

時間変化する因果関係の抽出に基づいた高速将来予測

千原 直己^{†,††} 松原 靖子[†] 藤原 廉^{†,††} 櫻井 保志[†]

[†] 大阪大学産業科学研究所 〒 567-0047 大阪府茨木市

^{††} 大阪大学大学院情報科学研究科 〒 565-0871 大阪府吹田市

E-mail: [†]{naoki88,yasuko,r-fujiwr88,yasushi}@sanken.osaka-u.ac.jp

あらまし 本論文では、大規模時系列データストリーム中の時間変化する因果関係の抽出および将来予測を同時に行うための最新手法 MODEPLAIT を提案する。提案手法は以下の優れた特性を全て満たす。(a) 時々刻々と変化する環境の移り変わりに従って変化する因果関係を明らかにする。(b) 時間変化する因果関係の抽出および将来予測を同時かつ正確に行う。(c) 計算時間は時系列データストリーム全体の長さに依存せず、高速に処理を行う。人工データおよび実データを用いた評価実験により、提案手法が最新の既存手法に比べて因果探索、将来予測の両方の観点において高精度であること、そして、計算効率の良い高速な処理が可能であることを明らかにした。

キーワード 時系列データ分析, ストリームデータ, 動的システム, 因果探索

1 はじめに

時系列データは、Internet of Things (IoT) [1,2], ソーシャルネットワーク [3,4], 医療情報分析 [5,6], ユーザー行動分析 [7] など多岐にわたるイベントおよびアプリケーションから大量に生成されている。特に、実世界においてはこれらのデータは高速かつ継続的に生成され続けるため、ストリーミング方式で処理することの需要が増している。

また、時系列データの観測値間にはさまざまな関係性 (e.g., 相関, 独立性) が存在し、それらはクラスタリング [8,9], 将来予測 [10], 欠損値補完 [11], など [12], 幅広い時系列分析にとって重要な特徴である。その中でも因果関係 [13,14] は特に価値のある洞察を提供することで知られており、多くの研究がその調査に注力している [15,16]。また、因果関係を帰納バイアスとして取り入れることで、より汎化性の高い表現を学習し、下流タスクの改善を図る研究も盛んである [17,18]。しかし、大半の研究では多次元時系列データ中の因果関係は定常であると仮定している。新しい原因を迅速に検出し、ストリーミング形式で正確に予測するには、時間変化する関係性の発見が重要である。例えば、感染症の流行では、ある国で新規のウイルス株が出現すると、その国内での感染者数が急増するだけでなく、国境を越える移動などといった特定の活動が他国の感染者数の増加を引き起こす可能性が存在する。また、このような原因となる国は時間の経過とともに変化する。

時間変化する因果関係を捉えるために、これらの変化の要因ともいえる、時々刻々と変化する環境の変化を捉える必要がある。これらは、特徴的な時系列パターンの遷移という形で現れる。具体的には、web の検索データの場合、時系列パターンの遷移は新商品の発売のようなさまざまな理由で生じる。そして、これらのイベントは既存の商品の売り上げなどに影響を与える、すなわち、因果関係の時間変化にまでつながるといえる。本論文ではこのような時系列パターンのことを“レジーム”と呼ぶ。

本研究では、上記で述べた時間変化する因果関係を構造方程式モデル [19] に基づいて表現し、時系列データストリームをリアルタイムに予測する手法として MODEPLAIT [20] を提案する。より具体的な本論文の問題は以下のとおりである。

問題: d 次元の観測データで構成される半無限長の時系列データストリーム $\mathbf{X} = \{\mathbf{x}(1), \dots, \mathbf{x}(t_c), \dots\}$ が与えられ、現時刻を t_c としたとき、(a) 潜在的な時系列パターン (レジーム) の遷移の検出、(b) その遷移に従って時間変化する因果関係の抽出 (c) l_s ステップ先の値 $\mathbf{v}(t_c + l_s)$ の予測を全て達成する。

1.1 具体例

図 1 は疫病データストリーム (#1) *covid19* における MODEPLAIT の予測結果の様子を示している。このデータセットは、5 か国 (日本, 米国, 中国, イタリア, 南アフリカ共和国) の COVID-19 感染者数で構成される。

図 1 (a) は MODEPLAIT が抽出した時間変化する因果関係を示している。図中の各矢印の基部は原因を、先端は結果を意味する。MODEPLAIT は、疫病データストリームから国間の時間変化する因果関係を正確に抽出する。たとえば、図 1 (a-i) は、南アフリカ共和国が他国に影響を及ぼしたことを示している。これは、南アフリカ保健省が 2020 年 12 月 18 日に南アフリカ共和国で COVID-19 の新たな変異株、つまり 501.V2 の発見を発表したという事実 [21] に対応している、故に、新たな変異株による影響を MODEPLAIT が適応的に検出したといえる。さらに、図 1 (a-ii) は、図 1 (a-i) とは対照的に、中国が他の国々に影響を与えたことを示している。これは、2022 年 4 月初旬から 2022 年 6 月 1 日まで続いた上海で最も長く厳しいロックダウンの期間と一致している [22]。これは、MODEPLAIT が上海での厳格で長期的なロックダウンにつながる程の COVID-19 の感染拡大の影響を検出したことを意味する。

図 1 (b) は、外生変数における潜在的なダイナミクスの固有値を示している。これらの図は複素平面であり、灰色の点線は単位円で、色付きの点は潜在的ダイナミクスの固有値であり、そ

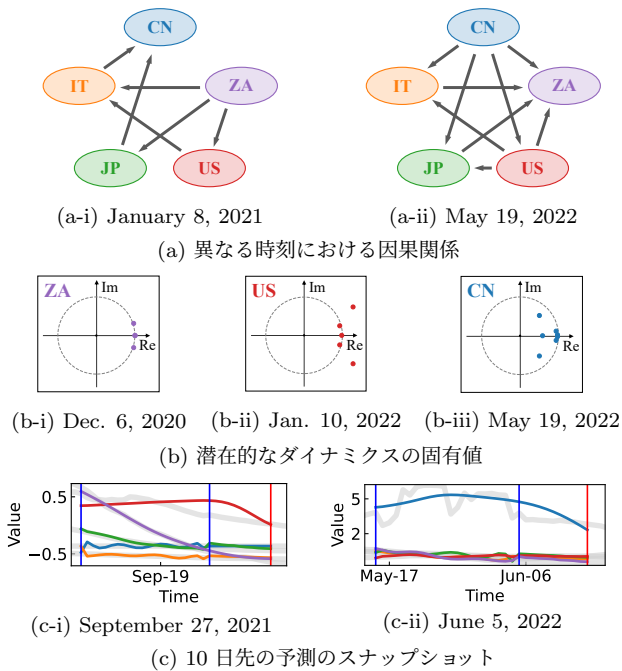


図 1: 疫病データストリームに対する MODEPLAIT の出力例。

それぞれが特定のモードの減衰率と時間周波数を示している。具体的には、固有値の絶対値が 1 より大きい場合、対応するモードは増幅を、1 より小さい場合は減衰を示す。図 1 (b-i) は、南アフリカ共和国の外生変数の弱い増幅モードを示しており、南アフリカ共和国での感染者数が前述の新たな変異株によって増加したことを意味する。図 1 (b-ii) は、米国における外生変数の強い増幅モードを示しており、新規感染者数が 1 日で初めて 100 万人を超えた状況を反映している [23]。図 1 (b-iii) は、中国に対応する外生変数の減衰モードを示している。この期間は上海のロックダウンの終わり頃で、感染の拡大が緩和し始めている事実があり、提案手法はこれを正確に捉えているといえる。

図 1 (c) は、現在のウィンドウが与えられた場合の $l_s = 10$ ステップ先の将来予測のスナップショットを示している。青い縦軸は、現在のウィンドウの開始 t_m と現在の時点 t_c を、赤い縦軸は l_s ステップ先の時点 $t_c + l_s$ を示している。また、推定された値は色付きの太線で示し、元の値は灰色の線で示す。MODEPLAIT は現在の特徴的な時系列パターンを正確に捉え、任意の時点で連続的に将来の値を生成する。

1.2 本論文の貢献

提案手法 MODEPLAIT の主要な貢献点を以下に示す。なお、本論文の内容は [24] に依拠しているため、全文については同文献を参照されたい。

- 時系列データ中のレジームの推移を逐次的に捉え続けることにより、時間変化する因果関係を発見する。
- 因果関係を理論的に発見し、高精度に将来を予測する。
- 計算時間はデータストリーム全体の長さに依存しない。

2 関連研究

時系列モデリングと将来予測. 時系列モデリングおよび将来予測は、多くの分野で大きな関心を集めている非常に重要な分

野である。自己回帰和分移動平均 (ARIMA) [25] やカルマンフィルター (KF) [26] は、従来のモデリングおよび予測方法の代表的な例であり、それらの派生型に関する研究も数多く行われている [27–30]。TICC [8] は、マルコフ確率場に基づいて異なる観測間の相互依存性を特徴付けるが、因果関係を捉えることはできない。また、ストリーミング形式に対応可能な手法は、時間やメモリの制約下で膨大なデータを処理するために不可欠な存在となっており、データマイニングやデータベース分野において非常に重要であることが実証されている [3, 31–33]。OrbitMap [34] は、ストリーム予測に焦点を当てた最新の手法で、主要な動的時系列パターン間の遷移を見つけることができる。ただし、観測間の因果関係を発見することはできない。さらに、深層学習モデルによる時系列予測の研究は、近年非常に活発である [10, 35–38]。深層学習モデルは有力ではあるものの、時系列分析に伴う計算コストが非常に高いため、最新の観測データによるモデルの継続的な更新が困難となり、ストリーミング形式の将来予測への適用は限定的である。

因果推論および探索. 因果推論や因果探索 [15, 16, 39–41] そして、課題解決のための因果の概念の応用 [17, 42, 43] などに関する幅広い研究が長年行われている。NOTEARS [44] は、有向非巡回グラフ (DAG) の推定問題を、非巡回制約項を組み込んだ滑らかな連続最適化問題として定式化した微分可能なフレームワークである。グレンジャー因果 [45] は、時間的な因果関係を解析するために広く利用されている。しかしながら、グレンジャー因果は変数間の予測的因果関係を示すものであり、典型的な因果関係とは異なる [46]。具体的には、従来の因果関係はある観測が原因となって別の観測を引き起こすかどうかを表すのに対し、グレンジャー因果性はある観測が別の観測の予測に役立つかどうかを示す [47]。本論文では、時系列データストリーム中の時間変化する因果関係の抽出に焦点を当てる。

3 提案モデル

本章では、提案モデルを紹介する。本題に入る前に、MODEPLAIT の基本的な概念について簡潔に説明する。提案モデルは $\mathbf{X}_{\text{sem}} = \mathbf{B}_{\text{sem}}\mathbf{X}_{\text{sem}} + \mathbf{E}_{\text{sem}}$ のように記述される構造方程式モデル (SEM) [19] に基づいて設計される。ここで、 \mathbf{X}_{sem} は観測変数、 \mathbf{B}_{sem} は因果隣接行列、 \mathbf{E}_{sem} は非ガウス分布を持つ相互に独立した外生変数である。本論文では、データ生成過程は線形であり、因果関係を有向非巡回グラフで表現し、未観測共通原因が存在しないことを仮定する。構造方程式モデルは典型的な因果関係を表現できるが、実社会に存在する特徴的な時系列パターンの遷移に応じて時間変化する因果関係は表現できない。そこで、提案モデルは外生変数が動的システムとして振る舞うと仮定する。しかし、外生変数は互いに独立しているため、それらを単一の動的システムとして考えるのは不適切である。これらを踏まえて、本研究では、様々な特徴的な時系列パターン (レジーム) を含む時系列データストリームが与えられたとき、時系列データを逐次的に要約し、ストリーミング形式で時間変化する因果関係を発見することを目指す。

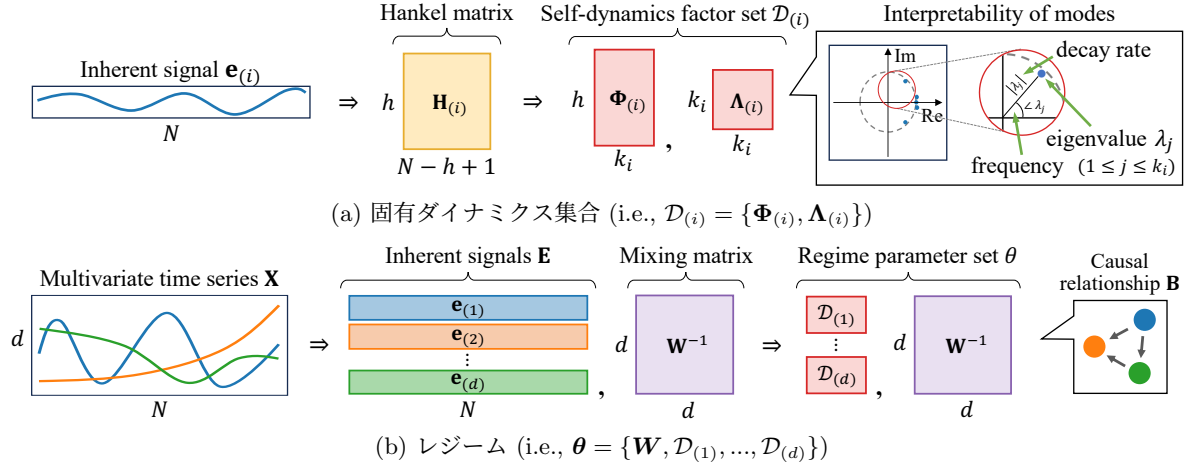


図 2: MODEPLAIT のモデル概要図: (a) 動的システムに従う i 番目の変数に固有な単変量信号 $e_{(i)}$ から潜在的なダイナミクスを抽出する. (b) 多次元時系列データは混合行列 \mathbf{W}^{-1} と d 個の自己駆動因子 $\{\mathcal{D}_{(1)}, \dots, \mathcal{D}_{(d)}\}$ によって表現される.

3.1 ModePlait モデル

本節では提案モデルの詳細を述べる. 初めに必要な概念の定義について説明する.

定義 1 (固有信号: \mathbf{E}). 非ガウス分布に従う d 個の相互に独立した要素を持つ信号 $\mathbf{E} = \{e_{(i)}\}_{i=1}^d$ を固有信号と呼ぶ. ただし, $e_{(i)} = \{e_{(i)}(1), \dots, e_{(i)}(t)\}$ は i 番目の単変量時系列である. これは, 時間の経過に従って変化するという特徴がある.

図 2 は提案モデルの全体図である. 提案手法は, 以下のような特徴を捉えることで目的を達成する.

(P1) 固有信号の潜在的な時間ダイナミクス

(P2) 単一レジーム内の特徴的な時系列パターン

(P1) は, 固有信号を基底ベクトル (モード) の重ね合わせで表現する. そして, 上記の要素を組み合わせて (P2) を捉える.

3.1.1 固有信号中の潜在的な時間ダイナミクス (P1)

初めに, i 番目の固有信号 $e_{(i)} = \{e_{(i)}(1), \dots, e_{(i)}(t)\}$ から潜在的な時間ダイナミクスを捉える方法について説明する. 問題点としては, システム内の潜在的なダイナミクスが一般に多次元であるため, システムを十分に表現するためには, 単次元なデータではしばしば不十分であることが挙げられる. この問題点を補うために, 状態空間の拡張手法を活用する. 特に, 本研究では非線形なダイナミクスの抽出に有効な時間遅れ埋め込みを採用する. 具体的には, これは一般的な観測量 $g(e_{(i)}(t)) := (e_{(i)}(t), e_{(i)}(t-1), \dots, e_{(i)}(t-h+1)) \in \mathbb{R}^h$ に基づいており, 非線形システムのアトラクタを幾何学的に再構成するための確立された手法である. ただし, h は埋め込み次元である. 上記の $g(\cdot)$ を用いてハンケル行列を形成する.

$$\mathbf{H}_{(i)} = \begin{bmatrix} | & | & & | \\ g(e_{(i)}(h)) & g(e_{(i)}(h+1)) & \cdots & g(e_{(i)}(t)) \\ | & | & & | \end{bmatrix} \quad (1)$$

式 (1) のとおり, 各状態ベクトルは過去情報を付与して拡張されている. さらに, Takens の埋め込み定理 [48] によれば, 特定の条件下において, 時間遅れ埋め込みによって生成されるベ

クトルは, 元の状態と微分同相なダイナミクスを持つことが保証される. 直感的に説明をすると, この再構成は元の力学系の特性を理論的に保つ, つまり, ハンケル行列 $\mathbf{H}_{(i)}$ の解析を通じて, 元のデータからは直接抽出できない重要な特徴を明らかにすることを可能にする. 多くの場合, 微分同相写像を犠牲にすることなく埋め込み次元を選択できる.

ここで, i 番目の固有信号 $e_{(i)}$ の動的システムのために, k_i 次元の複素数値の潜在状態 $\mathbf{s}_{(i)}(t) \in \mathbb{C}^{k_i}$ を導入する. ただし, k_i はモードの数である. したがって, i 番目の固有信号 $e_{(i)}$ は以下の式で記述される.

モデル 1. $\mathbf{s}_{(i)}(t)$ を時刻 t における k_i 次元の潜在状態とする. i 番目の単変量固有信号 $e_{(i)}$ は次の式で表現される.

$$\begin{aligned} \mathbf{s}_{(i)}(t+1) &= \Lambda_{(i)} \mathbf{s}_{(i)}(t) \\ e_{(i)}(t) &= g^{-1}(\Phi_{(i)} \mathbf{s}_{(i)}(t)) \end{aligned} \quad (2)$$

ここで, $g^{-1}(\cdot)$ は観測量 $g(\cdot)$ の逆写像, $\Phi_{(i)}$ の各列が各モードで, $\Lambda_{(i)}$ が k_i 個の固有値である.

$\mathbf{s}_{(i)}(t)$ は k_i 個のモードの重ね合わせで表現される. そして, 固有値 $\Lambda_{(i)} \in \mathbb{C}^{k_i \times k_i}$ が時間ダイナミクスを示し, モード $\Phi_{(i)} \in \mathbb{C}^{h \times k_i}$ および $g^{-1}(\cdot)$ は時刻 t における i 番目の固有信号 $e_{(i)}(t)$ を生成するための射影を示す. まとめて, 以下を得る.

定義 2 (固有ダイナミクス集合: $\mathcal{D}_{(i)}$). モード $\Phi_{(i)}$ と固有値 $\Lambda_{(i)}$ による集合 $\mathcal{D}_{(i)} = \{\Phi_{(i)}, \Lambda_{(i)}\}$ を固有ダイナミクス集合と呼ぶ. これは, i 番目の単変量固有信号 $e_{(i)}$ の潜在的な時間ダイナミクスを表現する.

3.1.2 単一レジーム内の特徴的な時系列パターン (P2)

続いて, 時系列データストリーム中の時間変化する因果関係を考慮した特徴的な時系列パターン (レジーム) を表現する方法について述べる. 時刻 t における推定値 $\mathbf{v}(t) \in \mathbb{R}^d$ を生成するためのモデルを, d 個の固有ダイナミクス集合 $\mathcal{D}_{(1)}, \dots, \mathcal{D}_{(d)}$ で構築する. また, d 個の潜在状態を $\mathbf{S}(t) = \{\mathbf{s}_{(i)}(t)\}_{i=1}^d$ と表記する. したがって, 多変量時系列データはモデル 1 を拡張した以下の式で記述される.

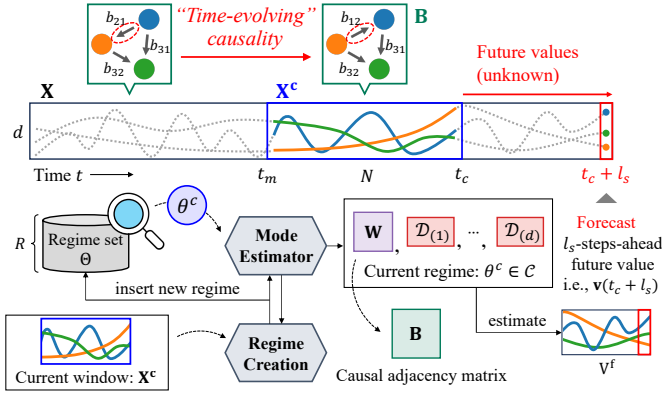


図 3: MODEPLAIT のアルゴリズム概要図: 逐次的に最新のデータ $\mathbf{x}(t_c)$ を入力として受け取り、カレントウィンドウ \mathbf{X}^c に対するを最もよく表現する θ^c を探索する。その後、 θ^c を用いて因果隣接行列 \mathbf{B} を探索し、 l_s ステップ先の値 $\mathbf{v}(t_c + l_s)$ を予測する。 \mathbf{X}^c 中に未知のパターンを検出した場合、新たなレジームを推定し、レジームセット Θ に追加する。

モデル 2. $\mathbf{s}_{(i)}(t)$ を時刻 t の i 番目の固有信号 $e_{(i)}(t)$ のための k_i 次元の潜在状態、 $\mathbf{e}(t)$ を時刻 t の d 次元の固有信号 ($\mathbf{e}(t) = \{e_{(i)}(t)\}_{i=1}^d$)、 $\mathbf{v}(t)$ を時刻 t の d 次元の推定値とする。レジームは次の式で表現される。

$$\begin{aligned} \mathbf{s}_{(i)}(t+1) &= \mathbf{\Lambda}_{(i)} \mathbf{s}_{(i)}(t) \quad (1 \leq i \leq d) \\ e_{(i)}(t) &= \mathbf{g}^{-1}(\mathbf{\Phi}_{(i)} \mathbf{s}_{(i)}(t)) \quad (1 \leq i \leq d) \\ \mathbf{v}(t) &= \mathbf{W}^{-1} \mathbf{e}(t) \end{aligned} \quad (3)$$

モデル 2 のために、新たなパラメータである分離行列 \mathbf{W} を導入する。これは、 d 個の固有信号間の関係性を表現し、因果関係の特定のために重要な役割を果たす。 \mathbf{W} から \mathbf{B} を特定するためのアルゴリズムについては 4.2.3 節にて説明する。まとめると、以下を得る。

定義 3 (レジーム: θ). $\theta = \{\mathbf{W}, \mathcal{D}_{(1)}, \dots, \mathcal{D}_{(d)}\}$ をレジームを表現するパラメータ集合とする。ここで、 \mathbf{W} は因果隣接行列 \mathbf{B} を生成するための基盤となる要素である。

最終的な目的は、複数のレジーム θ が時々刻々と遷移する様子を捉えることである。なお、このレジームの遷移が因果関係の時間変化を誘発する。時刻 t までの適切なレジーム数を R とすると、時系列データストリーム \mathbf{X} は R 個のレジーム $\{\theta^1, \dots, \theta^R\}$ によって要約される。これらを踏まえて、レジームセットおよび時間変化する因果関係を以下のように定義する。

定義 4 (レジームセット: Θ). 複数のレジームで構成されるモデルパラメータ集合 $\Theta = \{\theta^1, \dots, \theta^R\}$ として定義し、レジームセットと呼ぶ。これは、時系列データストリーム全体の特徴的な複数の時系列パターンを表現する。

定義 5 (時間変化する因果関係: \mathcal{B}). 因果隣接行列の集合を $\mathcal{B} = \{\mathbf{B}^1, \dots, \mathbf{B}^R\}$ として定義し、時間変化する因果関係と呼ぶ。ここで、 \mathbf{B}^i は i 番目のレジーム θ^i に対応する因果隣接行列である。

4 アルゴリズム

本章では、時間変化する因果関係 \mathcal{B} およびレジームセット Θ を推定するための効率的なアルゴリズムを提案する。図 3 は、提案アルゴリズムの概要図である。初めに、単一のレジームのみを持つと仮定された多変量時系列データからレジームを算出する効果的な方法を提示する。その後、複数の異なる特徴的な時系列パターンを含む時系列データストリームに対して、 \mathcal{B} を同定しながら、 Θ を逐次的に更新するアルゴリズムを説明する。

4.1 RegimeCreation

初めに、レジーム $\theta = \{\mathbf{W}, \mathcal{D}_{(1)}, \dots, \mathcal{D}_{(d)}\}$ を推定するためのアルゴリズム REGIMECREATION を提案する。このアルゴリズムは、以下の 2 つの主要な手順から構成される。(i) \mathbf{X} を分離行列 \mathbf{W} と固有信号 \mathbf{E} に分解する。(ii) 式 (2) に従って、 d 個の固有ダイナミクスセット $\{\mathcal{D}_{(1)}, \dots, \mathcal{D}_{(d)}\}$ を計算する。最適な因果関係を捉えるために、独立成分分析 (ICA) を使用して \mathbf{X} の分解を分解する。次に、 i 番目の固有ダイナミクスセット $\mathcal{D}_{(i)}$ の計算に関しては、ハンケル行列 $\mathbf{H}_{(i)}$ に基づく以下のデータ行列を使用する。

$$\begin{aligned} \mathbf{L}_{(i)} &= \begin{bmatrix} g(e_{(i)}(h+1)) & \dots & g(e_{(i)}(t)) \end{bmatrix} \in \mathbb{R}^{h \times (t-h)} \\ \mathbf{R}_{(i)} &= \begin{bmatrix} g(e_{(i)}(h)) & \dots & g(e_{(i)}(t-1)) \end{bmatrix} \in \mathbb{R}^{h \times (t-h)} \end{aligned}$$

これを基に、本論文では以下の重み付き損失関数を使用する。

$$\begin{aligned} \min_{\mathbf{A}} \sum_{t'=h}^{t-1} \mu^{2(t-1-t')} \|g(e_{(i)}(t'+1)) - \mathbf{A}_{(i)} g(e_{(i)}(t'))\|_2^2 \\ = \min_{\mathbf{A}} \|(\mathbf{L}_{(i)} - \mathbf{A}_{(i)} \mathbf{R}_{(i)}) \mathbf{M}\|_F^2 \end{aligned} \quad (4)$$

ここで、 $\mathbf{A}_{(i)}$ は遷移行列であり、これは固有値分解が与えるモード $\mathbf{\Phi}_{(i)}$ と対応する固有値分解 $\mathbf{\Lambda}_{(i)}$ により構成される。また、 $\mathbf{M} = \text{diag}(\mu^{t-h-1}, \dots, \mu^0) \in \mathbb{R}^{(t-h) \times (t-h)}$ は忘却行列である。これは逐次最小二乗法の原理に基づいている。加えて、Koopman 作用素論によると、古典的な線形時不変システムのモード分解とは異なり、遷移行列は線形であるが非線形動的システムにも適用可能である。まとめると、以下を得る。

- I. 独立性分析を使用して $\mathbf{X} = \mathbf{W}^{-1} \mathbf{E}$ を計算する。
- II. 式 (1) に従い、ハンケル行列 $\mathbf{H}_{(i)}$ を構成する。
- III. データ行列のペア $(\mathbf{L}_{(i)}, \mathbf{R}_{(i)})$ を計算する。
- IV. 特異値分解 (SVD) を使用して $\mathbf{R}_{(i)} \mathbf{M} = \mathbf{U}_{(i)} \mathbf{\Sigma}_{(i)} \mathbf{V}_{(i)}^\top$ を計算する。特異値の数は [49] に従って決定する。
- V. 遷移行列 $\mathbf{A}_{(i)}$ を左特異値ベクトル $\mathbf{U}_{(i)}$ が張る k_i 次元部分空間に射影する。

$$\tilde{\mathbf{A}}_{(i)} = \mathbf{U}_{(i)}^\top \mathbf{A}_{(i)} \mathbf{U}_{(i)} = \mathbf{U}_{(i)}^\top \mathbf{L}_{(i)} \mathbf{M} \mathbf{V}_{(i)} \mathbf{\Sigma}_{(i)}^{-1} \in \mathbb{R}^{k_i \times k_i}$$

- VI. 固有値分解 $\tilde{\mathbf{A}}_{(i)} \mathbf{Z}_{(i)} = \mathbf{Z}_{(i)} \mathbf{\Lambda}_{(i)}$ を計算する。ここで、 $\mathbf{U}_{(i)}$ は直交行列であるため、固有値行列 $\mathbf{\Lambda}_{(i)}$ は $\mathbf{A}_{(i)}$ の支配的な k_i 個の固有値と一致する。
- VII. モード $\mathbf{\Phi}_{(i)} = \mathbf{U}_{(i)} \mathbf{Z}_{(i)}$ を計算する。

定理 1. REGIMECREATION の計算量は $O(N(d^2 + h^2) + k^3)$ である。ただし、 $k = \max_i(k_i)$ である。証明は [24] を参照。

4.2 ストリーミングアルゴリズム

続いて、提案モデルを用いてストリーミング方式で分離行列 $\mathbf{W} \in \Theta$ から因果隣接行列 \mathbf{B} を特定し、将来を予測する方法を提案する。本題に入る前に、いくつかの重要な概念を定義する。

定義 6 (更新用パラメータ: ω)。レジーム θ を更新するためのパラメータ集合を $\omega = \{\{\mathbf{P}_{(i)}\}_{i=1}^d, \{\epsilon_{(i)}\}_{i=1}^d\}$ と定義し、更新用パラメータと呼ぶ。ただし、 $\mathbf{P}_{(i)} = (\mathbf{R}_{(i)}\mathbf{M}\mathbf{R}_{(i)}^\top)^{-1}$ 、 $\epsilon_{(i)}$ はエネルギーである。

定義 7 (モデルパラメータ集合: \mathcal{F})。 $\mathcal{F} = \{\Theta, \Omega\}$ を MODEPLAIT のパラメータ集合とする。ただし、 Θ および Ω は R 個のレジーム $\Theta = \{\theta^1, \dots, \theta^R\}$ 、更新用パラメータ集合 $\Omega = \{\omega^1, \dots, \omega^R\}$ によって構成される。

これらの定義に基づくと、問題定義は以下のとおりである。

問題 1. 時刻 t_c における最新の値 $\mathbf{x}(t_c)$ として、時系列データストリーム \mathbf{X} が与えられたとき、

- 最適なモデルパラメータ集合 $\mathcal{F} = \{\Theta, \Omega\}$ を発見する、
- 時間変化する因果関係 \mathcal{B} を抽出する、
- l_s ステップ先の値 $\mathbf{v}(t_c + l_s)$ を予測する。

ここで、カレントウィンドウ $\mathbf{X}^c = \mathbf{X}[t_m : t_c]$ のレジームを θ^c と、 θ^c に対応する更新パラメータを ω^c と呼ぶ。さらに、 l_s ステップ先の値 $\mathbf{v}(t_c + l_s)$ を予測するには、現在の時刻 t_c での潜在ベクトル $\mathbf{S}(t_c)$ が必要であるため、これを \mathbf{S}_{en}^c として保持する。要約すると、提案アルゴリズムは、これらをモデル候補 $\mathcal{C} = \{\theta^c, \omega^c, \mathbf{S}_{en}^c\}$ として保持する。

4.2.1 全体像

MODEPLAIT は以下のアルゴリズムによって構成される。

- MODEESTIMATOR: 最適なモデルパラメータ集合 \mathcal{F} およびモデル候補 \mathcal{C} を推定する。
- MODEGENERATOR: モデル候補 \mathcal{C} より、 l_s ステップ先の値 $\mathbf{v}(t_c + l_s)$ を予測し、因果隣接行列 \mathbf{B} を抽出する。
- REGIMEUPDATER: 現在の更新用パラメータ ω^c と最新の値 $\mathbf{x}(t_c)$ を用いて、現在のレジーム θ^c を更新する。

4.2.2 MODEESTIMATOR

現在の時刻 t_c の観測値 $\mathbf{x}(t_c)$ が与えられたとき、最初にモデルパラメータ集合 \mathcal{F} およびカレントウィンドウ \mathbf{X}^c を最も表現するモデル候補 \mathcal{C} を逐次的に更新する。ここで、 $f(\mathbf{X}^c; \mathbf{S}_0^c, \theta^c)$ はカレントウィンドウ \mathbf{X}^c と推定ウィンドウ \mathbf{V}^c の誤差を最小化することによって、最適なモデルパラメータ集合を算出する (i.e., $f(\mathbf{X}^c; \mathbf{S}_0^c, \theta^c) = \sum_{t=t_m+h-1}^{t_c} \|\mathbf{x}(t) - \mathbf{v}(t)\|$)。式 (3) に基づくと、 \mathbf{S}_0^c を計算する最も簡便な方法は $\{\Phi_{(i)}^\dagger g(e_{(i)}(t_m + h - 1))\}_{i=1}^d$ を用いることである。しかし、過度なノイズが含まれた初期値は適切な予測が達成できない。これを対処するために、LM (Levenberg-Marquardt) アルゴリズム [50] を使用して \mathbf{S}_0^c を最適化し、観測におけるノイズの影響を除去する。まとめると、MODEESTIMATOR は次の手順に

従う。

- I. カレントウィンドウ \mathbf{X}^c と現在のレジーム θ^c 間の誤差を最小化するように初期値 \mathbf{S}_0^c を最適化する
- II. $f(\mathbf{X}^c; \mathbf{S}_0^c, \theta^c) > \tau$ のとき、最適なレジーム $\theta \in \Theta$ を得る。
- III. $f(\mathbf{X}^c; \mathbf{S}_0^c, \theta^c) > \tau$ のとき、REGIMECREATION を用いて新たなレジームを生成し、レジームセット Θ に追加する。

4.2.3 MODEGENERATOR

続いて、逐次的に因果隣接行列 \mathbf{B} を抽出し、 l_s ステップ先の値 $\mathbf{v}(t_c + l_s)$ を予測するアルゴリズム MODEGENERATOR を提案する。予測については、式 (3) に従って $\mathbf{v}(t_c + l_s)$ を推定する。一方で、因果隣接行列 \mathbf{B} については前述のとおり分離行列 $\mathbf{W} \in \Theta^c$ から抽出する。独立成分分析で得られる混合行列 (分離行列の逆行列) には独立成分の順序および尺度という 2 つの主要な不定性が存在する。しかし、最適な因果隣接行列を抽出するためには、これらの問題を解決しなければならない。上記の不定性を解消し、因果隣接行列 \mathbf{B} を特定するアルゴリズムは以下のとおりである。

- I. \mathbf{W} の行を並べ替えることで、主対角線上にゼロを含まない行列 $\tilde{\mathbf{W}}$ を得る。
- II. $\tilde{\mathbf{W}}$ の各行を対応する対角要素で割ることで、主対角線上にすべて 1 を持つ新たな行列 $\tilde{\mathbf{W}}'$ を得る。
- III. \mathbf{B} の推定値を $\hat{\mathbf{B}} = \mathbf{I} - \tilde{\mathbf{W}}'$ によって算出する。
- IV. 最後に、因果順序を得るために、 $\hat{\mathbf{B}}$ の置換行列 \mathbf{K} を用いて $\tilde{\mathbf{B}} = \mathbf{K}\hat{\mathbf{B}}\mathbf{K}^\top$ を計算する。これは、 $\tilde{\mathbf{B}}$ の上三角行列の要素の総和を最小化する。

定理 2. MODEPLAIT における因果探索は、MODEGENERATOR での因果隣接行列 \mathbf{B} の抽出と同値である。証明は [24] を参照。

4.2.4 REGIMEUPDATER

最後に、既存のレジーム θ の表現力を向上させるための最新の値 $\mathbf{x}(t_c)$ を使用したレジームのパラメータの更新方法について説明する。REGIMEUPDATER は主に、(i) 分離行列 \mathbf{W} の更新、および (ii) 固有ダイナミクス集合 \mathcal{D} の更新、の 2 つの手順から構成される。手順 (i) では、適応フィルタに基づいたアルゴリズムを使用する [51, 52]。これは、計算とメモリの両方の観点で非常に効率的である。更新手順は以下のとおりである。

- I. 現在時刻 t_c において、更新前の \mathbf{W} の i 番目の行ベクトル \mathbf{w}_i に $\mathbf{x}(t_c)$ を射影することで、 i 番目の固有信号 $g(e_{(i)}(t_c))$ を算出する。
 - II. $g(e_{(i)}(t_c))$ を用いて、復元誤差およびエネルギー $\epsilon_{(i)}$ を計算する。
 - III. 誤差およびエネルギー $\epsilon_{(i)}$ を用いて \mathbf{w}_i を更新する。
- 一方で、手順 (ii) では以下の再帰式を用いる。

$$\begin{aligned} \mathbf{A}_{(i)}^{new} &= \mathbf{A}_{(i)}^{prev} + (g(e_{(i)}(t_c)) - \mathbf{A}_{(i)}^{prev} g(e_{(i)}(t_c - 1))) \gamma_{(i)} \\ \gamma_{(i)} &= \frac{g(e_{(i)}(t_c - 1))^\top \mathbf{P}_{(i)}^{prev}}{\mu + g(e_{(i)}(t_c - 1))^\top \mathbf{P}_{(i)}^{prev} g(e_{(i)}(t_c - 1))} \end{aligned} \quad (5)$$

$$\mathbf{P}_{(i)}^{new} = \frac{1}{\mu} (\mathbf{P}_{(i)}^{prev} - \mathbf{P}_{(i)}^{prev} g(e_{(i)}(t_c - 1)) \gamma_{(i)})$$

ここで、 $\Phi_{(i)}$ および $\Lambda_{(i)}$ はそれぞれ $\mathbf{A}_{(i)}$ の固有ベクトル、固

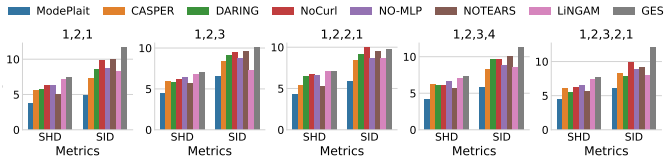


図 4: 人工データに対する因果探索の結果

有値を表す。この再帰式は、式 (4) に示した重み付き損失関数を最小化し、最新の値 $\mathbf{x}(t_c)$ に着目することで、時系列パターンの変化に適応する。この式の導出については [24] を参照されたい。まとめると、REGIMEUPDATER は以下のとおり。

- I. 手順 (i) で述べたアルゴリズムに従って、最新の値 $\mathbf{x}(t_c)$ を用いて分離行列 \mathbf{W} を更新する。
- II. 更新後の分離行列 \mathbf{W} を用いて、カレントウィンドウ \mathbf{X}^c から現在の固有信号を \mathbf{E}^c を計算する。
- III. 式 (5) に従って、固有ダイナミクス集合 $\mathcal{D}_{(i)}$ を更新する。

定理 3. 定理 1 に基づくと、各プロセスにおける MODEPLAIT の計算時間量は少なくとも $O(N \sum_i k_i + dh^2)$ であり、高々 $O(RN \sum_i k_i + N(d^2 + h^2) + k^3)$ である。証明は [24] を参照。

この理論的解析により、提案手法はデータストリーム全体の長さ t_c に対して一定の計算時間しか必要としないことが示される。したがって、MODEPLAIT は実行速度の観点から半無限長のデータストリームに対して実用的であるといえる。

5 評価実験

本論文では、MODEPLAIT の有効性を検証するため、人工データおよび実データを用いた実験を行った。本章では以下の項目について検証する。

- Q1** 提案手法が捉える時間変化する因果関係の有効性
- Q2** 因果探索および予測に対する提案手法の精度の検証
- Q3** データストリームの予測に対する計算時間の検証

実験には Intel Xeon Platinum 8268 2.9GHz 24 core CPU, 512GB DDR4 RAM, NVIDIA RTX A6000 GPU を搭載した Linux マシンを使用した。

データセット. 使用した人工データおよび実データは以下のとおりである。また前処理として、各データセットは平均値と分散で正規化 (z-normalization) して使用した。

- (#0) *synthetic*: 構造方程式モデル [19] に基づいて生成した。詳細については [24] を参照されたい。
- (#1) *covid19*: Google COVID-19 Open Data [53] から収集されたデータであり、日本、アメリカ、中国、イタリア、南アフリカ共和国の感染者数によって構成されている。日ごとに記録された 900 日以上のデータが含まれている。
- (#2) *web-search*: GoogleTrends から 10 年間にわたり毎週収集された、ビールに関連する検索クエリの検索件数で構成されている [54]。
- (#3) *chicken-dance*, (#4) *exercise*: CMU motion capture database [55] から得られるデータであり、左脚、右脚、左腕、右腕に対応する 4 次元ベクトルで構成される。

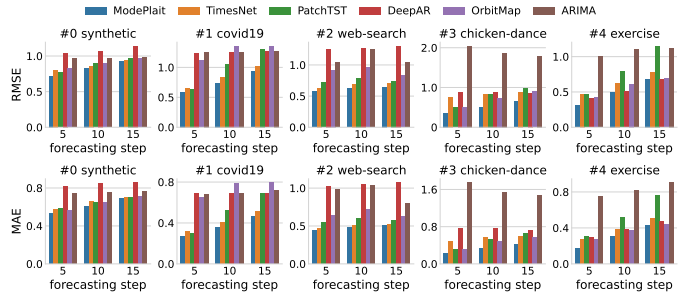


図 5: 多次元時系列予測の結果

比較手法. 本研究では、因果探索のベースラインとして以下の 7 つの手法と比較を行った。具体的には、CASPER [41], DARING [15], NoCurl [56], NOTEARS-MLP (以下、NO-MLP と略記) [57], NOTEARS [44], LiNGAM [39], および GES [58] である。また、時系列予測の比較手法として、TimesNet [38], PatchTST [10], DeepAR [35], OrbitMap [34], および ARIMA [25] の 5 つの手法を採用した。実験設定の詳細については、[24] に記載している。

5.1 Q1: 提案手法の有効性

疫病データストリームに対する提案手法の結果は 1 章の図 1 に示したとおりである。MODEPLAIT は複雑な社会の情勢の変化が反映された時間変化する因果関係の探索を実現している。加えて、この特徴を用いて高精度な将来予測を達成した。

5.2 Q2: 提案手法の精度

因果探索精度. 提案手法の因果探索の精度を検証するために、人工データを用いた実験を行った。評価指標については structural Hamming distance (SHD) および structural intervention distance (SID) を採用した。SHD は因果隣接行列の差異を定量化する指標であり、欠落した辺、余分な辺、逆転した辺の数を計測する。一方、SID は因果探索の精度を評価するために特に適しており、推定された因果隣接行列を使用した場合に、介入分布 $p(x_j | \text{do}(X_i = \bar{x}))$ が誤って計算されるような組 (i, j) の数を測定する。図 4 は、複数の人工データセットに対する各手法の因果探索の精度を示している。提案手法はすべてのデータセットにおいて一貫して比較手法を上回る性能を示す。この結果は定理 2 の解析結果と一致している。比較手法については、いずれの手法も時系列データストリーム内の時間変化する因果関係を処理できないため精度が低下したといえる。

予測精度. 続いて、MODEPLAIT の l_s ステップ先の予測精度を検証する。評価指標は、推定値の二乗平均誤差 (RMSE: root mean square error) および平均絶対誤差 (MAE: mean absolute error) の 2 種類を利用した。図 5 は、時系列データストリームにおける MODEPLAIT および比較手法の予測精度を示している。MODEPLAIT は最新の比較手法を上回る予測精度を示している。深層学習モデルは高い表現能力を有しているが、逐次的なパラメータ更新を達成できないため、予測精度が低下する。OrbitMap は複数の特徴的な時系列パターンを扱うことが可能であるが、時間変化する因果関係を捉えることはできないことが性能低下の原因であると考えられる。

表 6: アブレーション研究の結果

Datasets		#0 synthetic		#1 covid19		#2 web-search		#3 chicken-dance		#4 exercise	
Metrics		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
MODEPLAIT (full)	5	0.722	0.528	0.588	0.268	0.573	0.442	0.353	0.221	0.309	0.177
	10	0.829	0.607	0.740	0.361	0.620	0.481	0.511	0.325	0.501	0.309
	15	0.923	0.686	0.932	0.461	0.646	0.505	0.653	0.419	0.687	0.433
w/o causality	5	0.759	0.563	0.758	0.374	0.575	0.437	0.391	0.262	0.375	0.218
	10	0.925	0.696	0.848	0.466	0.666	0.511	0.590	0.398	0.707	0.433
	15	1.001	0.760	1.144	0.583	0.708	0.545	0.821	0.537	0.856	0.533

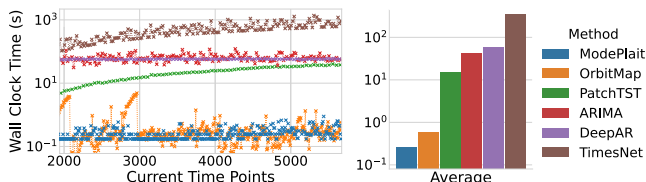


図 7: 各時刻 t_c における計算コスト (左) と平均値 (右): これらの図の y 軸は対数スケールで表示されている。

アブレーション研究. 因果関係が時系列予測の精度に与える影響を定量的に評価するために、分離行列 \mathbf{W} を単位行列に固定した *w/o causality* との比較に基づいてアブレーション研究を行った。表 6 は、人工データおよび実データの両方を用いて行ったアブレーション研究の結果である。*w/o causality* はすべての実験設定において予測精度が低下していることが確認された。したがって、時系列データストリーム内の時間変化する因果関係を抽出することが予測精度を向上させるといえる。

5.3 Q3: 提案手法の計算時間

最後に、提案手法の計算コストについて検証する。図 7 は、MODEPLAIT と比較手法の計算効率を比較したものである。具体的には、図 7 の左図は (#4) *exercise* に対する各時点 t_c における計算コストを、右図は時系列データストリーム全体の計算時間の平均値を示している。これらの図の y 軸は対数スケールで表示している。逐次更新のおかげで、提案手法は比較手法により高速に動作することが可能であり、これは定理 3 の内容と一致している。

6 むすび

本論文では、大規模時系列データストリーム中の時間変化する因果関係の抽出および将来予測を効率的かつ適応的に実現するための時系列解析技術として MODEPLAIT を提案した。MODEPLAIT は冒頭で確認した以下の優れている特性を全て達成している。

- 時系列データストリーム中の時間変化する因果関係という重要な特徴を逐次的に抽出する。
- 評価実験において、提案手法が時間変化する因果関係の抽出および将来予測を正確に行うことを実証した。
- 計算コストはデータストリームの長さに依存することなく、高速に処理が可能である。

謝辞 本研究の一部は大阪大学博士課程教育リーディングプログラム、JST CREST JPMJCR23M3 の助成を受けたものです。

文 献

- [1] Gianmarco De Francisci Morales, Albert Bifet, Latifur Khan, Joao Gama, and Wei Fan. Iot big data stream mining. In *Proceedings of the 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2119–2120, 2016.
- [2] Mohammad Saeid Mahdavejad, Mohammadreza Rezvan, Mohammadamin Barekatin, Peyman Adibi, Payam Barnaghi, and Amit P Sheth. Machine learning for internet of things data analysis: A survey. *Digital Communications and Networks*, 4(3):161–175, 2018.
- [3] Koki Kawabata, Yasuko Matsubara, Takato Honda, and Yasushi Sakurai. Non-linear mining of social activities in tensor streams. In *Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2093–2102, 2020.
- [4] Kota Nakamura, Yasuko Matsubara, Koki Kawabata, Yuhei Umeda, Yuichiro Wada, and Yasushi Sakurai. Fast and multi-aspect mining of complex time-stamped event streams. In *Proceedings of the ACM Web Conference*, pages 1638–1649, 2023.
- [5] Tasuku Kimura, Yasuko Matsubara, Koki Kawabata, and Yasushi Sakurai. Fast mining and forecasting of co-evolving epidemiological data streams. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3157–3167, 2022.
- [6] Yasuko Matsubara, Yasushi Sakurai, Willem G Van Panhuis, and Christos Faloutsos. Funnel: automatic mining of spatially coevolving epidemics. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 105–114, 2014.
- [7] Yasuko Matsubara, Yasushi Sakurai, Christos Faloutsos, Tomoharu Iwata, and Masatoshi Yoshikawa. Fast mining and forecasting of complex time-stamped events. In *Proceedings of the 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 271–279, 2012.
- [8] David Hallac, Sagar Vare, Stephen Boyd, and Jure Leskovec. Toeplitz inverse covariance-based clustering of multivariate time series data. In *Proceedings of the 23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 215–223, 2017.
- [9] Kohei Obata, Koki Kawabata, Yasuko Matsubara, and Yasushi Sakurai. Dynamic multi-network mining of tensor time series. In *Proceedings of the ACM on Web Conference*, pages 4117–4127, 2024.
- [10] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023.

- [11] Dingsu Wang, Yuchen Yan, Ruizhong Qiu, Yada Zhu, Kaiyu Guan, Andrew Margenot, and Hanghang Tong. Networked time series imputation via position-aware graph enhanced variational autoencoders. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2256–2268, 2023.
- [12] Veronica Tozzo, Federico Ciech, Davide Garbarino, and Alessandro Verri. Statistical models coupling allows for complex local multivariate time series analysis. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1593–1603, 2021.
- [13] Kenneth A Bollen. *Structural equations with latent variables*. John Wiley & Sons, 1989.
- [14] Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT press, 2000.
- [15] Yue He, Peng Cui, Zheyang Shen, Renzhe Xu, Furui Liu, and Yong Jiang. Daring: Differentiable causal discovery with residual independence. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 596–605, 2021.
- [16] Daigo Fujiwara, Kazuki Koyama, Keisuke Kiritoshi, Tomomi Okawachi, Tomonori Izumitani, and Shohei Shimizu. Causal discovery for non-stationary non-linear time series data using just-in-time modeling. In *Conference on Causal Learning and Reasoning*, pages 880–894. PMLR, 2023.
- [17] Enyan Dai and Jie Chen. Graph-augmented normalizing flows for anomaly detection of multiple time series. In *The Tenth International Conference on Learning Representations*, 2022.
- [18] Yuxiao Cheng, Runzhao Yang, Tingxiong Xiao, Zongren Li, Jinli Suo, Kunlun He, and Qionghai Dai. Cuts: Neural causal discovery from irregular time-series data. In *The Eleventh International Conference on Learning Representations*, 2023.
- [19] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge university press, 2009.
- [20] <https://github.com/C-Naoki/ModePlait>, Source codes.
- [21] S. Fink. South africa announces a new coronavirus variant. *New York Times*, Dec. 21, 2020.
- [22] C. Buckley. Relief, reunions and some anxiety as shanghai (mostly) reopens. *New York Times*, Jun. 1, 2022.
- [23] H. Ellyatt. U.s. reports over 1 million new daily covid cases as omicron surges. *CNBC*, Jan. 4 2022.
- [24] Naoki Chihara, Yasuko Matsubara, Ren Fujiwara, and Yasushi Sakurai. Modeling time-evolving causality over data streams. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2025.
- [25] George EP Box and Gwilym M Jenkins. *Time series analysis: forecasting and control (Revised Edition)*. John Wiley & Sons, 1976.
- [26] James Durbin and Siem Jan Koopman. *Time series analysis by state space methods*, volume 38. OUP Oxford, 2012.
- [27] Spiros Papadimitriou, Anthony Brockwell, and Christos Faloutsos. Adaptive, hands-off stream mining. In *Proceedings of 29th International Conference on Very Large Data Bases*, pages 560–571. Elsevier, 2003.
- [28] Lei Li, B Aditya Prakash, and Christos Faloutsos. Parsimonious linear fingerprinting for time series. *Proceedings of the VLDB Endowment*, 3(1-2):385–396, 2010.
- [29] Alysha M De Livera, Rob J Hyndman, and Ralph D Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association*, 106(496):1513–1527, 2011.
- [30] Qiquan Shi, Jiaming Yin, Jiajun Cai, Andrzej Cichocki, Tatsuya Yokota, Lei Chen, Mingxuan Yuan, and Jia Zeng. Block hankel tensor arima for multiple short time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5758–5766, 2020.
- [31] Charu C. Aggarwal, editor. *Data Streams - Models and Algorithms*, volume 31 of *Advances in Database Systems*. Springer, 2007.
- [32] Michael Hahsler and Matthew Bolaños. Clustering data streams based on shared density between micro-clusters. *IEEE transactions on knowledge and data engineering*, 28(6):1449–1461, 2016.
- [33] Yasuko Matsubara and Yasushi Sakurai. Regime shifts in streams: Real-time forecasting of co-evolving time sequences. In *Proceedings of the 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1045–1054, 2016.
- [34] Yasuko Matsubara and Yasushi Sakurai. Dynamic modeling and forecasting of time-evolving data streams. In *Proceedings of the 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 458–468, 2019.
- [35] David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020.
- [36] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11106–11115, 2021.
- [37] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.
- [38] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *The Eleventh International Conference on Learning Representations*, 2023.
- [39] Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, Antti Kerminen, and Michael Jordan. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(10), 2006.
- [40] Song Jiang, Zijie Huang, Xiao Luo, and Yizhou Sun. Cf-gode: Continuous-time causal inference for multi-agent dynamical systems. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, page 997–1009, 2023.
- [41] Fangfu Liu, Wenchang Ma, An Zhang, Xiang Wang, Yueqi Duan, and Tat-Seng Chua. Discovering dynamic causal space for dag structure learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1429–1440, 2023.
- [42] Jonathan G Richens, Ciarán M Lee, and Saurabh Johri. Improving the accuracy of medical diagnosis with causal machine learning. *Nature communications*, 11(1):3923, 2020.
- [43] Chenwang Wu, Xiting Wang, Defu Lian, Xing Xie, and Enhong Chen. A causality inspired framework for model interpretation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, page 2731–2741. ACM, 2023.
- [44] Xun Zheng, Bryon Aragam, Pradeep K Ravikumar, and Eric P Xing. Dags with no tears: Continuous optimization for structure learning. In *Advances in neural information processing systems*, volume 31, pages 9492–9503, 2018.
- [45] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.
- [46] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.
- [47] Clive William John Granger and Paul Newbold. *Forecasting economic time series*. Academic press, 2014.

- [48] Floris Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980: proceedings of a symposium held at the University of Warwick 1979/80*, pages 366–381. Springer, 2006.
- [49] Matan Gavish and David L Donoho. The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*, 60(8):5040–5053, 2014.
- [50] Jorge J Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis: proceedings of the biennial Conference held at Dundee, June 28–July 1, 1977*, pages 105–116. Springer, 2006.
- [51] Bin Yang. Projection approximation subspace tracking. *IEEE Transactions on Signal processing*, 43(1):95–107, 1995.
- [52] Simon Haykin. Adaptive filter theory. *Prentice Hall google schola*, 2:67–94, 2002.
- [53] <https://health.google.com/covid-19/open-data/>, Google COVID-19 Open Data Repository.
- [54] <https://trends.google.co.jp/trends/>, Google Trends.
- [55] <http://mocap.cs.cmu.edu/>, CMU motion capture database.
- [56] Yue Yu, Tian Gao, Naiyu Yin, and Qiang Ji. Dags with no curl: An efficient dag structure learning approach. In *Proceedings of the 38th International Conference on Machine Learning*, pages 12156–12166. PMLR, 2021.
- [57] Xun Zheng, Chen Dan, Bryon Aragam, Pradeep Ravikumar, and Eric P. Xing. Learning sparse nonparametric DAGs. In *International Conference on Artificial Intelligence and Statistics*, 2020.
- [58] David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of machine learning research*, 3(Nov):507–554, 2002.