \backslash (\{k\} \cup De(k))$ と表記する。さらに、因果的順序 (causal oredering) について定義する。因果的順序とは、その順序に従って変数を並び替えると、すべての矢線 $(j, k) \in E$) について、k が j の原因になることがない順序のことであり、$\pi = (\pi_1, \ldots, \pi_p)$ と表記する。DAG で表現される因果グラフには、このような順序が (一意とは限らないが) 存在するという特徴がある。つまり、因果グラフを同定することは、因果的順序を同定することとスケルトンを同定することという 2 つの工程に分解することができる。

有向グラフ G における頂点上の標本空間 $\mathcal{X}_V$ の確率分布に従う確率変数の集合 $X \equiv (X_j)_{j \in V}$ について考える。ここで、確率変数ベクトル X は、同時確率密度関数 $f_G(X) = f_G(X_1, X_2, \ldots, X_p)$ で与えられていると仮定する。V の任意の部分集合 S について、$X_S \equiv \{X_j : j \in S \subset V\}$ と $\mathcal{X}_S \equiv \times_{j \in S} \mathcal{X}_j$ を定義する。ただし、$\mathcal{X}_j$ は X_j の確率空間である。また、任意の頂点 $j \in V$ について、確率変数ベクトル X_S を与えたときの変数 X_j の条件付き確率を $f_j(X_j | X_S)$ と表記する。すると、DAG G によるモデルは以下のように因数分解することができる [3]。

$$f_G(X) = f_G(X_1, X_2, \ldots, X_p) = \prod_{j=1}^{p} f_j(X_j|X_{Pa(j)}) \tag{1}$$

ここで、$f_j(X_j|X_{Pa(j)})$ は、X_j の親変数 $X_{Pa(j)} \equiv \{X_k : k \in Pa(j) \subset V\}$ を与えた条件付き確率である。

また、本論文では観察データから因果グラフを同定するという問題を扱うため、因果グラフの識別可能性について定義する。識別可能性を直感的に説明すると、条件付き確率分布 $f_j(X_j|X_{Pa(j)})$ に対してある仮定を置くと、同時確率密度関数 $f_G(X)$ を与えた DAG G の構造を一意に決定付けることができるということである。

識別可能性について詳細に定義するために、すべての $j \in V$ に関する条件付き確率分布 $f_j(X_j|X_{Pa(j)})$ の集合を $\mathcal{P}$ と表記する。また、グラフ $G = (V, E)$ について、グラフ G に関する同時分布のクラスと、分布 $\mathcal{P}$ のクラスを以下で定義する。

$$\mathcal{F}(G; \mathcal{P}) \equiv \{f_G(X) = \prod_{j \in V} f_j(X_j|X_{Pa(j)}); \text{where} f_j(X_j|X_{Pa(j)}) \in \mathcal{P} \quad \forall j \in V\} \tag{2}$$

続いて、p 個の変数からなる非巡回的有向グラフの集合を $\mathcal{G}_p$ と表記する。そこで、DAG $\mathcal{G}_p$ の空間上の確率分布のクラス $\mathcal{P}$ における識別可能性を以下のように定義する。

定義 2.1 (識別可能性)**.** 条件付き分布のクラス $\mathcal{P}$ が $\mathcal{G}_p$ において識別可能であるとは、$G, G' \in \mathcal{G}_p$ において $G \neq G'$ であるならば、$f_G = f_{G'}$ を満たすような、$f_G \in \mathcal{F}(G; \mathcal{P})$ と $f'_G \in \mathcal{F}(G'; \mathcal{P})$ が存在しないことである。

2.2 2 次分散関数 (QVF) DAG モデル

本節では、Park and Raskutti(2017)[2] によって提案された 2 次分散関数 (QVF) DAG モデルについて概説する。

QVF-DAG モデルは、各頂点の親による条件付き分布 $\mathcal{P}$ の分散が、平均の 2 次式で与えられているというモデルであり、以下のように定義される。

定義 2.2 (QVF-DAG モデル [2])**.** 2 次分散関数 (Quadratic variance function, QVF)DAG モデルは、各頂点の親による条件付き分布の分散が、平均の 2 次式で表現できる DAG モデルである。つまり、すべての $j \in V$ について、以下を満たすような $\beta_{j0}, \beta_{j1} \in \mathbb{R}$ が存在する。

$$Var(X_j|X_{Pa(j)}) = \beta_{j0} E(X_j|X_{Pa(j)}) + \beta_{j1} E(X_j|X_{Pa(j)})^2 \tag{3}$$

2 次分散関数による確率分布は、Morris(1982)[1] によって正準型指数分布族の文脈において導入され、ポアソン分布、二項分布、負の二項分布、ガンマ分布が含まれる。DAG モデルの文脈においては、各頂点の分布がその頂点の親集合からの影響を受けていると考えるため、各頂点の親による条件付き分布は以下のように記述することができる。

$$P(X_j|X_{Pa(j)}) = \exp\left(\theta_{jj} X_j + \sum_{(k,j) \in E} \theta_{jk} X_k X_j - B_j(X_j) - A_j\left(\theta_{jj} + \sum_{(k,j) \in E} \theta_{jk} X_k\right)\right) \tag{4}$$

ここで、$A_j(\cdot)$ は対数分配関数 (log-partition function)、$B_j(\cdot)$ は指数分布族によって決まる関数、$\theta_{jk} \in \mathbb{R}$

は頂点 j に対応するパラメータである。DAG モデルの因数分解 (1) 式により、QVF-DAG モデルの同時確率分布は、以下のように記述することができる。

$$P(X) = \exp\left(\sum_{j\in V}\theta_{jj}X_j + \sum_{(k,j)\in E}\theta_{jk}X_kX_j - \sum_{j\in V}B_j(X_j) - \sum_{j\in V}A_j\left(\theta_{jj} + \sum_{(k,j)\in E}\theta_{jk}X_k\right)\right) \quad (5)$$

参考文献

[1] Carl N Morris. Natural exponential families with quadratic variance functions. *Ann. Stat.*, Vol. 10, No. 1, pp. 65–80, 1982.

[2] Gunwoong Park and Garvesh Raskutti. Learning quadratic variance function (QVF) DAG models via overdispersion scoring (ODS). *J. Mach. Learn. Res.*, Vol. 18, No. 1, pp. 8300–8342, January 2017.

[3] Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, USA, 2nd edition, 2009.

[4] Wei Wenjuan, Feng Lu, and Liu Chunchen. Mixed causal structure discovery with application to prescriptive pricing. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pp. 5126–5134. ijcai.org, 2018.

謝辞

ありがとうございました.