

# 時系列データに対する予測モデルの介入効果の推定

Estimation of interventinal effect on prediction models for time series data

切通 恵介 \*1

Keisuke Kiritoshi

紅林 亘 \*2

Wataru Kurebayashi

泉谷 知範 \*1

Tomonori Izumitani

小山 和輝 \*1

Kazuki Koyama

木村 大地 \*1

Daichi Kimura

大川内 智海 \*1

Tomomi Okawachi

清水 昌平 \*2

Shohei Shimizu

\*1 NTT コミュニケーションズ株式会社

NTT Communications corporation

\*2 滋賀大学

Shiga University

For prediction models of machine learning including operation variables, there are applications to specify operation variables with high contribution from the weights and partial derivatives of models and to connect them to actual operations. On the other hand, since prediction models do not consider the causal relationship behind the data, the intervention effect on the model may differ from the actual system even if the prediction accuracy is high. Blöbaum et al. propose the framework which obtains the intervention value which approaches the prediction value of the prediction model to the desired value through the causal graph by combining the regression model and causal graph. In this paper, we propose a method to obtain the intervention effect for the time series model and the time series causal graph using this framework. We evaluate the error between the predicted value and the desired value by the intervention obtained only by the regression model and the proposed method for the time series data by experiments of artificial data and simulation data.

## 1. はじめに

機械学習モデルの説明を得る手法は、その実応用におけるモデルの信頼性の調査やモデルの予測結果を元にした操作を行うための意思決定の観点で重要性が高く、近年多くの研究が提案されている [Adadi 18]。機械学習の予測モデルにおいて、説明としてモデルの各特徴の重みや偏微分値などを貢献度としてモデルや入力サンプルの情報のみから定義することが多い。

変数間の関係性、特に因果関係を持つシステムやモデルに対して、特徴のある変数のある値を固定することを介入といい、介入が他の変数にどれくらい影響を与えるかをその変数間の介入効果という。介入効果を調べることで、例えばある操作変数が他の観測値に対してどのくらい影響を及ぼすかを表すことで、目標値に観測値を近づけるためにどの程度の操作を行えば良いかを判断する指標となる。

一方で、機械学習モデル、特に線形回帰やニューラルネットワークといった入力変数と出力変数の関係のみから関数を学習するモデルは予測対象の裏にある因果関係やデータの生成プロセスを考慮していない、よってモデルの予測値の精度が高い場合でも、重みなどのモデル単体から得られる説明は、データの生成プロセスから得られる実際の介入効果と一致しない可能性が高い。したがって、予測モデル単体から得られる説明を操作などの意思決定に用いることはリスクが高いといえる。

Blöbaum らは、因果グラフと予測モデルで表される変数間の因果方向をもとに介入効果を伝播させ、予測モデルの予測値が目標値に至る時の説明変数への介入の値を得る手法を提案している。本研究では、Blöbaum らのフレームワークを用いて時系列データに応用する手法を提案する。また、人工データと酢酸ビニル製造プラントのシミュレータを用いて提案手法を評価した。

## 2. 関連研究

変数間に因果関係を持つ構造データから因果グラフを抽出する手法は複数提案されている、LiNGAM [Shimizu 06] は、データ生成プロセスの線形性と変数が非ガウシアンノイズを含むことを仮定して因果構造を抽出する手法であり、その拡張として VAR モデルを利用して時系列性を持つデータに対して因果グラフを抽出する VAR-LiNGAM [Hyvärinen 10] が提案されている。

また、機械学習で用いられるモデルに対する介入効果を推測する手法として、Blöbaum らは因果グラフと機械学習モデルを接続するフレームワークを提案している [Blöbaum 17]。

本研究では、これらの研究を用いて、時系列データに対する Blöbaum らのフレームワークの有用性を評価する。VAR モデルから生成される時系列人工データと、酢酸ビニル製造プラントシミュレータから生成される時系列データに対して VAR-LiNGAM で因果グラフを抽出し、さらに線形回帰モデルを学習した上で両者を接続した場合とモデル単体から得られる介入効果を比較する。

## 3. 背景とフレームワーク

### 3.1 機械学習モデル

機械学習モデルのシステムへの応用において、予測だけではなく制御などの目的で入力と出力の振る舞いをすることは重要である。例えば機械学習モデルをシミュレータとして出力をコントロールしたい時、出力を目標値まで変動させるために入力をどの値まで変動させれば良いかを知ること、機械学習モデルを通じた操作やそこから得られた操作変数を通じたモデルの信頼性を議論することができる。モデル単体からこの値を求める時、線形回帰モデル  $y = \sum_i w_i x_i + w_0$  において、目標値  $d$  にするための特徴量  $x_k$  の値は以下の式で求まる。

連絡先: 切通 恵介, NTT コミュニケーションズ株式会社,  
k.kiritoshi@ntt.com

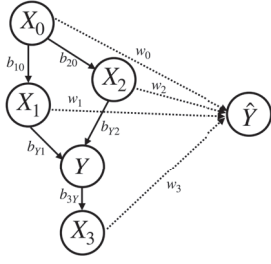


図 1: 入力変数と予測値の因果グラフ例:  $X_i$  は説明変数,  $\hat{Y}$  は予測モデルによる目的変数  $Y$  の予測値を表す. また実線と破線の矢印は, 実際のデータの生成過程による因果方向と予測モデルによる予測値への因果方向をそれぞれ示す.  $b_{ji}$  は  $X_i \rightarrow X_j$  の因果の強さを,  $w_i$  は変数  $X_i$  に対する予測モデルの重みを表す.

$$x_k = \frac{d - \sum_{j \neq k} w_j x_j - w_0}{w_k} \quad (1)$$

### 3.2 介入効果

あるデータの変数の集合  $V$  の変数  $X_i$  の関係を因果グラフとして構築した際, ある変数に対してその変数へのパスを無視して一定値を入力することを介入という. 図 1 に示す因果グラフの実線に着目した場合, 例えば  $X_0$  を定数  $c$  に固定した場合, その効果は子孫である変数  $X_1, X_2, Y, X_3$  に波及する. このような介入が他の変数に及ぼす影響 (介入効果) を調べることは因果グラフやその因果グラフに関連するシステムの振る舞いを知る上で重要である. 変数  $X_i$  に定数  $c$  で介入した時の変数  $X_j$  への介入効果は, 以下の式で表される.

$$\mathbb{E}[X_j | do(X_i = c)] \quad (2)$$

ただし, 記号  $do$  は  $X_i$  への介入を表す. 具体的な計算方法は 3.3 章で示す.

### 3.3 フレームワーク

あるシステムから得られたデータをもとに機械学習モデル構築し, そのモデルから式 (1) を利用して, 目標値  $d$  に近づける  $x_k$  を直接的に計算し, その値を実際のシステムに対して適用したとしても予測値が目標値に近づけることは難しい. なぜなら, 機械学習でモデル化するようなシステムの入力変数や予測する変数は一般的に因果構造を持つためである. 例として図 1 に元々のシステムの変数間の因果関係 (実線) と, 回帰モデルによる因果関係 (破線) を表す. この図からわかる通り,  $X_0$  に値を入力, すなわち介入を行うとその子孫である  $X_1, X_2, X_3$  も変動し, その両変数の変化した値を持って予測値  $\hat{Y}$  が決まる. 回帰モデルのみから計算された  $x_k$  は入力変数間の因果関係を考慮していないため, 予測値  $\hat{Y}$  の算出に  $X_1, X_2, X_3$  の介入効果は考慮されない. したがって, 実際のシステムにおけるモデルの入力への介入による予測値の変化を見積もることができない.

Blöbaum らは回帰モデルへの介入効果を測るためのフレームワークを提案している [Blöbaum 17]. 線形の因果グラフ  $V$  とモデルが与えられた時, 予測値  $\hat{Y}$  を目標値  $d$  とするための変数  $X_i$  に対する介入の値  $c$  を求めるフレームワークは以下の式で表現される.

$$c = \arg \min_{c^*} (\mathbb{E}[\hat{Y} | do(X_i = c^*)] - d)^2 \quad (3)$$

さらに, 因果グラフの介入効果は以下の式で逐次的に計算される.

$$\mathbb{E}[X_j | do(X_i = c)] = \begin{cases} c, & \text{if } j=i, \\ \mathbb{E}[X_j], & \text{if } X_j \text{ is a root variable,} \\ \sum_{X_k \in pa(X_j)} w_k \mathbb{E}[X_k | do(X_i = c)] + \mathbb{E}[N_j], & \text{otherwise,} \end{cases} \quad (4)$$

ただし  $pa(X_i)$  は変数  $X_i$  の親の変数の集合を,  $w_k$  は  $X_i$  から  $X_k$  への因果グラフ上の係数を示す. また,  $\mathbb{E}[N_j]$  は変数  $X_j$  に対するノイズの期待値であり, 以下の式で計算される.

$$\mathbb{E}[N_j] = \mathbb{E}[X_j] - \sum_{X_k \in pa(X_j)} w_k \mathbb{E}[X_k] \quad (5)$$

以上より, フレームワーク全体は以下の工程からなる.

1. 観測データから説明変数  $X$  と目的変数  $Y$  を定め, 予測モデルを作成し, さらに全ての変数を用いて因果グラフを推論する,
2. 介入する変数  $X_c$  を一つ選び, 式 (4) を用いて入力へ逐次的な介入効果を伝播させる.
3. 目標値  $d$  を定め, 式 (3) に従って介入後の予測値  $\hat{Y}$  が  $d$  に最も近づくような介入  $c$  を求める. 線形モデルの場合, 式 (3) は式 (1) によって求まる.

### 3.4 時系列データへの応用

時系列データに対する因果グラフを推論する手法として VAR-LiNGAM が提案されている [Hyvärinen 10]. VAR-LiNGAM で推論される因果グラフは以下の式で表現される.

$$\mathbf{x}(t) = \sum_{\tau=0}^T \mathbf{B}_\tau \mathbf{x}(t-\tau) + \mathbf{e}(t) \quad (6)$$

ここで  $\mathbf{x}(t) = [x_0(t), x_1(t), \dots, x_n(t)]^T$  は時刻  $t$  における変数ベクトル,  $\tau$  は時間遅れを表す. また,  $\mathbf{B}_\tau$  は変数  $x_i(t-\tau)$  間の因果関係を表す  $n \times n$  の係数行列である. この式で表される因果グラフは, 遅れ時間から現在時刻への変数間の因果関係  $x_i(t-\tau) \rightarrow x_j(t), (\tau = 1, 2, \dots, T)$  と, 現在時刻の変数間の因果関係  $x_i(t) \rightarrow x_j(t)$  を表している. VAR-LiNGAM では, 前者を自己回帰モデル (VAR モデル), 後者を自己回帰モデルとデータの残差に対して LiNGAM [Shimizu 06] を適用してそれぞれ推論する.

本研究では Blöbaum らのフレームワークを時系列データに適用するために, 章 3.3 の工程 1 において, 観測された時系列データに対して VAR-LiNGAM を利用して式 (6) で定義される因果グラフを推論する. さらに, 説明変数  $\mathbf{X}(t-i)$  から目的変数  $Y(t)$  を回帰する機械学習モデルを学習する.

## 4. 実験

本章では時系列の人工データおよび酢酸ビニル製造のプラントシミュレーターを用いて, 回帰モデル単体から数値計算的に得られる介入効果と, 本手法を用いた因果グラフと回帰モデルを接続して得られる介入効果を比較し, 時系列データに対する本手法の有用性を評価する.

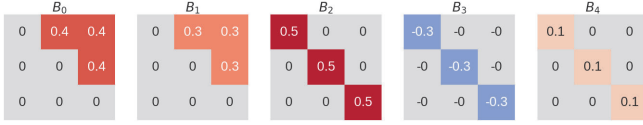


図 2: 生成モデル  $g$  の係数: 縦軸は  $\mathbf{x}(t)$ , 横軸は  $B_i$  における  $\mathbf{x}(t-i)$  を示し, 行列内の数値は因果関係の強さを表す係数で, 横軸の変数から縦軸の変数への因果を示している. 例えば  $B_1$  は  $x_1(t-1) \rightarrow x_0(t)$ ,  $x_2(t-1) \rightarrow x_0(t)$ ,  $x_2(t-1) \rightarrow x_1(t)$  の因果を持つ.



図 3: VARLiNGAM で予測された係数

#### 4.1 人工データ

人工データによる実験では, VAR-LiNGAM[Hyvärinen 10] で提案されている構造, VAR モデルと同時刻での線形変換から構成される以下の生成モデル  $G$  で時系列データを生成する.

$$\mathbf{x}'(t) = \sum_{\tau=1}^4 \mathbf{B}_{\tau} \mathbf{x}(t-\tau) + \mathbf{e}(t) \quad (7)$$

$$\mathbf{x}(t) = \mathbf{B}_0 \mathbf{x}'(t) + \mathbf{x}'(t) \quad (8)$$

ここで  $x_t$  は時刻  $t$  の 3 次元の変数ベクトルを表し,  $e_t$  はラプラス分布から生じるノイズを表す. また, 行列  $B_{\tau}$  は各時刻の変数の係数行列を表す (図 2). また, 時間遅れの次数は 4 とした.

実験ではまず,  $x_0(t)$  を目的変数とし,  $x_0(t-\tau)$ ,  $x_1(t-\tau)$ ,  $x_2(t-\tau)$  ( $\tau = 1, 2, \dots, 5$ ) を説明変数とした線形回帰モデルを学習する. さらに, 生成したデータにより VAR-LiNGAM を用いて因果グラフを抽出した. 抽出された因果関係を図 3 に示す. ここで,  $d = 20$  となる  $x_1(t-1)$  に対して (1) 回帰モデルから直接得られる  $c_m$ , (2) 提案手法によって求める  $c_{cm}$  を実際の生成モデルに入力し, その介入効果と  $d$  との差分 (平均二乗誤差) を評価する. 評価値は式 (7) で定義される生成モデル  $g$  とデータから学習された回帰モデル  $f$  により以下の式で表される.

$$\sqrt{\sum_i^n (d - f(g(\mathbf{x}_i | do(x_c = c))))^2} \quad (9)$$

ただし,  $n$  はサンプル数,  $x_c$  は介入した変数を表し, 本実験では  $x_c = x_1(t-1)$  である.

実験では, 生成モデル  $g$  から 10000 サンプルを生成し, 回帰モデルと生成モデル  $g$  に対する介入  $c_m, c_{cm}$  によって算出された式 (9) で表される評価値を表 1 に示す. 結果として, 線形回帰モデルのみから素朴に求めた値  $c$  と VAR-LiNGAM を用いた手法で求めた値  $c_{mc}$  により介入を行なった結果, 提案手法の方が回帰モデルの予測値が所期の値に近づくことを確認した.

表 1: 回帰モデルへの介入効果の評価

	回帰モデル単体	提案手法
評価値 (RMSE)	45.0	1.28

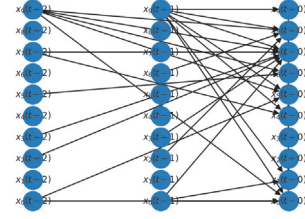


図 4: VAR-LiNGAM で抽出されたシミュレータの因果グラフ: 各ノードと矢印は変数と因果方向を表す.

#### 4.2 シミュレーションデータ

時系列データ実応用を狙いとして, 我々は酢酸ビニル製造のプラントシミュレータ [Chen 03, Kimura 19, Luyben 98] を用いて, 人工データと同様に, 介入により目標値に近い値が得られるかを評価した. 酢酸ビニル製造プラントは反応器 (リアクター), 蒸留塔などの複数の機器からなる系であり, シミュレータでは  $H_2O$ ,  $C_2H_2$  といった材料の量や酢酸ビニルを生成する際の温度や流量などの観測値, それを制御する MV や SV といった制御のパラメータを基に酢酸ビニルの収量や反応器の出口温度をシミュレートすることができる. SV は設定値を表し, MV は SV と実際の観測値との差分を元に, その差分が小さくなるように制御を与える, 自動計算される変数である.

本実験においては, 酢酸ビニル製造の全行程のうちリアクターのデータを用いた. リアクターの入出力及び制御パラメータとその説明を表 2 に示す. シミュレーションにおいては, リアクター出口温度 ( $x_0$ ) がリアクター目標値 SV に近づくように 1 時刻前の MV ( $x_9(t-1)$ ) を制御する. 本実験では SV の値を Chen らのデフォルトの設定値 159.17 で固定し, サンプリングレートを 1/180 分, データ取得間隔 1 分で 3600 分のデータを生成した [Chen 03].

生成されたリアクターのデータ ( $X_t$ ) を正規化し, VAR-LiNGAM を用いて因果グラフを構築した. さらに, リアクターの出力温度  $x_0(t=0)$  を目的変数として, 目的変数を除く変数を入力とした回帰モデルを学習した. VAR-LiNGAM で得られた因果グラフを図 4 に, 目的変数  $x_0(t)$  への各変数の因果関係の強さを図 5 に示す.

人工データと同様に, 目標値  $d = 159.3$  とした際の制御パラメータ MV に対する介入を算出する. 回帰モデル単体と提案手法から得られた介入値  $c_m, c_{cm}$  を用いて, シミュレータ  $g$  に対して介入を行い, 同様のパラメータ設定で 3600 分のシミュレーションを行なった. 評価値は式 (9) を用いた. 表 3 に示す結果から, 提案手法で得られた  $c_{mc}$  は回帰モデル単体で得られる  $c_m$  よりも, 予測値と目的値の誤差を小さくすることが確認できた. 一方で, 図 6 で示すように, 平均的には目標値に近い値を得られているものの, トレンドとしては下がり傾向にあり目標値に対して安定的な予測値を得ることができなかった. 時定数を考慮した介入を行うことが今後の課題である.

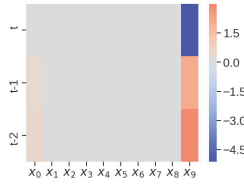


図 5:  $x_0(t)$  に対する因果関係の強さ: 縦軸は時刻, 横軸は変数, 図中の色は因果関係の強さを表す.  $x_0(t)$  に対する因果は  $x_9$  (MV) が支配的であることがわかる.

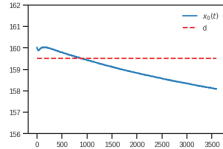


図 6: MV を  $c_{mc}$  で介入したシミュレーションで得られた  $x_0(t)$  の値

## 5. 終わりに

本研究では, Blöbaum らが提案した因果グラフと予測モデルを組み合わせることで予測モデルへの介入効果を得るフレームワークを用いて, 時系列データに適用する手法を提案した. 時系列データの因果グラフの抽出に VAR-LiNGAM を用いて時系列人工データと酢酸プラントのシミュレーションデータにフレームワークを適用し, 有用性を確認した. 今後の課題としては, より非線形性の強いデータに対する本手法の適用や, 予測モデルに時系列モデルを利用した場合の評価を行なっていく.

## 参考文献

- [Adadi 18] Adadi, A. and Berrada, M.: Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI), *IEEE Access*, Vol. 6, pp. 52138–52160 (2018)
- [Blöbaum 17] Blöbaum, P. and Shimizu, S.: Estimation of interventional effects of features on prediction, in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6IEEE (2017)
- [Chen 03] Chen, R., Dave, K., McAvoy, T. J., and Luyben, M.: A nonlinear dynamic model of a vinyl acetate process, *Industrial & engineering chemistry research*, Vol. 42, No. 20, pp. 4478–4487 (2003)
- [Hyvärinen 10] Hyvärinen, A., Zhang, K., Shimizu, S., and Hoyer, P. O.: Estimation of a structural vector autoregression model using non-gaussianity, *Journal of Machine Learning Research*, Vol. 11, No. May, pp. 1709–1731 (2010)
- [Kimura 19] Kimura, D., Shimada, K., and Tomonori, I.: Deep Neural Networks for a Chemical Plant Simulator, in *The 33rd Annual Conference of the Japanese Society for Artificial Intelligence*, pp. 4J2J1301–4J2J1301The Japanese Society for Artificial Intelligence (2019)

表 2: リアクターのセンサー

変数	説明
$x_0$	リアクター出口温度
$x_1$	リアクター出口流量
$x_2$	リアクター入力組成 ( $O_2$ )
$x_3$	リアクター入力組成 ( $CO_2$ )
$x_4$	リアクター入力組成 ( $C_2H_4$ )
$x_5$	リアクター入力組成 ( $C_2H_6$ )
$x_6$	リアクター入力組成 (VAc)
$x_7$	リアクター入力組成 ( $H_2O$ )
$x_8$	リアクター入力組成 (HAc)
$x_9$	リアクター出口温度操作パラメータ (MV)

表 3: シミュレーション実験における回帰モデルへの介入効果の評価

	回帰モデル単体	提案手法
評価値 (RMSE)	3.41	1.18

[Luyben 98] Luyben, M. L. and Tyréus, B. D.: An industrial design/control study for the vinyl acetate monomer process, *Computers & Chemical Engineering*, Vol. 22, No. 7-8, pp. 867–877 (1998)

[Shimizu 06] Shimizu, S., Hoyer, P. O., Hyvärinen, A., and Kerminen, A.: A linear non-Gaussian acyclic model for causal discovery, *Journal of Machine Learning Research*, Vol. 7, No. Oct, pp. 2003–2030 (2006)