情報処理学会論文誌　データベース　Vol.14　No.2　10–19 (Apr. 2021)

大規模疫病データのための将来予測アルゴリズム

木村 輔[1,a)]　松原 靖子[1]　川畑 光希[1]　櫻井 保志[1]

受付日 2020年9月10日, 採録日 2021年1月4日

概要：本論文では，大規模疫病データのための高速予測手法である EpiCast について述べる．EpiCast は，様々な地域の大規模疫病データストリームが与えられたときに，その中から疫病の特徴を表現，要約，共有し，長期的かつ継続的に将来の感染者数予測を行う．提案手法は (a) 疫病の複雑な拡散過程を非線形モデルで表現し，(b) それらの中に含まれる重要な特徴を各地域で共有し，適切なモデルを選択することで，感染拡大予測を実現する．ここで，提案手法は (c) データストリームの長さに依存せず，一定の計算時間で感染者数を推定する．COVID-19 の実データを用いた実験では，EpiCast が大規模疫病データストリームの中から疫病の重要な特徴を発見，共有することで感染者数を長期的に予測し，さらに，既存手法と比較し大幅な精度，性能向上を達成していることを確認した．

キーワード：疫病，時系列データストリーム，非線形動的システム，将来予測

Real-time Forecasting of Co-evolving Epidemics

Tasuku Kimura[1,a)]　Yasuko Matsubara[1]　Koki Kawabata[1]　Yasushi Sakurai[1]

Received: September 10, 2020, Accepted: January 4, 2021

Abstract: Given a large collection of co-evolving epidemics, how can we forecast their future characteristics? In this paper, we propose a streaming algorithm, EpiCast, which is able to model, understand and forecast future epidemic outbreaks as well as pandemics. Our method has the following features for the effective and efficient modeling of the dynamics of spreading viruses. (a) *Non-linear*: we incorporate a non-linear equation that is suitable for complex epidemic modeling. (b) *Dynamic*: it maintains multiple such non-linear models to share important patterns among locations, and chooses the non-linear model for the forecast while monitoring a co-evolving epidemic data stream. (c) *Scalable*: it can quickly forecast future phenomena at any time in a practically constant time. In extensive experiments using real COVID-19 datasets over major countries, we demonstrate that our proposed method outperforms existing methods for time series in terms of forecasting accuracy, and significantly reduces the required computational time.

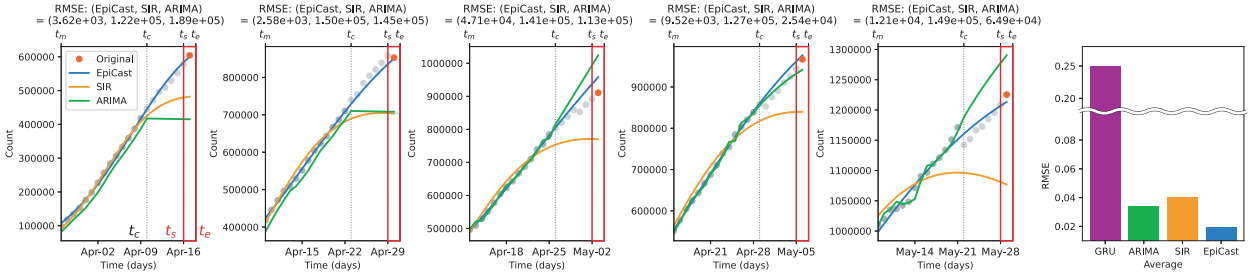
Keywords: epidemics, time series, non-linear dynamical systems, real-time forecasting

1. まえがき

新型ウイルスである COVID-19 の感染は世界中で拡大しており [27]，働き方やコミュニケーションの取り方など，人々の生活に強い影響を与えている [3], [16]．将来起こりうるアウトブレイク（感染爆発）やパンデミック（世界的大流行）に対して，政府や人々が最良の意思決定を下すためには，疫病の感染状況について，より正確な予測を得ることが重要である．もし，今後の感染者数を把握できれば，通勤などの頻繁な社会活動を抑制したり，病床の稼働率をコントロールすることで，パンデミックのリスクを事前に管理することが可能となる [1]．感染状況を正確に把握するためには，信頼性が高い検査を迅速かつ大規模に実施することが重要である．各国は感染者情報管理システム[*1] を運用することで，疫病の統計情報を把握する環境整備を進めているが，現在の主な検査手法である PCR 法の検査時

[1] 大阪大学産業科学研究所産業科学 AI センター
ISIR, Osaka University, Ibaraki, Osaka 567–0047, Japan
[a)] tasuku@sanken.osaka-u.ac.jp

[*1] https://protect-public.hhs.gov/,
https://www.rki.de/EN/Home/homepage_node.html,
https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000121431_00129.html



(a) 提案手法と既存手法による 7 日先の感染者数予測結果のスナップショット　　(b) 予測値の平均エラー値

図 1　アメリカ合衆国における COVID-19 の感染者数に対する 7 日先の (a) 予測結果と (b) 予測精度の比較

Fig. 1　The seven-days-ahead forecasting results of EpiCast and its competitors over the number of COVID-19 cases in the United States.

間が長く，また，検査できる場所が限られるため，1 日あたりの検査数に限界が生じている．一方，PCR 法に変わる新規の検査手法の開発を，多くの企業や研究機関が進めている．特に，近年の医療 ICT 化と連動し，検査時間が短く，誰でもどこでも検査できる，スマートデバイスと連携した検査キット[*2]の開発を進める企業が多く存在する．これらの新規検査キットの開発・導入と疫病情報の管理・把握システムが組み合わさることで，大規模かつリアルタイム性の高い疫病データセットの構築環境が整うことが予想される．

強力な感染力を持つ COVID-19 が引き起こす，さらなるパンデミックを回避するには，感染の拡大や収束をすでに経験した地域から，他の地域でも適用可能な形式で疫病の特徴を抽出し，蓄積することで，拡散過程に関する知識を互いに共有し，活用することが重要である．さらに，アウトブレイクを抑制するために施行された対策によって，疫病のダイナミクスは，時間の経過とともに変化する可能性がある [26]．よって，拡散過程の動的な変化をとらえ，今後の感染状況を推定可能なストリーミング手法の確立は，きわめて重要な研究課題である．

本論文では，大規模疫病データのための将来予測アルゴリズムである EpiCast について述べる．EpiCast は，疫病テンソルストリームに潜む疫病の振舞いを非線形モデルで表現し，それらの中に含まれる重要な特徴を各地域で共有することで，複雑な拡散過程を柔軟に表現する．

より具体的には，次の問題を扱う．

現在時刻 t_c において，r カ所の場所で観測された d 次元の属性で構成される疫病テンソルストリーム $\mathcal{X} = \{x_{tij}\}_{t,i,j=1}^{t_c,r,d}$ が与えられたとき，l_s ステップ先の未知の疫病テンソルストリームを継続的に予測する．

[*2] https://www.sanofi.com/en/media-room/press-releases/2020/2020-04-16-14-00-00,
https://attheu.utah.edu/facultystaff/portable-test-for-covid-19/,
https://coughvid.epfl.ch/

1.1　具体例

図 1 (a) は，5 つのスナップショットを含むアメリカ合衆国における COVID-19 の感染者数のダイナミクスを示している．灰色の丸印は，感染者数の実データを示しており，赤色の丸印は，現在時刻 t_c から 7 日先で観測される感染者数を表している．ここで，黒色の点線は現在時刻 t_c を，赤色の矩形は，7 日先の予測ウィンドウ $[t_s : t_e]$ を示している．最後に，各実線は，3 種類の手法で推定した値を示している．青色の線は，提案手法の EpiCast，オレンジ色と緑色の線は，比較手法の SIR と ARIMA である．図 1 (a) から分かるように，提案手法が疫病の拡散過程をとらえることに成功しているのに対し，どの比較手法もそのダイナミクスを上手くとらえられていない．線形モデルの ARIMA は，疫病の持つ複雑なパターンを表現することが困難なため，アウトブレイクやパンデミックの予測には効果的ではない．一方，他の比較手法である SIR は，疫病の時系列を想定した非線形モデルであるものの，疫病パターンの動的変化には対応できなかったために，モデルパラメータの学習が不十分となり，ARIMA と同様の結果となった．

図 1 (b) は，予測結果の評価指標である，7 日先の実データと予測値との二乗平均誤差（RMSE: root mean square error）を示している．提案手法 EpiCast と，既存手法である GRU，ARIMA および SIR を比較すると，提案手法は，既存手法と比較して予測誤差を大幅に改善した．これは EpiCast が，4 章で説明する，ストリーミング手法による感染予測を実現するうえで望ましい性質である，(**P1**) 複雑な疫病テンソルストリームの非線形モデルリングと (**P2**) 複数の地域間におけるモデル共有の仕組みを持つためである．

1.2　本論文の貢献

本論文では，大規模疫病データのための将来予測アルゴリズムである EpiCast を提案する．EpiCast は以下の特長を持つ．

(1) 非線形方程式に基づいた疫病モデルを構築し，疫病の複雑な拡散過程をモデルパラメータとして抽出する．
(2) 各地域の疫病モデルを共有し，適切なモデルを選択することで，感染拡大予測を実現する．
(3) 計算コストはデータストリームの長さに依存しない．

2. 関連研究

この章では，データマイニングと統計疫学の関連文献を，時系列解析とデータストリームマイニングの 2 つの観点から紹介する．

時系列のモデル化と予測は，重要なトピックである [19]．既存手法としては，自己回帰モデル（AR: autoregressive model）およびカルマンフィルタ（KF: Kalman filters）などがあり [7]，これらの手法は様々に拡張されてきた [18], [20], [21]．データセットの仮定が少ない，より一般的な問題設定のための非線形モデル [13], [23] に加えて，適切な非線形微分方程式を選択することで，最近のデータや過去のデータでは観測されていない複雑なダイナミクスの予測を可能にするドメイン知識をモデルに適用することができる [14], [22]．

Deep Neural Network（DNN）は，入力データから高次元の時間領域の特徴を取得し，様々なコンテキストやドメインにおける将来のイベントを予測する新たな手法としてさかんに研究されている [5], [8], [17], [28], [29]．最新手法である EpiDeep [2] は，インフルエンザのデータセットにおいて，季節的に発生する感染症の拡散過程のモデル化に，DNN を適用することに成功した．しかし，これら DNN ベースの手法は，数多くのモデルパラメータを推定するために，膨大な量の学習データを必要とする．そのため，これまでのウイルスとは拡散過程が大きく異なり，かつ，学習データが少ない新型ウイルスのモデル化は困難である．

計算時間や使用メモリの制約の下で，大量のデータを処理・解析するオンラインアルゴリズムやストリーミングアルゴリズムの重要性が高まっている [4], [10], [11], [12]．一方で，新型ウイルスの出現により，パンデミックのピーク時期の予測を重要な目標とする，リアルタイム性の高い予測も必要とされている [24], [25]．しかし残念なことに，これら既存手法による疫病モデルの学習において，そのスケーラビリティについては議論されていない．

まとめると，いずれの既存手法も，疫病テンソルストリームの非線形な振舞いのモデル化，複数の地域間の拡散過程の共有および高速な感染者数予測に対応していない．本論文ではこれらの要件を満たすことで，COVID-19 のような新型ウイルスの拡散過程を予測するためのストリーミングアルゴリズムを提案する．

3. 提案モデル

本章では，疫病テンソルストリームのためのモデルを提案する．まず，具体的な問題設定について述べ，その後，提案モデルについて詳細に説明する．

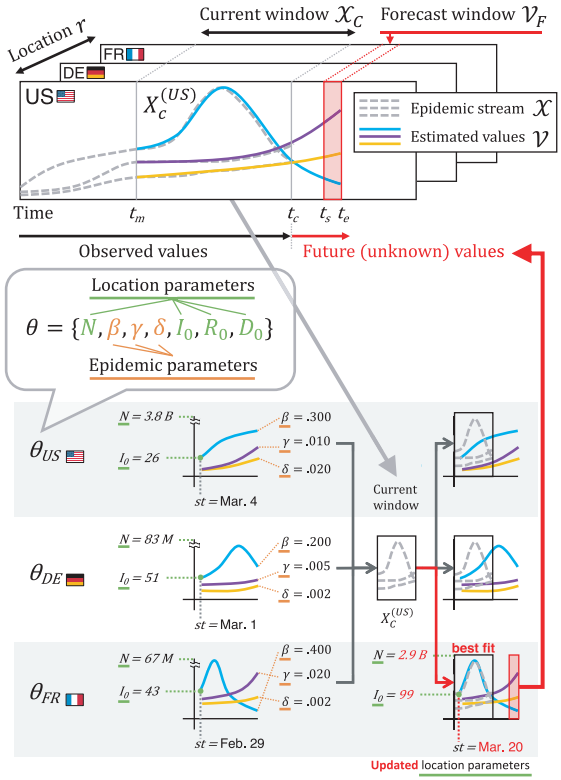


図 2　EpiCast の様子：提案手法は，オリジナルの疫病テンソルストリーム $\mathcal{X}$（灰点線）が与えられたとき，共有する様々な地域の疫病モデルから最良のモデルを選択し，カレントウィンドウの時系列パターン（色太線）を推定することで，l_s ステップ先の予測ウィンドウ（赤矩形内）を高速かつ継続的に出力する

Fig. 2　Illustration of EpiCast: Given an epidemic streams $\mathcal{X}$, it incrementally maintains the current window $[t_m : t_c]$ in each i-th location (i.e., $\mathcal{X}_C = \{X_C^{(i)}\}_{i=1}^{r}$), captures co-evolving epidemic patterns with shared non-linear models among r locations/countries, and then forecasts the l_s-steps-ahead future window $[t_s : t_e]$.

現在時刻 t_c までに観測された，r カ所の地域における，d 次元の疫病データから構成される $\mathcal{X} \in \mathbb{N}^{t_c \times r \times d}$ を，疫病テンソルストリームとする．本論文では，1 日の感染者数，回復者数および死亡者数に対応するために次元数を $d = 3$ とした．したがって，$\mathcal{X}$ の要素 $\{x_{tij}\}_{t=1,i=1,j=1}^{t_c,r,d}$ は，時刻 t において i 番目の地域で観測された j 番目の属性を表す．毎時刻において新たに $X_{t_c+1} \in \mathbb{N}^{r \times d}$ が観測され，$\mathcal{X}$ の総量は増加する．ストリーミング手法では，処理速度が重要であるため，計算時間と使用メモリが小さくなるように制約され，多くの場合，これまで観測したデータの一部のみから将来を予測する必要がある．よって，**図 2** の上部に示すように，最新の $\mathcal{X}$ のみを用いて予測するための短い時間窓を定義する．より具体的には，疫病テンソルストリームの最新の時系列を含むカレントウィンドウを $\mathcal{X}_C$ と定義する．同様に，$\mathcal{V}_F$ は，推定値を生成したい期間に対応する

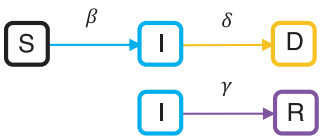


図 3　EpiCast-base の状態遷移の様子
Fig. 3　EpiCast-base diagrams.

予測ウィンドウである．本論文で取り組む問題を以下のように定義する．

問題 1　長さ l_c のカレントウィンドウ $\mathcal{X}_C = \{X_t\}_{t=t_m}^{t_c}$ が与えられたとき，l_s ステップ先の予測ウィンドウ $\mathcal{V}_F = \{V_t\}_{t=t_s}^{t_e}$ を予測する．

3.1　EpiCast—単一地域の疫病ストリームの場合

1 章の COVID-19 の疫病ストリームのスナップショットから確認できるように，疫病は，非線形に拡散する傾向がある．まず解決するべき問題は，アウトブレイクからパンデミックまでの複雑なダイナミクスをモデル化する方法を見つけることである．ここで，単純化のために，疫病テンソルストリームの単一の地域にのみ焦点を当て，SIR ベースのモデル EpiCast-base を提案する．提案モデルは，次の 4 つの状態から構成される．

- Susceptible（感受性保持者）：時間の経過によって疫病を患う可能性のある状態
- Infected（感染者）：感染率 β で疫病に感染した状態
- Recover（回復者）：回復率 γ で感染から回復した状態
- Die（死亡者）：死亡率 δ で感染から死亡した状態

これら 4 つのクラスの時間依存性は，次の式で表される．

$$\begin{aligned}\frac{dS}{dt} &= -\beta S(t)I(t),\\ \frac{dI}{dt} &= \beta S(t)I(t) - \gamma I(t) - \delta I(t),\\ \frac{dR}{dt} &= \gamma I(t), \quad \frac{dD}{dt} = \delta I(t).\end{aligned} \tag{1}$$

図 3 は，EpiCast-base の状態遷移の様子を示す．ここで，モデルは，各推定値を生成するために，感染者数，回復者数および死亡者数について，初期値 I_0，R_0，D_0 をそれぞれ必要とする．また，総人口の概念に従って，初期の感受性集団 S_0 は，疫病の潜在的人口 N を用いて，$S_0 = N - (I_0 + R_0 + D_0)$ と計算することで求めることができる．

まとめると，推定したいパラメータ集合全体は以下のようになる．

モデル 1（EpiCast-base）　EpiCast-base のパラメータ集合を $\boldsymbol{\theta} = \{N, I_0, R_0, D_0, \beta, \gamma, \delta\}$ とする．各要素は以下のように定義される．

- N：疫病の潜在的人口（$0 \le N$）
- I_0：感受性保持者の初期人口（$0 \le I_0$）
- R_0：回復者の初期人口（$0 \le R_0$）
- D_0：死亡者の初期人口（$0 \le D_0$）
- β：疫病の感染率（$0 \le \beta \le 1$）
- γ：疫病の回復率（$0 \le \gamma \le 1$）
- δ：疫病の死亡率（$0 \le \delta \le 1$）

ここで，モデル 1 が生成する記号 $\mathcal{V} \in \mathbb{N}^{l_e \times r \times d}$, $V \in \mathbb{N}^{l_e \times d}$ および $\mathbf{v} \in \mathbb{N}^d$ を使用する．たとえば，i 番目の地域のウィンドウ $V^{(i)} \subset \mathcal{V}$ は，モデル 1 を用いて，各初期値 I_0，R_0 および D_0 から値を推定し続けることで生成される．

3.2　EpiCast—複数地域の疫病ストリームの場合

提案手法においてより重要な目標は，異なる時刻において発生する可能性のある地域間の類似するダイナミクスを特定することである．これは各地域が，感染の初期段階のような共通する感染傾向を，潜在的に保持していると考えられるためである．単一の地域でモデルを推定するための十分な観測値が得られない場合には，別の地域で得られたモデルを予測に適用すべきであるというのが，本手法が持つべき最も重要な性質である．

そこで本研究では，地域ごとに人口や文化のような特徴に違いがあることを考慮し，$\boldsymbol{\theta}$ のパラメータを，地域パラメータ $\boldsymbol{\theta_L}$ および疫病パラメータ $\boldsymbol{\theta_E}$ の 2 つのグループに分ける．ここで，$\boldsymbol{\theta} = \boldsymbol{\theta_L} \cup \boldsymbol{\theta_E}$ である．疫病パラメータ $\boldsymbol{\theta_E}$ は，どの地域でも共有可能であるのに対し，地域パラメータ $\boldsymbol{\theta_L}$ は，地域ごとに最適化されている．たとえば，異なる地域で同じ感染対策が講じられていても，それを施行するタイミングが地域によって異なる場合，ダイナミクスが変化する時期は異なるが，共通のダイナミクスを保持する可能性が高い．したがって，図 2 で示すように，複数の EpiCast-base のパラメータ集合を保持し，共有することで，各地域で得られた知識を，他の地域の予測に反映させることが可能となる．最終的に，EpiCast-full のパラメータ集合を以下のように定義する．

モデル 2（EpiCast-full）　g 種類の疫病モデルから構成される EpiCast-full のパラメータ集合を $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_g\}$ とする．これは疫病テンソルストリーム内のすべての非線形ダイナミクスを記述しており，モデル $\boldsymbol{\theta}$ は，2 つのパラメータグループから構成されている．

- 地域パラメータ：$\boldsymbol{\theta_L} = \{N, I_0, R_0, D_0\}$
- 疫病パラメータ：$\boldsymbol{\theta_E} = \{\beta, \gamma, \delta\}$

4.　アルゴリズム

これまで，疫病テンソルストリームにおける非線形の拡散過程をモデル化する方法を述べた．本章では，複数の疫病モデルを時間発展とともに獲得するストリーミングアルゴリズム EpiCast を提案する．本アルゴリズムは，以下に示す 2 つの重要な特性を満たす．

(**P1**)　複雑な疫病テンソルストリームの非線形モデリング
(**P2**)　複数の地域間におけるモデル共有の仕組み

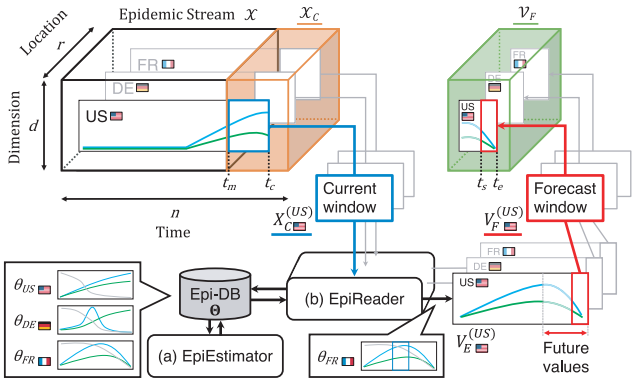


図 4　EpiCast のアルゴリズム概要

Fig. 4　Overview of the EpiCast algorithm.

まず，非線形に振る舞う感染データにおける，複雑なダイナミクスをとらえる必要がある．(**P1**) に対して，非線形微分方程式を利用したストリーミングアルゴリズム（EpiEstimator）を提案する．また，(**P2**) に対して，地域間の類似したダイナミクスを検出し，別の地域で得られた疫病モデルを，他の地域の予測に利用できるアルゴリズム（EpiReader）を提案する．図 4 は，以下の 3 つのアルゴリズムで構成される EpiCast の概要を示している．

- EpiEstimator：i 番目の地域において，$\boldsymbol{\Theta}$ 内の各モデル $\boldsymbol{\theta}$ とは異なるダイナミクスが観測されたとき，新たなモデル $\boldsymbol{\theta}$ を推定し，そのモデルから，推定ウィンドウ $V_E^{(i)} = V^{(i)}[t_m : t_e]$ を生成する（Algorithm 1）．
- EpiReader：i 番目の地域のカレントウィンドウ $X_C^{(i)}$ およびモデルパラメータ集合 $\boldsymbol{\Theta}$ が与えられたとき，最良のモデル $\boldsymbol{\theta}$ を推定し，推定ウィンドウ $V_E^{(i)}$ を生成する（Algorithm 2）．
- EpiCast：各地域 i における最適なモデル $\boldsymbol{\theta}$ から推定ウィンドウ $\mathcal{V}_E = \{V_E^{(1)}, \ldots, V_E^{(r)}\}$ を生成する．その後，l_s ステップ先の予測ウィンドウ $\mathcal{V}_F = \mathcal{V}[t_s : t_e]$ を報告する．さらに，モデルパラメータ集合 $\boldsymbol{\Theta}$ を更新する（Algorithm 3）．

4.1　モデル推定—EpiEstimator

議論の単純化のため，まずは，疫病ストリーム上の非線形モデルを効果的に推定する方法を説明する．具体的には，疫病テンソルストリーム $\mathcal{X}$ 内の単一の地域に焦点を当て，カレントウィンドウ X_C のモデルパラメータ $\boldsymbol{\theta}$ を推定できる EpiEstimator を提案する．Algorithm 1 は，EpiEstimator の処理の流れを示している．$X_C \in \mathbb{N}^{l_c \times d}$ を単一の地域のカレントウィンドウとし，$X_C^{(i)} \subset \mathcal{X}_C$ $(i = 1, \ldots, r)$ を i 番目の地域のカレントウィンドウとする．EpiEstimator は，$X_C^{(i)}$ が与えられたとき，次の式を最小化することで，最適な $\boldsymbol{\theta}$ を発見する．

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}'} ||X_C^{(i)} - V_C^{(i)}||, \quad V_C^{(i)} = f(\boldsymbol{\theta}'), \tag{2}$$

Algorithm 1 EpiEstimator(X_C)

Input: Current epidemic stream $X_C = \{\boldsymbol{x}_t\}_{t=t_m}^{t_c}$
Output: Estimated values $V_E = \{\mathbf{v}_t\}_{t=t_m}^{t_e}$ and Model parameter set $\boldsymbol{\theta} = \{N, I_0, R_0, D_0, \beta, \gamma, \delta\}$
$\boldsymbol{\Theta}_C \leftarrow \emptyset$; // Candidate parameter set
for $t'_m = t_m : t_c$ **do**
　$X'_C = X_C[t'_m : t_c]$;
　$\boldsymbol{\theta} \leftarrow \arg\min_{\boldsymbol{\theta}'} ||X'_C - V_C'||$; // $V_C' = f(\boldsymbol{\theta}')$
　$\boldsymbol{\Theta}_C \leftarrow \boldsymbol{\Theta}_C \cup \boldsymbol{\theta}$;
end for
/* Choose the best model */
$\boldsymbol{\theta} \leftarrow \arg\min_{\boldsymbol{\theta}' \in \boldsymbol{\Theta}_C} ||X_C - V_C'||$; // $V_C' = f(\boldsymbol{\theta}')$
Compute V_E using $\boldsymbol{\theta}$; // $V_E = f(\boldsymbol{\theta})$
return $\{V_E, \boldsymbol{\theta}\}$;

ここで，$||\cdot||$ は平均二乗誤差を表し，$V_C^{(i)} = f(\boldsymbol{\theta}')$ は $\boldsymbol{\theta}'$ を入力として式 (1) で生成された $X_C^{(i)}$ と同じ期間を持つウィンドウを表す．最適な $\boldsymbol{\theta}$ を得るためには，非線形最小二乗問題を解く必要がある．$\boldsymbol{\theta}$ のすべてのパラメータは，非線形性を有する学習に適した Levenberg-Marquardt（LM）アルゴリズム [15] によって最適化する．また，予測ウィンドウの各次元の値は，4 次のルンゲ・クッタ法 [9] に基づき生成する．

次に，疫病ストリームにおいて，各地域のアウトブレイクはランダムに発生する．そのため，カレントウィンドウの初期値 $\{I_0, R_0, D_0\} \in \boldsymbol{\theta}$ を推定する際に，適切な感染の開始時刻を選択することは重要である．なぜならば，この開始時刻 t'_m の選択が，モデルの品質に影響を与えるためである．

そこで本研究では，最適な開始時刻 $t'_m \in [t_m : t_c]$ を探索する．t'_m を $t_m, t_m+1, t_m+2, \ldots$ と変化させながら $X[t'_m : t_c]$ を用いて新しいパラメータ $\boldsymbol{\theta}$ を推定し，$t_c - t'_m + 1$ 個の候補モデル $\boldsymbol{\theta}$ を取得する．$X_C^{(i)}$ と $V_C^{(i)}$ の二乗誤差にしたがって，$X_C^{(i)}$ に最適なモデルを選択する．この処理において，探索の候補となる時刻のうち，$t'_m = \{t \mid X_t = 0\}$ となる範囲では，時刻 t において，アウトブレイクが発生しないことが保証されているので，開始時刻 t'_m を探索する範囲から排除することができる．これにより，より効率的に $\boldsymbol{\theta}$ を推定することが可能となる．

4.2　モデル選択—EpiReader

カレントウィンドウ $\mathcal{X}_C$ のダイナミクスが，時間と場所によって変化する疫病テンソルストリームでは，単一の地域における疫病の振舞いに関する情報が著しく不足しているため，任意の地域から得られる複数の非線形モデルを利用する必要がある．そこで本研究では，既存モデルのパラメータ集合 $\boldsymbol{\Theta}$ からカレントウィンドウ $\mathcal{X}_C$ に最適なモデル $\boldsymbol{\theta}$ を選択する EpiReader を提案する．Algorithm 2

Algorithm 2 EpiReader(X_C, Θ)

Input: Current epidemic stream $X_C = \{\boldsymbol{x}_t\}_{t=t_m}^{t_c}$ and Current full parameter set $\Theta = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_g\}$
Output: Estimated values $V_E = \{\mathbf{v}_t\}_{t=t_m}^{t_e}$ and Model parameter set $\boldsymbol{\theta} = \{N, I_0, R_0, D_0, \beta, \gamma, \delta\}$
$\Theta_C \leftarrow \emptyset$; // Candidate parameter set
for $\boldsymbol{\theta}$ in Θ **do**
  for $t'_m = t_m : t_c$ **do**
    $X'_C = X_C[t'_m : t_c]$; $\boldsymbol{\theta_E} \subset \boldsymbol{\theta}$;
    /* Estimate only location parameters */
    $\boldsymbol{\theta_L} \leftarrow \arg\min_{\boldsymbol{\theta'_L}} ||X'_C - V_C'||$; // $V_C' = f(\boldsymbol{\theta'_L}, \boldsymbol{\theta_E})$
    /* Estimate full parameters */
    $\boldsymbol{\theta}' \leftarrow \arg\min_{\boldsymbol{\theta}''} ||X'_C - V_C'||$; // $V_C' = f(\boldsymbol{\theta}''|\boldsymbol{\theta_L}, \boldsymbol{\theta_E})$
    $\Theta_C \leftarrow \Theta_C \cup \boldsymbol{\theta}'$;
  end for
end for
/* Choose the best model */
$\boldsymbol{\theta} \leftarrow \arg\min_{\boldsymbol{\theta}' \in \Theta_C} ||X_C - V_C||$; // $V_C = f(\boldsymbol{\theta}')$
Compute V_E using $\boldsymbol{\theta}$; // $V_E = f(\boldsymbol{\theta})$
return $\{V_E, \boldsymbol{\theta}\}$;

Algorithm 3 EpiCast($\mathcal{X}_C, \Theta$)

Input: Current epidemic streams $\mathcal{X}_C = \{X_t\}_{t=t_m}^{t_c}$ and Model parameter set $\Theta = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_g\}$
Output: l_s-steps-ahead values $\mathcal{V}_F$ and Updated model parameter set $\Theta' = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{g'}\}$
for $i = 1 : r$ **do**
  /* (1) Extract current window at i-the location */
  $X_C^{(i)} = \{\boldsymbol{x}_{ti}\}_{t=t_m}^{t_c}$;
  /* (2) Parameter fitting for local epidemics */
  $\{V_E^{(i)}, \boldsymbol{\theta}\} \leftarrow$EpiReader($X_C^{(i)}, \Theta$);
  /* (3) Estimate new regimes (if required) */
  $V_C^{(i)} = V^{(i)}[t_m : t_c]$; // Estimated values from t_m to t_c
  if $||X_C^{(i)} - V_C^{(i)}|| > \epsilon$ **then**
    $\{V_E^{(i)}, \boldsymbol{\theta}'\} \leftarrow$EpiEstimator($X_C^{(i)}, \Theta$);
    $\Theta' \leftarrow \Theta \cup \boldsymbol{\theta}'$; $g' = g + 1$;
  end if
end for
/* (4) l_s-steps-ahead future value generation */
$\mathcal{V}_E = \mathcal{V}[t_m : t_c]$; $\mathcal{V}_F = \mathcal{V}[t_s : t_e]$;
return $\{\mathcal{V}_F, \Theta'\}$;

は，EpiReader の詳細を示している．

まず，カレントウィンドウ $X_C \subset \mathcal{X}_C$ および既存モデルのパラメータ集合 Θ が与えられた場合を考える．より具体的に，i 番目の地域のための $X_C^{(i)} \subset \mathcal{X}_C$ をカレントウィンドウとする．$X_C^{(i)}$ について，アルゴリズムは，$X_C^{(i)}$ および $\boldsymbol{\theta}$ で生成される $V_C^{(i)}$ の二乗誤差に従って最適なモデル $\boldsymbol{\theta}$ を選択する．しかし，このステップにおいて，i 番目の地域から推定されたモデルではなく，別の地域で推定された異なる人口規模などを持つモデル $\boldsymbol{\theta}$ を使用する可能性がある．そのような場合，他の地域のモデルを適用するには効果的な微調整が必要である．

そこで本研究では，まず疫病に関するパラメータ $\boldsymbol{\theta_E} = \{\beta, \gamma, \delta\}$ を固定した状態で，地域に関するパラメータ $\boldsymbol{\theta_L} = \{N, I_0, R_0, D_0\}$ のみを推定することで，$\boldsymbol{\theta}$ 全体が推定対象の地域に過適合しないようにする．具体的には，次の目的関数を最小化する．

$$\boldsymbol{\theta_L} = \arg\min_{\boldsymbol{\theta'_L}} ||X_C^{(i)} - V_C^{(i)}||, \quad V_C^{(i)} = f(\boldsymbol{\theta'_L}, \boldsymbol{\theta_E}), \ (3)$$

ここで，$\boldsymbol{\theta_E} \subset \boldsymbol{\theta}$ であり，また，$V_C^{(i)} = f(\boldsymbol{\theta'_L}, \boldsymbol{\theta_E})$ は，更新された $\boldsymbol{\theta'_L}$ および固定された $\boldsymbol{\theta_E}$ を用いて，式 (1) によって計算される．部分最適化後，$\boldsymbol{\theta}$ 内の全パラメータを同時に更新するために式 (2) を最小化する．これによって，初期値から推定する場合と比較して，非線形パラメータを効果的に収束することができる．なお，EpiEstimator と同様の方法で，各候補の初期時刻について推定を反復する．

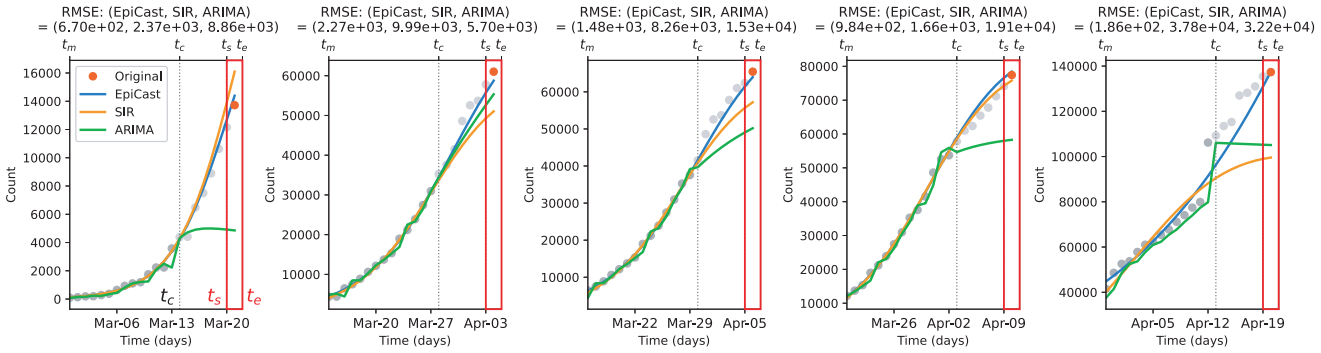
4.3 ストリームアルゴリズム—EpiCast

我々の最終目的は，様々な地域で発生した疫病モデル $\Theta = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_g\}$ をとらえ，すべての地域の予測ウィンドウ $\mathcal{V}_F = \{V_F^{(1)}, \ldots, V_F^{(r)}\}$ を推定することである．そこで，時間発展する疫病の非線形ダイナミクスを考慮した，疫病テンソルストリームの高速な予測を実現するストリーミング手法 EpiCast を提案する．Algorithm 3 は，EpiCast の手順をまとめたものである．カレントウィンドウ $\mathcal{X}_C$ と Θ が与えられたとき，EpiEstimator および EpiReader を組み合わせることで，最良の非線形モデルを決定し，予測ウィンドウ $\mathcal{V}_F$ を生成する．ここでは，2 つのアルゴリズムを組み合わせるうえで重要な最適なパラメータの推定手順および Θ の更新ルールについて説明する．i 番目の地域について，アルゴリズムは $X_C^{(i)} \subset \mathcal{X}_C$ を抽出し，EpiReader を用いて最適なモデル $\boldsymbol{\theta}$ を探索することで，時刻 t_m から t_e までの $V_E^{(i)}$ を得ることができる．もし Θ に最適なモデルが存在しないとき，アルゴリズムは，EpiEstimator を用いて新たなモデル $\boldsymbol{\theta}'$ を推定する必要がある．アルゴリズムは，$X_C^{(i)}$ と $V_C^{(i)}$ の誤差が，ϵ[*3]以上のとき，EpiEstimator を実行する．新たなモデルは，Θ に追加され，$g' = g + 1$ となる．この手順を各地域に対して繰り返し行うことで，最終的に $\mathcal{V}_F$ を得ることができる．

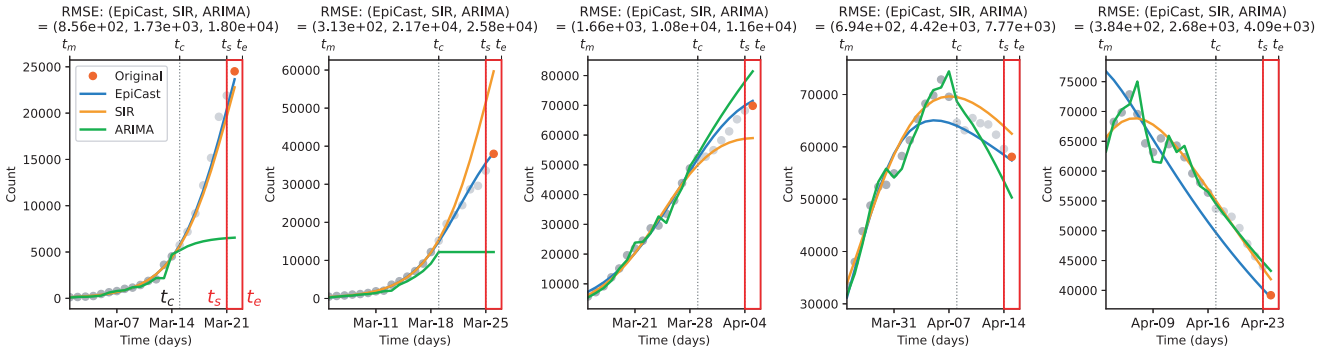
補助定理 1　各時刻における EpiCast の計算時間は $O(g)$ となる．

証明 1　EpiCast のアルゴリズムにおいて，EpiReader

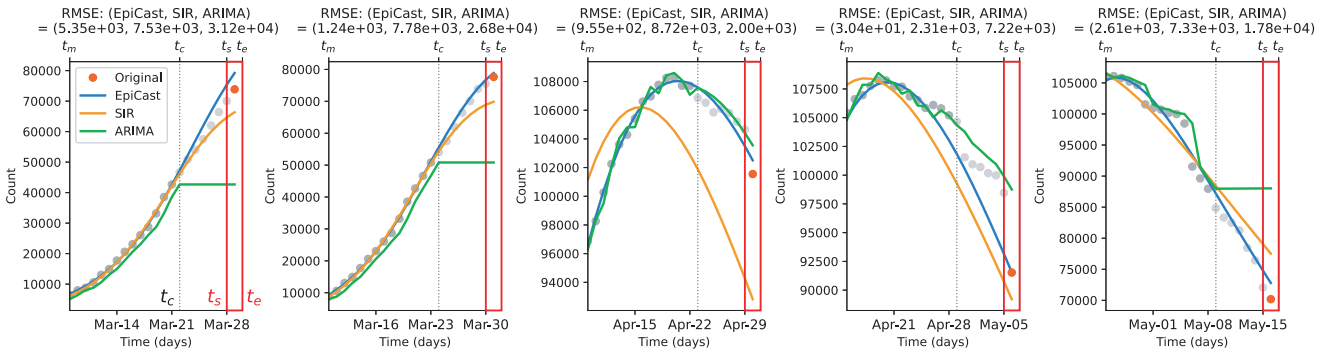
[*3] 本論文では，$\epsilon = 1/2||X_C^{(i)}||$ とする．



(a) フランスにおける 7 日先の感染者数の予測結果



(b) ドイツにおける 7 日先の感染者数の予測結果



(c) イタリアにおける 7 日先の感染者数の予測結果

図 5　COVID-19 の感染者数に対する提案手法および比較手法の 7 日先の予測結果

Fig. 5　The seven-days-ahead forecasting results of EpiCast and its competitors over the number of COVID-19 cases in major countries.

および EpiEstimator は，$\Theta = \{\theta_1, \ldots, \theta_g\}$ のパラメータの推定に $O(g \cdot l_c \cdot r \cdot d)$ の計算時間を要し，最大で l_c 回繰り返される．l_c, r, d によって与えられるカレントウィンドウのサイズは無視できるほど小さい定数値であるので，EpiCast の計算時間は，時間点あたりの $O(g)$ 時間である．

5.　評価実験

本論文では EpiCast の有効性を検証するため，COVID-19 の実データ [6] を用いた実験を行った．本データセットは，200 日にもわたる疫病データの 3 次元のベクトル（感染者数，回復者数および死者数）で構成される[*4]．本章では，以下の項目について検証する．

Q1 疫病ストリームの予測に対する提案手法の有効性

Q2 疫病の拡散過程の予測に対する提案手法の精度の検証

Q3 疫病ストリームの予測に対する計算時間の検証

実験は，16 GB のメモリ，Intel Core i7 2.8 GHz quad core の CPU を搭載したマシン上で実施した．

5.1　Q1. 提案手法の有効性

本節では，疫病テンソルストリームに対する EpiCast の予測能力を検証する．

図 1，図 5 は，実際の COVID-19 データストリームに対する提案手法および既存手法の予測結果である．ここで，各国の人口および感染ステージ（拡大期，増加期および減少期）はそれぞれ異なる．灰色の丸印はオリジナルデータを，青色の実線は EpiCast の予測結果を，オレンジ色と緑色の実線は比較手法である ARIMA および SIR の予測結

[*4] https://github.com/CSSEGISandData/COVID-19

果をそれぞれ示している．また，赤色の丸印は，各時刻 t_c の観測から 7 日後の感染者数を表している．ここで，GRU の結果は，エラー値がきわめて高いため省略した．

すでに 1 章の図 1 においても示したように，EpiCast は，長期的な将来の感染者数の効果的な予測に成功している．図 5 (a) および図 1 は，フランスおよびアメリカ合衆国の 5 つの異なる時期における，7 日先の感染者数をそれぞれ示している．この 2 つの国は，感染者数の増加傾向は同じであるが，人口規模が大きく異なる．図から分かるように，EpiCast は，非線形モデルを共有し，各国の人口に関連するパラメータ $\boldsymbol{\theta_L}$ を更新することで，効果的な予測を実現している．同様に，図 5 (b) および (c) から分かるように，EpiCast は，複数の感染段階から構成されるドイツおよびイタリアの感染者数についても予測できることが分かる．この結果は，EpiCast が，感染症の発生・増加・減少のすべての段階を含む，将来の感染者数を正確にとらえる能力を有することを示す．提案手法の結果と比較して，既存手法は，非線形のダイナミクスをとらえるのには，不向きであることが確認できる．線形モデルである ARIMA は，感染数の急激な変化を予測することができていない．また SIR は，疫病の時系列をモデル化する非線形手法であるものの，疫病の振舞いの動的な変化には対応できていない．

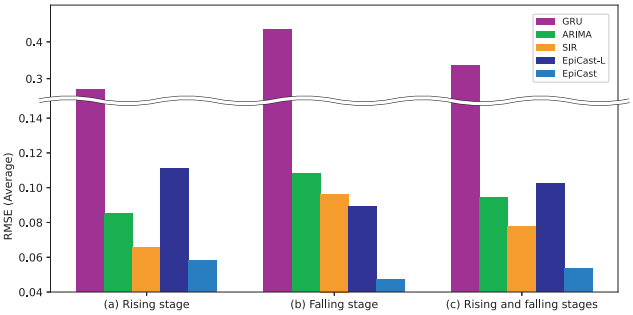


図 6　COVID-19 の感染者数に対する 7 日先の予測精度の比較
Fig. 6　Forecasting error (RMSE) between original and forecast values of EpiCast and its competitors.

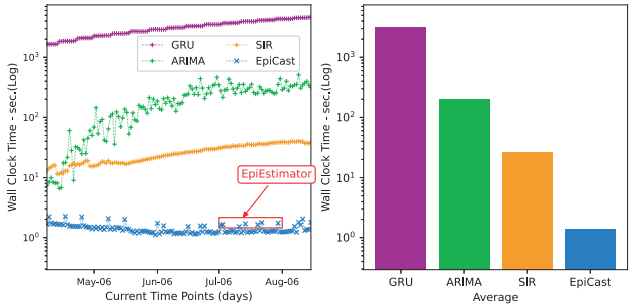


図 7　EpiCast および比較手法の (a) 各時刻における計算時間と (b) 平均計算時間
Fig. 7　(a) Wall clock time vs. data stream length t_c and (b) average time consumption.

5.2　Q2. 提案手法の精度

次に，本論文では，EpiCast の予測精度を検証するため，既存手法である線形の時系列予測手法である (a) ARIMA，一般的な非線形の疫病モデルである (b) SIR，再帰型ニューラルネットワークモデルである (c) GRU と比較した．これらのベースラインは，すべてオフラインの手法であるため，各時刻における，すべての過去データを用いてパラメータを推定し，その時刻の感染者数を予測した．ここで，ARIMA のパラメータ数は AIC を用いて選択した．また，GRU は，ユニット数 30 の 2 層の RNN 層およびユニット数 30 の 4 層の全結合層で構成し，最適化アルゴリズムに Adam を使用した．ここではさらに，(**P2**) 複数の地域間におけるモデル共有の効果を検証するため，EpiReader を用いず単一の地域のみで予測する場合の精度も検証した．これを (d) EpiCast-L と呼ぶ．

図 6 は，疫病テンソルストリームにおける EpiCast の予測精度を示している．具体的には，オリジナルデータと，7 日先の予測感染者数の推定値の平均二乗誤差（RMSE: root mean square error）を示している．ここで，3 種類の感染ステージについてそれぞれ評価するために，データを次の 3 区間に区切った．(a) Rising stage：各国における感染者数が最大となった日以前の 2 カ月の区間，(b) Falling stage：各国における感染者数が最大となった日以降の 2 カ月の区間，(c) Rising and Falling stages：感染段階 (a) および (b) を合わせた区間．図に示すとおり，EpiCast は，3 種類の区間すべてで高い予測精度を達成した．これは，EpiCast が，疫病テンソルストリームの上昇と下降の両方の感染段階をモデル化する能力を持っていることを意味する．また，EpiCast-L と比較して精度が向上したことから，複数の地域間におけるモデル共有の効果が確認できる．

5.3　Q3. 提案手法の計算時間

続いて，提案アルゴリズムの計算コストを検証する．**図 7** は，図 6 の結果を得たときの各時点 t_c における計算コストを，提案手法の EpiCast と既存手法である ARIMA，SIR および GRU とで比較したものである．図 6 で示した結果を得たときの各時点での計算時間を既存手法と比較した．なお，グラフは，対数スケールで表示されており，また，公正な比較のために，単一の CPU を使用した．図に示すとおり，EpiCast は既存手法と比較し，長期的なイベント予測に対する大幅な性能向上を達成した．具体的には，GRU と比較し 2,200 倍の高速化を実現している．図 7 (a) において，赤枠で囲まれた場所にいくつかのスパイクが見られるが，これは，EpiEstimator が新しいモデルを生成したことを表している．図 7 (b) において，疫病ストリーム全体の計算時間の平均値を示している．図に示すとおり，本研究の提案手法は，モデルパラメータのストリーミング推定により，精度，性能ともに向上しているこ

とが分かる．

6. むすび

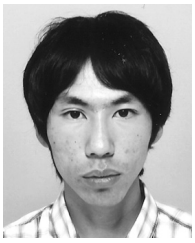
本論文では，大規模疫病データのための高速予測手法である EpiCast について述べた．EpiCast は，(a) 疫病の複雑な拡散過程を非線形モデルで表現し，(b) それらの中に含まれる重要な特徴を各地域で共有し，適切なモデルを選択することで，効果的な感染拡大予測を実現する．また，(c) データストリームの長さに依存せず，一定の計算時間で効率的に感染者数を推定する．公開されている COVID-19 の実データセットを用いて，提案手法が，既存手法よりも計算時間を大幅に短縮しながら，疫病の拡散過程における感染者数の上昇・下降パターンの予測精度を向上させることを実証した．

謝辞　本研究の一部は JSPS 科研費 JP17H04681，JP18H03245，JP19J11125，JP20H00585，JST さきがけ JPMJPR1659，JST 未来社会創造事業 JPMJMI19B3，総務省 SCOPE 192107004 の助成を受けたものです．

参考文献

[1] Adhikari, B., Lewis, B., Vullikanti, A., Jiménez, J.M. and Prakash, B.A.: Fast and near-optimal monitoring for healthcare acquired infection outbreaks, *PLoS Computational Biology*, Vol.15, No.9, p.e1007284 (2019).
[2] Adhikari, B., Xu, X., Ramakrishnan, N. and Prakash, B.A.: EpiDeep: Exploiting Embeddings for Epidemic Forecasting, *KDD*, pp.577–586 (2019).
[3] Andreadis, Georgios and Quirós Gámez, Ana Isabel: Prospective analysis of the impact of a pandemic in Industry 4.0, *MATEC Web Conf.*, Vol.318, p.01037 (online), DOI: 10.1051/matecconf/202031801037 (2020).
[4] Beyazit, E., Alagurajah, J. and Wu, X.: Online Learning from Data Streams with Varying Feature Spaces, *AAAI* (2019).
[5] Chen, P., Liu, S., Shi, C., Hooi, B., Wang, B. and Cheng, X.: NeuCast: Seasonal Neural Forecast of Power Grid Time Series, *IJCAI*, pp.3315–3321 (2018).
[6] Dong, E., Du, H. and Gardner, L.: An interactive web-based dashboard to track COVID-19 in real time, *The Lancet Infectious Diseases*, Vol.20, No.5 (online), DOI: 10.1016/S1473-3099(20)30120-1 (2020).
[7] Durbin, J. and Koopman, S.J.: *Time Series Analysis by State Space Methods*, Oxford University Press, 2 edition (2012).
[8] Islam, M., Muthiah, S., Adhikari, B., Prakash, B. and Ramakrishnan, N.: DeepDiffuse: Predicting the 'Who' and 'When' in Cascades (online), DOI: 10.1109/ICDM.2018.00134 (2018).
[9] Jackson, E.: *Perspectives of Nonlinear Dynamics*, Cambridge University Press (1992).
[10] Kawabata, K., Matsubara, Y., Honda, T. and Sakurai, Y.: Non-Linear Mining of Social Activities in Tensor Streams, *KDD*, pp.2093–2102 (2020).
[11] Liu, C., Hoi, S.C., Zhao, P. and Sun, J.: Online arima algorithms for time series prediction, *AAAI* (2016).
[12] Matsubara, Y. and Sakurai, Y.: Dynamic Modeling and Forecasting of Time-Evolving Data Streams, *KDD*, pp.458–468 (2019).
[13] Matsubara, Y., Sakurai, Y., Prakash, B.A., Li, L. and Faloutsos, C.: Rise and fall patterns of information diffusion: Model and implications, *KDD*, pp.6–14 (2012).
[14] Matsubara, Y., Sakurai, Y., van Panhuis, W.G. and Faloutsos, C.: FUNNEL: Automatic mining of spatially coevolving epidemics, *KDD*, pp.105–114 (2014).
[15] Moré, J.J.: The Levenberg-Marquardt algorithm: Implementation and theory, *Numerical Analysis*, pp.105–116 (1978).
[16] Prem, K., Liu, Y., Russell, T., Kucharski, A., Eggo, R., Davies, N., Jit, M., Klepac, P., Flasche, S., Clifford, S., Pearson, C., Munday, J., Abbott, S., Gibbs, H., Rosello, A., Quilty, B., Jombart, T., Sun, F., Diamond, C. and Hellewell, J.: The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study, *The Lancet Public Health*, Vol.5 (2020).
[17] Qin, Y., Song, D., Cheng, H., Cheng, W., Jiang, G. and Cottrell, G.W.: A Dual-stage Attention-based Recurrent Neural Network for Time Series Prediction, *IJCAI*, pp.2627–2633, AAAI Press (2017).
[18] Rogers, M., Li, L. and Russell, S.J.: Multilinear Dynamical Systems for Tensor Time Series, *NIPS*, pp.2634–2642 (2013).
[19] Sakurai, Y., Matsubara, Y. and Faloutsos, C.: Mining and Forecasting of Big Time-series Data, *SIGMOD, Tutorial*, pp.919–922 (2015).
[20] Shaman, J. and Karspeck, A.: Forecasting seasonal outbreaks of influenza, *Proc. National Academy of Sciences*, Vol.109, No.50, pp.20425–20430 (2012).
[21] Shi, Q., Yin, J., Cai, J., Cichocki, A., Yokota, T., Chen, L., Yuan, M. and Zeng, J.: Block Hankel Tensor ARIMA for Multiple Short Time Series Forecasting, *AAAI* (2020).
[22] Song, H.A., Hooi, B., Jereminov, M., Pandey, A., Pileggi, L.T. and Faloutsos, C.: PowerCast: Mining and Forecasting Power Grid Sequences, *ECML/PKDD* (2017).
[23] Taghvaei, A., De Wiljes, J., Mehta, P.G. and Reich, S.: Kalman filter and its modern extensions for the continuous-time nonlinear filtering problem, *Journal of Dynamic Systems, Measurement, and Control*, Vol.140, No.3 (2018).
[24] Tizzoni, M., Bajardi, P., Poletto, C., Ramasco, J.J., Balcan, D., Gonçalves, B., Perra, N., Colizza, V. and Vespignani, A.: Real-time numerical forecast of global epidemic spreading: Case study of 2009 A/H1N1pdm, *BMC Medicine*, Vol.10, No.1, p.165 (2012).
[25] Venna, S.R., Tavanaei, A., Gottumukkala, R.N., Raghavan, V.V., Maida, A.S. and Nichols, S.: A novel data-driven model for real-time influenza forecasting, *IEEE Access*, Vol.7, pp.7691–7701 (2018).
[26] Wang, C.J., Ng, C.Y. and Brook, R.H.: Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing, *JAMA*, Vol.323, No.14, pp.1341–1342 (2020).
[27] WHO: Coronavirus disease 2019 (COVID-19): Situation report, 72 (2020).
[28] Ye, J., Sun, L., Du, B., Fu, Y., Tong, X. and Xiong, H.: Co-Prediction of Multiple Transportation Demands Based on Deep Spatio-Temporal Neural Network, *SIGKDD*, pp.305–313 (2019).
[29] Zhang, C., Song, D., Chen, Y., Feng, X., Lumezanu, C., Cheng, W., Ni, J., Zong, B., Chen, H. and Chawla, N.: A Deep Neural Network for Unsupervised

Anomaly Detection and Diagnosis in Multivariate Time Series Data, *AAAI*, Vol.33, pp.1409–1416 (online), DOI: 10.1609/aaai.v33i01.33011409 (2019).



木村 輔　（正会員）

2013 年京都産業大学コンピュータ理工学部インテリジェントシステム学科卒業．2016 年同大学大学院博士前期課程修了．2020 年同大学院先端情報学研究科先端情報学専攻博士後期課程修了．博士（先端情報学）．2020 年 4 月より大阪大学産業科学研究所特任助教．時系列データマイニング，自然言語処理，自動テキスト要約の研究に従事．日本データベース学会会員．

松原 靖子　（正会員）

2006 年お茶の水女子大学理学部情報科学科卒業．2009 年同大学大学院博士前期課程修了．2012 年京都大学大学院情報学研究科社会情報学専攻博士後期課程修了．博士（情報学）．2012 年 NTT コミュニケーション科学基礎研究所 RA．2013 年熊本大学大学院自然科学研究科日本学術振興会特別研究員（PD）．2014 年同大学院助教．この間，カーネギーメロン大学客員研究員．2016 年 12 月国立研究開発法人科学技術振興機構さきがけ研究者．2019 年 5 月より大阪大学産業科学研究所准教授．2016 年度日本データベース学会上林奨励賞，情報処理学会山下記念研究賞．2018 年度 IPSJ/ACM Award for Early Career Contributions to Global Research 受賞．大規模時系列データマイニングに関する研究に従事．ACM，電子情報通信学会，日本データベース学会各会員．

川畑 光希　（学生会員）

2016 年熊本大学工学部情報電気電子工学科卒業．2018 年同大学大学院博士前期課程修了．現在，大阪大学大学院情報科学研究科情報システム工学専攻博士後期課程に在籍，日本学術振興会特別研究員（DC2）．第 8 回データ工学と情報マネージメントに関するフォーラム（DEIM 2016）最優秀論文賞，第 11 回 Web とデータベースに関するフォーラム（WebDB Forum 2018）最優秀論文賞，学生奨励賞，企業賞受賞．2019 年度コンピュータサイエンス領域奨励賞（データベースシステム）等受賞．データマイニング，データストリーム処理の研究に従事．日本データベース学会学生会員．

櫻井 保志　（正会員）

1991 年同志社大学工学部電気工学科卒業．1991 年日本電信電話（株）入社．1999 年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了．博士（工学）．2004～2005 年カーネギーメロン大学客員研究員．2013 年熊本大学大学院自然科学研究科教授．2019 年より大阪大学産業科学研究所産業科学 AI センター センター長・教授．本会平成 18 年度長尾真記念特別賞，平成 16 年度および平成 19 年度論文賞，電子情報通信学会平成 19 年度論文賞，日本データベース学会上林奨励賞，ACM KDD best paper awards（2008，2010）等受賞．データマイニング，データストリーム処理，センサーデータ処理，Web 情報解析技術の研究に従事．ACM，IEEE，電子情報通信学会，日本データベース学会各会員．

（担当編集委員　上田 高徳）