

大規模 Web データにおけるキーワード・地域ごとの拡散パターン抽出

村山 太一[†] 松原 靖子[†] 櫻井 保志[†]

[†] 大阪大学産業科学研究所 〒 567-0047 大阪府茨木市

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

あらまし Web 検索データは人々の実社会上の動きを反映することから、ソーシャルセンサとして重要な役割を担っている。このようなデータから、各キーワード間の関係や各地域間の影響力の流れを発見し、将来のキーワードの検索量の変化の遷移を予測することは有用である。そこで、本研究では複雑に共進化する Web データを予測しつつ、キーワード間や地域間の関係が解釈できる新たなモデル “FluxCube” を提案する。FluxCube は、地域同士の影響の流れを捉える拡散反応方程式と、キーワード間の関係を生態系としてとらえるロトカ・ヴォルテラ方程式、2 つの数理モデルに基づいている。また、Physics-informed neural networks の概念を活用することで、推定されるパラメータに基づいた高い解釈性と高い予測性能を同時に達成する。Google Trend を用いた実験の結果、予測精度は比較手法を上回り、各構成要素が性能向上に寄与していることが示された。最後に、Web データからクエリと地域間の潜在的な相互作用を抽出したケーススタディを示す。

キーワード 時系列解析, Web データ, データマイニング

1 はじめに

Web 検索データの増加は、ユーザ行動モデリング [1]、インフルエンザ予測 [2]、業務パフォーマンス測定などの社会・経済の推定や予測のための機会を多く提供する [3, 4, 5]。ソーシャルセンサの役割を持つ Web データは、潜在パターンやインタラクションの発見や、将来のオンラインユーザーの行動量の予測など、人々の意思決定に洞察を提供し利益を生み出すこともある。例えば、マーケティング担当者は、適切な広告戦略や新商品開発を達成するため、オンラインデータから将来どのような商品が人気になっていくか、近い商品や地域との関係について明らかにしたいと考えており、オンラインデータを正確にモデリングすることでこれらの要望は達成できる。

しかしながら、複数のドメインやダイナミクスを持ち日々刻々とデータ量が増加しているオンラインデータのモデリングは依然として困難である。例えば、3 次テンソル（タイムスタンプ、場所、キーワード）で構成される Web データのモデリングを行う場合、以下の課題に対処する必要がある。

(a) **潜在的相互作用の獲得**：各地域やキーワード間の関係の時系列は独立ではない。例えば、同じカテゴリ内のキーワードの検索量はユーザのリソース・興味を巡って競合する [6]。また、ある場所のキーワードの人気は他の地域に影響を与え、異なる場所のキーワードの人気に影響を与える [7]。

(b) **時間変化パターンの獲得**：多くの時系列には、トレンドや季節性などの成分やパターンが含まれ、ブラックフライデーなどの人々の生活の動きが反映される。しかし、このようなパターンは複数あり、手作業で適切にモデル設計に含めることは困難である。隠れたデータ構造を捉えるため、隠れたパターンの推定が可能な自動モデリング手法が必要である。

(c) **正確な予測**：時系列予測は、マーケティング担当者にキー

ワードの将来動向に関する洞察を提供し、人的・物的資源の浪費を回避できる。時系列データの解釈性に関する前述 2 つの課題を解決しつつ、高い予測精度を達成することが重要である。

本論文では、共進化するオンラインデータの大規模集合を対象に、時系列データのモデリングと予測という課題を達成するマイニング手法 “FluxCube” を提案する。提案モデルでは、(a) と (b) の課題に対し、化学物質の時空間変化を表現する反応拡散モデル [8] と餌と捕食者の関係を表現するロトカ・ヴォルテラ方程式 [9, 10] の 2 つの数理モデルを組み合わせることで解決する。加えて、非線形偏微分方程式で記述される物理法則を教師あり学習で取り組む Physics-informed neural networks [11] を活用し (c) に取り組む。

ケーススタディ

図 1 と図 2 に 6 人のアーティスト (World#1) に関するオンラインデータのモデリング結果を示す。具体的には、提案手法によって以下のような傾向を獲得できる。

- **長期間のトレンド**：図 1(a) は、世界 50 カ国における 6 人のアーティストの検索量を表現したテンソルデータである。提案モデルは、キーワードと地域間の隠れた相互作用を自動的に捉え、2 年間の検索量の長期予測を達成する（赤色で示される）。

- **キーワード間の相互関係**：図 1(b) は、各キーワード間の潜在的な相互作用を示す。提案モデルによるキーワード間の相互作用の分類（競合、共生、寄生、相互依存）を矢の種類と向きによって示す。例えば、maroon 5 と beyonce はアメリカにおいて競合の関係であるが、日本では相互依存の関係にあることは、taylor swift と katy perry はどちらの杭においても競合の関係であることが示唆される。

- **地域間の拡散パターン**：図 1(c) では、地域のクラスタリングと各地域間の影響の流れを示す。提案モデルでは、グルー

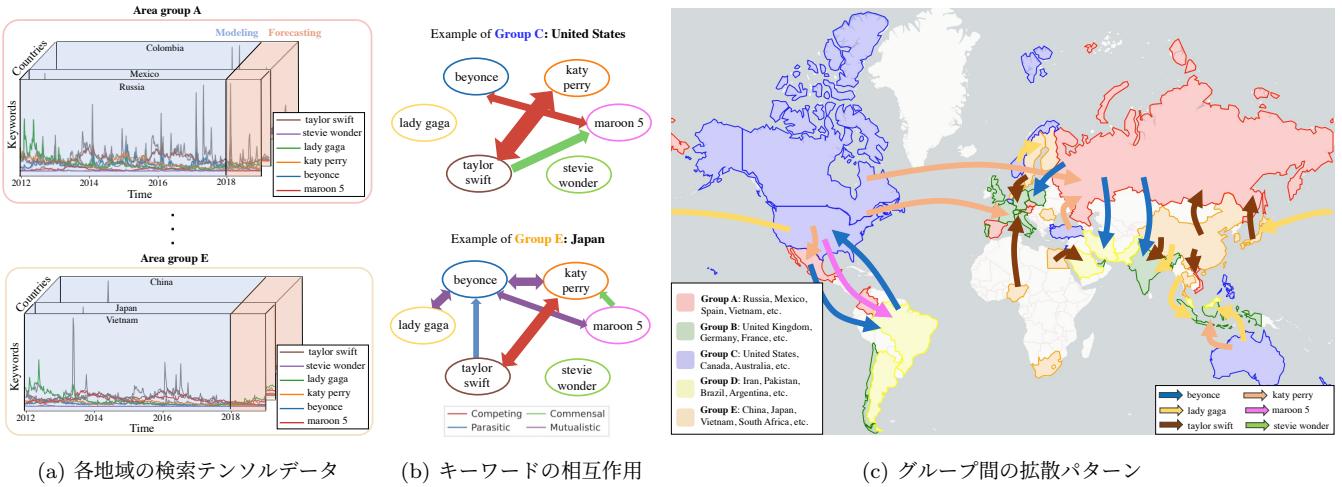


図 1: 6人のアーティスト (World#1) の Web 検索データのモデリング: (a) 検索データのダイナミックスを捉え (青), 長期先の予測結果を示す (赤). (b) 各地域のキーワード間の隠れた相互作用の種類と向きを矢の形式で示す. (c) キーワード同士の相互作用が類似する地域をグループとし, 影響力の流れを示す. ここでは, 2015 年の影響力の流れを可視化したものである.

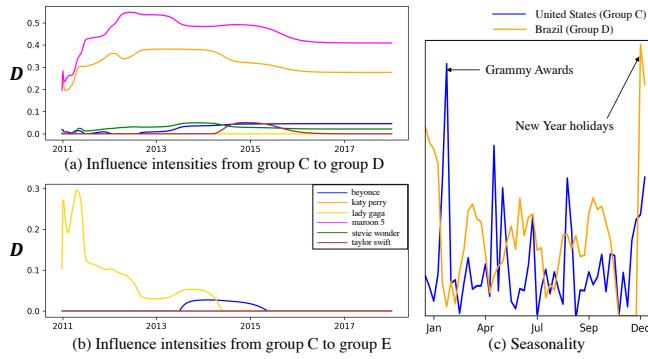


図 2: アメリカの6人のアーティストの潜在パターンの検出: (a) (b) はグループ C からの各アーティストの時間とともに変化する影響力を示す. (c) では, アメリカとブラジルにおける beyonce の検索量から検出された季節性パターンを示す.

グループ B はヨーロッパ、グループ C は北アメリカ、グループ D は東アジアが含まれるようなクラスタリング結果を示した. また、図 2 (a) と (b) で、時間とともに変化する地域間の影響の流れを示す. グループ C から D には maroon 5 と kety perry の影響力の流れがあり、グループ C から E には 2010 年代前半に taylor swift の影響力の流れが存在したが、現在は消滅していることが示される. このように、提案モデルは影響力の流れを直感的に理解できる値として提供する.

・ **季節性:** 図 2 (c) は検索量から抽出される季節性の変動を示す. 提案モデルは、アメリカとブラジルの検索量から「グラミー賞」と「年末年始」に起因する 2 種類のパターンを発見した. このような周期的パターンを発見することはダイナミクスを正確に予測・モデル化する上で重要である.

2 関連研究

時系列予測とモデリング: 従来の時系列予測とモデリングのアプローチの多くは、自己回帰 (AR)、状態空間モデル (SSM)、

カルマンフィルター (KF)，これらを拡張した統計モデルに基づいてる [12, 13]. また、これらの古典的手法に次元削除を組み合わせたモデルとして RMF [14] や SMF [15] などがあり、データマイニングで有用な結果を示しているが、これらの手法はデータの解釈可能性に関して限界がある. 近年、深層学習の急速な発展により、深層学習ベースのモデルが数多く提案されている. 時間変化を捉えられる Convolutional Neural Networks (CNNs) [16, 17] や Recurrent Neural Networks(RNNs) [18], Transformer [19] などが代表例である. 例えば、Transformer ベースのモデルの 1 つである Informer [20] は長期予測で高い精度を達成する. これらの深層学習ベースのモデルは高い予測性能を保証するが、解釈可能性という点では依然課題が残る. 本研究では、数理モデルと深層学習モデルの融合により、高い予測性能と解釈性を同時に実現する.

Web データモデリング: Web 検索データやソーシャルメディアの投稿などのオンラインデータは、ソーシャルセンサーとして人々の動きを捉えることができ、インフルエンザ予測 [2]、人材流入予測 [21]、アイテム人気予測 [22] などの実社会への応用の可能性から、Web データの動きを解明・モデリングすることに大きな関心が集まっている. 例えば、[23] では、Web 上の 2 つのキーワードの競合関係を調査し、感染症モデルでモデリングできることを示した. ソーシャルメディアプラットフォームにおけるコンテンツと視聴者の滞在時間のモデリング [24]、Web コミュニティへの入会脱退のモデリング [25]、検索データと外部イベントの関係抽出 [26] などの Web データモデリング研究が行われてきた. Web 検索データのモデリングとして、数理モデルに基づきウェブ検索データからキーワードの相互作用を捉えるための手法である EcoWeb [6] が提案されている. EcoWeb では、1 つの地域での検索キーワード同士の相互作用の検出に着目しているが、場所間のキーワードの相互作用については検討されていない.

表 1: 記号と定義

| 記号 | 定義 |
|--------------------------------|---------------------------------------|
| t_c, t_f | モデリングと予測のタイムスタンプ |
| $\mathcal{X}^c, \mathcal{X}^f$ | モデリングと予測のテンソルデータ |
| l_c, l_f | モデリングと予測の時間間隔 |
| L, K | 地域とキーワードの数 |
| d_l | 地域グループの数 |
| a_j | キーワード j の成長レート |
| b_j | キーワード j のキャパシティ |
| $c_{jj'}$ | j' 番目のキーワードから j 番目のキーワードへの相互作用の強さ |
| D_{mnk}^t | n 地域から m 地域への影響力の流れ |
| S | 季節性の潜在的パターンマトリックス |
| p | 季節の一周期 |
| α, β | 損失関数における正則化項 |
| σ^i | 地域 i の相互作用を表現するパラメータセット |
| g^1, \dots, g^{d_l} | 地域グループ |
| Θ'_{d_l} | d_l に関連するパラメータセット |

Physics-informed neural network: 近年、深層学習が万能近似能力を持つことから、偏微分方程式や力学モデルの解導出などの文脈で応用されている [11].

物理情報付きニューラルネットワークと呼ばれるこれらのモデルは、支配方程式の近似解を求める順問題 [27]、や、支配方程式に関わるパラメータを観測データから求める逆問題 [28] を解くことができる。私達は、偏微分方程式のパラメータの一部をニューラルネットワークで柔軟に表現するアイデアを活用する。

3 全 体 像

共進化するオンラインデータをモデリング・予測するための提案モデル FluxCube の全体像について説明する。最初に、記号・定義を紹介し、次に提案モデルの特徴を紹介する。

3.1 問題定義

表 1 に本論文で用いる記号を示す。タイムポイント数 T 、地域の数 L 、キーワードの数 K で構成する 3 次テンソル $\mathcal{X} \in \mathbb{R}^{T \times L \times K}$ のオンラインデータを扱う。 \mathcal{X} の要素である x_{tij} は i 地域の j 番目のキーワードの時間 t における検索量を示す。私達の目的は、隠れたパターンや潜在的な相互作用を抽出しつつ、テンソル \mathcal{X} の長期予測を行うことである。長さ l_c の観測したテンソルデータは、 $\mathcal{X}^c = \{x_{tij}\}_{t,i,j=1}^{t_c, L, K}$ と表現する。同様に、予測対象として、 t_c から t_f の部分テンソル $\mathcal{X}^f = \{x_{tij}\}_{t,i,j=t_c,1,1}^{t_f, L, K}$ を設定する。 l_c と l_f はそれぞれ、モデリングと予測のための時間間隔である。まとめると、私達の取り組む問題定義は以下の形である。

PROBLEM (t_f 先の予測). Given 時間 t_c までのテンソル $\mathcal{X}^c = \{x_{tij}\}_{t,i,j=1}^{t_c, L, K}$; Output: t_f 先の値予測 $\hat{\mathcal{X}}^f = \{\hat{x}_{tij}\}_{t,i,j=t_c,1,1}^{t_f, L, K}$, where $t_f = t_c + l_f$.

3.2 反応拡散方程式

キーワードと地域間の潜在的な相互作用を捉るために、提案モデルは反応拡散方程式から着想を得ている。反応拡散方程式は、化学物質の空間・時間変化などの物理現象を表現するモ

デルで、偏微分方程式として表現される。生物学、地質学、物理学など化学以外の分野の変化も記述できる [29, 30, 31, 32]。反応拡散方程式は以下の式で記述できる。

$$\frac{\partial \mathbf{u}}{\partial t} = f(\mathbf{u}, t) + \mathcal{D} \Delta \mathbf{u} \quad (1)$$

、 \mathbf{u} は化学物質の濃度、 t は現在の時間、 \mathcal{D} は拡散係数を示す。右辺の第一項 $f(\mathbf{u}, t)$ は局所的な反応を表現する反応項を表し、第二項 $\mathcal{D} \Delta \mathbf{u}$ は物質の広がりを表現する拡散項を表現する。

提案モデルは、オンラインデータの動きをモデリング、予測するために反応拡散方程式をベースに拡張したものである。例えば、各キーワードの検索量はお互いに影響し、ユーザの興味を奪い合う。このようなキーワード間の相互作用は反応拡散方程式の反応項として表現できる。また、ある地域でのキーワードの流行は別の地域のキーワード人気に影響を与えるかもしれない。このような場所間のキーワードの影響の流れは拡散項として表現できる。このように、地域とキーワードの相互作用は反応拡散方程式で表せると考える。

3.3 提案モデルの機能

Web データなどの 3 次テンソルデータの潜在的なパターンと相互作用を捉るために FluxCube を提案する。まず、潜在的なパターンと相互作用を捉えつつ、正確な予測するために、以下の 3 つ機能を備えていることが必要である。

- **潜在的な相互作用と拡散**: ユーザの行動を反映した Web データは、利用者の嗜好や習慣など多くの潜在的な要因によって時間とともに変化していく。これらのキーワードの潜在的な相互作用や場所の影響力の流れを明示的に捉えるために、提案モデルは微分方程式で表現されるロトカ・ヴォルテラ方程式 [9] を含んだ拡散反応方程式 [8] を用いることを提案する。

- **季節性**: オンラインデータには年時パターンが存在する。図 2 (c) が示すように、年末年始に検索量が増加する場合がある。このような季節性の変化を捉える適切なモデリングが必要である。

- **時変コンポーネント**: Web データは他の地域の影響などの外部イベントによって大きく変化することがある。時間とともに変化する外部イベントや他の地域との関係は、微分方程式の変数では容易に説明できないため、柔軟なモデル化が求められる。そこで、本研究では physics-informed neural networks [11] のアイデアを活用する。直感的には、万能近似能力を持つ深層学習 [33] と、解釈可能性の高い微分方程式を組み合わせたものである。

4 FluxCube

Web データなどのオンラインデータに対して適用する、反応拡散方程式 (式 (1)) に基づいた FluxCube の式は以下の形となる。

$$\frac{\partial x_{tij}}{\partial t} = f(x_{tij} | \mathcal{X}_{t,i,:}^c) + g(x_{tij} | \mathcal{X}_t^c, t) \quad (2)$$

$$\hat{x}_{t+1ij} = F\left(\frac{\partial x_{tij}}{\partial t} + x_{tij}\right) \quad (3)$$

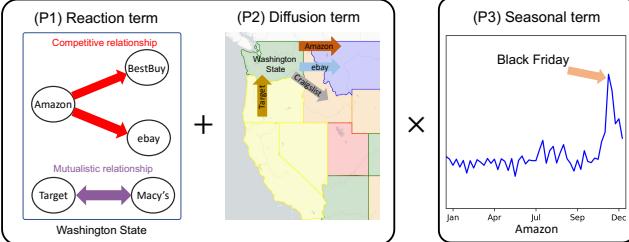


図 3: アメリカワシントン州の e-コマース関連の検索データ (US#1) を対象に提案モデルを適用した場合のイメージ図. (P1) 反応項でキーワード間の関係を捉え, (P2) 拡散項で地域間のキーワードの影響力の流れを表現する. そして, (P3) 季節項で各キーワードの季節の変動を捉える.

式 (2) の第 1 項は地域 i におけるキーワード間の相互作用を抽出するための反応項 (P1) として機能する. 第 2 項は, キーワード j は地域間の影響の流れを抽出するための拡散項 (P2) にある. 最後に, 季節性を捉えるための式 (3) を設定する (P3). 本モデリング手法の全体像を図 3 に示す. 次に, FluxCube の各構成要素を順に紹介する.

4.1 反応項 (P1)

反応項は, キーワード間の相互作用のモデリングを目的とする.

ウェブ上の検索キーワードはユーザの注目を集めるために競争している. ウェブ上のユーザの注目は有限であり,もし, そのキーワードが注目されないならばウェブで検索されない, このウェブ上のキーワードとユーザーの注目の関係は, ジャングルにおける種と食物資源の関係に非常に似てる. 例えば, リスザルとクモザルは果物を奪い合い, 食料がなければどちらも生き残れない. そこで, 私達は生物と資源の関係のモデリングに用いられるロトカ・ヴォルテラ方程式 [9, 10] を反応項として用いる.

キーワード j に関する反応項 (式 (2)), は以下のように表される.

$$f(x_{tij}|\mathcal{X}_{t,i,:}^c) = a_j x_{tij} \left(1 - \frac{\sum_{j'} c_{jj'} x_{tij'}}{b_j}\right), \quad (4)$$

ここで, $a_j > 0$, $b_j > 0$, $c_{jj} = 1$ である. 各パラメータのサイズはキーワード数 K に依存する: $\mathbf{a} \in \mathbb{R}^K$, $\mathbf{b} \in \mathbb{R}^K$, and $\mathbf{C} \in \mathbb{R}^{K \times K}$ 各パラメータは以下のように解釈できる.

- a_j : キーワード j の成長率;
- b_j : キーワード j のキャパシティ;
- $c_{jj'}$: j' 番目のキーワードから j 番目のキーワードへの相互作用の強さ. \mathbf{C} の非対角要素にあたる.

\mathbf{C} の非対角要素の符号から, 2 つのキーワードの潜在的な関係を解釈することができ

- $c_{jj'} > 0, c_{j'j} > 0$: 競合関係
- $c_{jj'} < 0, c_{j'j} > 0$: 寄生関係
- $c_{jj'} < 0, c_{j'j} = 0$: 共生関係
- $c_{jj'} < 0, c_{j'j} < 0$: 相互依存

We b 上のキーワード間の関係を, 上記のパラメータの関係に

基づいて推定できる.

4.2 拡散項 (P2)

拡散項は, 地域間の影響の流れのモデリングを目的とする. 偏微分方程式や数理モデルで, Web データに反映される地域間の関係を適切に捉えることは難しい. 拡散反応方程式では, 化学物質は外的影響を受けずに一定に広がるという観測に基づき, 拡散係数は定数で空間・時間的な変化を表現する. しかしながら, Web 上における場所間の影響は, 外的要因によって一定ではない. 相互作用の時間変化を捉えるには, 動的線形モデルを利用することが考えられるが, 複雑な現象や急激な変化を考慮することは困難である. 我々は, ニューラルネットワークを数理モデルの一部として活用し, 時間とともに変化する場所間の相互作用を表現する. これは, 一種の physics-informed neural network [11] に相当する. 具体的には, 位置 i のキーワード j に関する拡散項 (式 (2)) を以下の形で表現する.

$$\mathbf{D}^t = \text{ReLU}(\text{RNN}(1:t)) \quad (5)$$

$$\mathbf{E}_i^t = \mathbf{D}_i^t \odot \mathcal{X}_t, \quad (i = 1, \dots, d_l) \quad (6)$$

$$g(x_{tij}|\mathcal{X}_t^c, t) = \sum_{i'}^{d_l} \mathbf{E}_{i,i',j}^t \quad (7)$$

ここで, $\mathbf{D}^t \in \mathbb{R}^{d_l \times d_l \times K}$, $\mathbf{E}^t \in \mathbb{R}^{d_l \times d_l \times K}$, $\mathbf{D}_{i,i,:}^t = 0$. $\mathbf{D}_i^t \in \mathbb{R}^{d_l \times K}$ と $\mathbf{E}_i^t \in \mathbb{R}^{d_l \times K}$ はそれぞれ \mathbf{E}^t と \mathbf{D}^t の i 番目のベクトルを示す. 各ユーザの場所間の注目の流れを, 時間変動するパラメータを扱える RNN (式 (5)) で表現する. RNN は, 共変量の時間依存性を捉えつつ, 時間変動するパラメータを許容することで, 高い表現力を実現する. RNN の出力である \mathbf{D}^t はテンソルの形に変換される (\mathbf{D}_{mnk}^t は場所 m におけるキーワード k 人気が場所 n にどの程度寄与するかを示す). つまり, \mathbf{D}_{mnk}^t はキーワード k の地域 n から m への影響の強さを表す. 図 3 (a) と (b) に $D^t(t=1, \dots, t_c)$ の例を示す. 地域 i の時刻 t においてキーワード j が他の地域から受ける影響の総和は, 式 (6) と (7) が示すように, 式 (5) の出力 $\mathbf{boldsymbol{D}}$ を観測されるテンソルデータ $\mathbf{mathcal{X}}$ を乗算することで求まる. このように, 数理モデルの一部を深層学習を適用することで柔軟なモデリングと高い説明性の両方が達成できる.

4.3 季節項 (P3)

最後に, オンラインデータの季節性・周期パターンを考慮できるように拡張する. 各キーワードは常に特定のユーザの関心を集めてるが, クリスマスやブラックフライデーなどの季節イベントに応じて動的に関心の多さが変化する. そこで, オンラインデータに隠れた季節変動を検出するため, 拡散反応方程式から季節性のパターンを考慮できる形に定式化する:

$$F\left(\frac{\partial x_{tij}}{\partial t} + x_{tij}\right) = (\mathbf{1} + [\mathbf{S}_{t \bmod p}]_{i,j}) \odot \left(\frac{\partial x_{tij}}{\partial t} + x_{tij}\right) \quad (8)$$

ここで, $\mathbf{S} \in \mathbb{R}^{p \times L \times K}$, $\mathbf{S} \geq 0$, p は季節の一周期, \odot はアダマール積を示す. また, $\mathbf{boldsymbol{S}}$ は, 図 3 の季節項が示すような潜在的なパターンを表現する行列となる.

4.4 損失関数

モデルの学習には、正則化項を持つ Mean Squared Error (MSE) 損失を以下の形式で利用する。

$$\|\mathcal{X}^c - \hat{\mathcal{X}}\|^2 + \alpha \sum \|\mathbf{D}\|^2 + \beta \sum \|\mathbf{S}\|^2 \quad (9)$$

ここで、 α と β は正則化項の重みを制御するパラメータとなり、 $\hat{\mathcal{X}}$ はモデリングの値となる。第1項はモデリングの値と観測値の差にペナルティを与える MSE 損失、第2項と第3項は場所間の影響 ($\Theta \boldsymbol{D}$) と潜在的な季節パターン ($\Theta \boldsymbol{S}$) の大きな値の採用を抑制する正則化項となる。

5 最適化アルゴリズム

我々は、複数の地域のオンラインデータをモデリング・予測するためのモデル FluxCube を提案したが、多くの地域のオンラインデータに対し最適化するためには依然として2つの課題が存在する：(a) 推定する候補パラメータの数が多く、学習が非効率的で不安定になること、(b) 地域間の相互作用が複雑になり、結果が解釈しにくくなる。そこで、キーワード同士のパターンが類似した地域を同じグループにすることで、解釈しやすい形に変換するための学習アルゴリズムを紹介する。

5.1 自動的地域クラスタリング

オンラインデータに対して地域の観点でクラスタリングするための我々の仮説は単純である。それは、反応項 (P1) で表現されるキーワード間の相互作用が類似している場所は同じ地域グループとみなすことができるというものです。クラスタリングの手法として、先行研究 [34] に基づき、Uniform Manifold Approximation and Projection (UMAP) [35] と K-means を用いて、地域数 L を地域グループ数 d_l にクラスタリングする。式は以下の形となる。

$$\mathbf{o}^i = UMAP(\mathbf{a}^i, \mathbf{b}^i, \mathbf{C}^i) \quad (i = 1, \dots, L) \quad (10)$$

$$g^1, \dots, g^{d_l} = \text{K-means}(\mathbf{o}, d_l) \quad (11)$$

ここで、 $\mathbf{o}^i \in \mathbb{R}^2$ 、 $\mathbf{o} \in \mathbb{R}^{L \times 2}$ はキーワード相互作用の圧縮表現となる。反応項の式 (4) の地域 i のキーワード同士の相互作用を表現する、3つのパラメータ ($\mathbf{a}^i, \mathbf{b}^i, \mathbf{C}^i$) を利用する。これら3つのパラメータは UMAP により 2次元ベクトル \boldsymbol{b}^i に圧縮され、K-means により d_l 個の地域グループ (g^1, \dots, g^{d_l}) にクラスタリングする。

5.2 最適な地域グループ数

最適なグループ数 d_l を含んだパラメータ数が少ないモデルで観測データを十分に説明できると有用といえる。そこで、適切な d_l を見つけるための手法として、Minimum Description Length (MDL) [36, 37, 38] を利用する。MDL は、データ符号化コストとモデル記述コストの和を最小化することで、データの良い要約を提供するもので、以下の形で記述される。

$$\langle \mathcal{X}^c | \Theta' \rangle + \langle \Theta'_{d_l} \rangle + C \quad (12)$$

ここで、 $\langle \mathcal{X}^c | \Theta' \rangle$ はモデルパラメータ群 Θ' が与えられた時のデータ記述コスト、 $\langle \Theta'_{d_l} \rangle$ は d_l によって変化するパラメータ群 Θ'_{d_l} を記述するコスト、 C は d_l に影響されない定数である。データを圧縮しつつ、背後にあるパターンを説明できる時、良いモデルであるという仮定に基づいたコスト関数となる。

データエンコーディングコスト： 観測データ \mathcal{X}^c をハフマン符号化 [38] に基づいて Θ' で符号化する。この符号化方式では、平均 μ 、分散 σ^2 のガウス分布における負の対数尤度の形で \mathcal{X}^c の各値に以下の形でビット数を割り当てられる

$$\langle \mathcal{X}^c | \Theta' \rangle = \sum_{t,i,j=1}^{t_c, L, K} -\log_2 p_{\mu, \sigma}(x_{t,i,j}^c - \hat{x}_{t,i,j}^c) \quad (13)$$

ここで、 $\hat{x}_{t,i,j}^c \in \hat{\mathcal{X}}^c$ は提案モデルによって生成される $x_{t,i,j}^c \in \mathcal{X}^c$ の出力である。

モデルコスト： モデルコストは、モデルを記述するために必要なビット数を示す。より複雑なモデル構造を採用すると、総コストは高くなる。ここでは、最適な d_l を選択するという目的から、 d_l に関するパラメータ、すなわち、拡散項の \boldsymbol{D} にのみ着目する。 d_l に関係ないパラメータは定数値 C として扱い、コスト計算では無視できる。提案モデルのコストは以下の形として表される。

$$\langle \Theta'_{d_l} \rangle = \langle d_l \rangle + \langle \mathbf{D} \rangle \quad (14)$$

$$\langle d_l \rangle = \log^*(d_l) \quad (15)$$

$$\langle \mathbf{D} \rangle = |\mathbf{D}| \cdot (\log(t_c) + 2 \cdot \log(d_l) + \log(K) + c_F) + \log^*(|\mathbf{D}|) \quad (16)$$

ここで、 \log^* は整数のユニバーサルコード長、 $|\cdot|$ は非ゼロの要素数、 c_F は浮動小数点コスト（32ビット）を表す。

5.3 最適化

まとめると、キーワード間の相互作用が類似している場所は同じ地域グループとし、提案モデル FluxCube を学習する手順を Algorithm 1 で示す。基本的な方針は、地域グループ数 d_l を増やしながら、式 (12) のエンコーディングコストを最小となるモデルを探索する。

6 実験と結果

6.1 実験設定

6.1.1 データセット

2011年1月1日から2020年12月31日までの約10年間の検索量が含まれた、2種類の地域のGoogle Trendsを利用した。

- **US データ**：4種類の米国50州の検索データセット
- **World データ** GDPスコアが上位50の国の検索データセット

データセットに含まれるキーワードを表2に示す。

6.1.2 比較手法

時系列予測・モデリングのタスクにおける比較手法として

Algorithm 1 Optimization

Require: Input tensor \mathcal{X}^c

Ensure: Learned FluxCube with interpretable parameters

```

1: /* Initialize
2:  $d_l \leftarrow 1$ 
3:  $g^{d_l} \leftarrow \{1, \dots, L\}$ 
4: MinCost  $\leftarrow \infty$ 
5: while improving the cost do
6:   if  $d_l > 1$  then
7:     /* Automatic location clustering (Section 5.1)
8:      $g^1, \dots, g^{d_l} \leftarrow Clustering(d_l, \mathbf{a}_{d_l-1}, \mathbf{b}_{d_l-1}, \mathbf{C}_{d_l-1})$ 
9:   end if
10:  /* FluxCube training (Section 4)
11:   $FluxCube_{d_l} \leftarrow Training(\mathcal{X}^c, g^1, \dots, g^{d_l})$ 
12:  /* Calculate the encoding cost (Section 5.2)
13:  Cost_{d_l}  $\leftarrow f(FluxCube_{d_l}, \mathcal{X}^c)$  //  $f(\cdot)$ : Equation(12)
14:  /* Update  $d_l$ 
15:  if  $Cost_{d_l} < \text{MinCost}$  then
16:    MinCost  $\leftarrow Cost_{d_l}$ 
17:     $d_l \leftarrow d_l + 1$ 
18:  else
19:    return  $FluxCube_{d_l}$ 
20:  end if
21: end while

```

表 2: データセット

| ID | データセット | キーワード |
|---------|------------|--|
| US#1 | E-commerce | Amazon/Apple/BestBuy/Costco/Craigslist/Ebay/Homedepot/Kohls/Macys/Target/Walmart |
| US#2 | VoD | AppleTV/ESPN/HBO/Hulu/Netflix/Sling/Vudu/YouTube |
| US#3 | Sweets | Cake/Candy/Chocolate/Cookie/Cupcake/Gum/Icecream/Pie/Pudding |
| US#4 | Facilities | Aquarium/Bookstore/Gym/Library/Museum/Theater/Zoo |
| World#1 | Music | Beyonce/KatyPerry/LadyGaga/Maroon5/StevieWonder/TaylorSwift |
| World#2 | SNS | Facebook/LINE/Slack/Snapchat/Twitter/Viber/WhatsApp |
| World#3 | Apparel | Gap/H&M/Primark/Uniqlo/Zara |

以下のものを用いる. EcoWeb [6], SMF [15], Gated Recurrent Unit (GRU) [39], Informer [20].

6.1.3 実験設定

前処理: 2011年1月1日から2017年12月31日までを訓練（モデリング）期間（7年間 = t_c ），次に2018年1月1日から2018年12月31日までを検証期間（1年間），そして2019年1月1日から2020年12月31日をテスト期間（2年間）として設定し，検索量を0–1の範囲へ正規化した。

評価指標: これらのモデルの予測性能を13週（3ヶ月），26週（半年），52週（1年）先の3つの時点で評価した。評価指標として，Root Mean Squared Error(RMSE)とMean Absolute

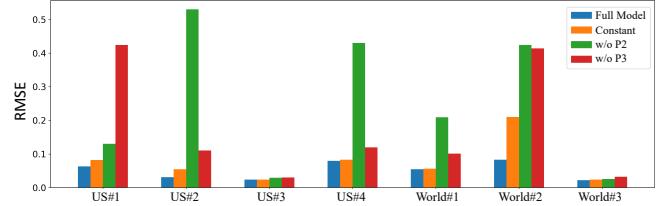


図 4: 52 週先予測におけるアプリケーション実験の結果

Error(MAE) の 2 つの評価指標を用いた。これらの評価指標の値が低いほど、予測精度が高いことを示す。

ハイパーパラメータ: 深層学習ベースのモデルのパラメータはすべて Adam [40] で更新される。DeepAR, GRU, Informer は、入力・予測長を 52，損失関数を MSE 損失と設定した。比較モデルの隠れ層数などのパラメータはグリッドサーチを適用した。FluxCube の学習における RNN の隠れ層の大きさは検証データを用いて (16, 32, 64) から選択した。また、Google Trend データが週次データであり、年次パターンを捉えるために、式 (8) の p を 52、式 (9) のハイパーパラメータ $alpha$ と $beta$ を 0.1、エポック数を 2,000 とし、early stopping を適用した。

6.2 実験結果

実験結果を表 3 に示す。FluxCube が比較モデルのほとんどにおいて高い精度を達成しており、提案モデルの有用性を確認できる。EcoWeb は提案モデルと同様のモチベーションで提案されたモデルであるが、キーワード間の相互作用のみを捉えるため、将来予測において高い精度を達成しなかった。季節パターンの抽出に重点を置いた SMF は、EcoWeb と比較して季節性の強いオンラインデータの予測に有用であることを示す。DeepAR や GRU などの RNN ベースのモデルは、ベースラインとして優れており、13 週先の予測では Informer とほぼ同じ性能を示したが、52 週先などの長期予測では不安定な結果を示す場合もあった。これらの結果は、長期予測の精度向上することは難しいことを示す。一方で、長期予測に特化した Informer は、比較対象モデルの中で最も高い精度を達成した。Informer は、52 週先予測の精度が 13 週先予測の精度を大きく下回らないという傾向を示している。

FluxCube は RMSE で 7 データセット中 5 データセット、MAE で 7 データセット中 6 データセットで最高精度を達成した。提案モデルは、キーワードの相互作用、地域間の拡散、季節性という 3 つの要素を考慮し、オンラインデータを明示的にモデル化することで、高い予測性能と高い解釈可能性を提供している。

アプリケーション実験: FluxCube の構成部分の有効性を示すため、3 つの設定 (Constant モデル, Full w/o P2, Full w/o P3) のアプリケーション実験を行う。Constant モデルは式 (5) の RNN 部分を時間不変のパラメータに置き換えたモデルで、時間変動する RNN の有効性について検証する。Full w/o P2 と Full w/o P3 は提案モデルからそれぞれ拡散項と季節項を取り除いたモデルとなっている。各データセットの 52 週先予

表 3: 予測性能比較

| Dataset | Model | 13 weeks | | 26 weeks | | 52 weeks | |
|---------|----------|---------------|---------------|---------------|---------------|---------------|---------------|
| | | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| US#1 | EcoWeb | 0.1470 | 0.0950 | 0.1554 | 0.1082 | 0.1654 | 0.1197 |
| | SMF | 0.0869 | 0.0620 | 0.0910 | 0.0654 | 0.1012 | 0.0674 |
| | DeepAR | 0.1003 | 0.0634 | 0.1302 | 0.0907 | 0.1385 | 0.1014 |
| | GRU | 0.1723 | 0.1175 | 0.1924 | 0.1374 | 0.2059 | 0.1525 |
| | Informer | 0.1477 | 0.1045 | 0.1375 | 0.0985 | 0.1575 | 0.1111 |
| | FluxCube | 0.0478 | 0.0257 | 0.0574 | 0.0323 | 0.0631 | 0.0365 |
| US#2 | EcoWeb | 0.1440 | 0.1133 | 0.1981 | 0.1621 | 0.1920 | 0.1684 |
| | SMF | 0.0621 | 0.0445 | 0.0713 | 0.0522 | 0.0760 | 0.0529 |
| | DeepAR | 0.1471 | 0.1026 | 0.1781 | 0.1314 | 0.1906 | 0.1474 |
| | GRU | 0.1518 | 0.1171 | 0.1619 | 0.1231 | 0.1683 | 0.1440 |
| | Informer | 0.1277 | 0.0878 | 0.1292 | 0.0876 | 0.1436 | 0.1012 |
| | FluxCube | 0.0245 | 0.0130 | 0.0276 | 0.0156 | 0.0310 | 0.0181 |
| US#3 | EcoWeb | 0.1555 | 0.1208 | 0.1730 | 0.1384 | 0.1754 | 0.1369 |
| | SMF | 0.0276 | 0.0186 | 0.0281 | 0.0170 | 0.0281 | 0.0190 |
| | DeepAR | 0.1107 | 0.0753 | 0.1267 | 0.0833 | 0.1309 | 0.0908 |
| | GRU | 0.1300 | 0.0869 | 0.1368 | 0.0943 | 0.1368 | 0.0939 |
| | Informer | 0.1322 | 0.0954 | 0.1311 | 0.0946 | 0.1279 | 0.0914 |
| | FluxCube | 0.0200 | 0.0121 | 0.0222 | 0.0136 | 0.0238 | 0.0148 |
| US#4 | EcoWeb | 0.0847 | 0.0573 | 0.1348 | 0.0950 | 0.1511 | 0.1182 |
| | SMF | 0.0905 | 0.0762 | 0.1100 | 0.0751 | 0.1206 | 0.1077 |
| | DeepAR | 0.0927 | 0.0682 | 0.1662 | 0.1119 | 0.2168 | 0.1639 |
| | GRU | 0.1199 | 0.0872 | 0.1737 | 0.1223 | 0.2319 | 0.1764 |
| | Informer | 0.1014 | 0.0720 | 0.0992 | 0.0690 | 0.1055 | 0.0762 |
| | FluxCube | 0.0495 | 0.0256 | 0.0610 | 0.0358 | 0.0794 | 0.0503 |
| World#1 | EcoWeb | 0.1259 | 0.0831 | 0.1422 | 0.1034 | 0.2101 | 0.1460 |
| | SMF | 0.0936 | 0.0783 | 0.0901 | 0.0602 | 0.1087 | 0.0787 |
| | DeepAR | 0.0900 | 0.0636 | 0.0929 | 0.0681 | 0.1395 | 0.0972 |
| | GRU | 0.0633 | 0.0452 | 0.0718 | 0.0501 | 0.0823 | 0.0572 |
| | Informer | 0.0704 | 0.0423 | 0.0719 | 0.0416 | 0.0738 | 0.0446 |
| | FluxCube | 0.0454 | 0.0274 | 0.0477 | 0.0286 | 0.0546 | 0.0331 |
| World#2 | EcoWeb | 0.0908 | 0.0300 | 0.1089 | 0.0570 | 0.1353 | 0.0742 |
| | SMF | 0.0841 | 0.0436 | 0.0799 | 0.0454 | 0.0826 | 0.0480 |
| | DeepAR | 0.0374 | 0.0098 | 0.0585 | 0.0199 | 0.0643 | 0.0209 |
| | GRU | 0.0401 | 0.0159 | 0.0588 | 0.0174 | 0.0739 | 0.0254 |
| | Informer | 0.0371 | 0.0159 | 0.0595 | 0.0196 | 0.0642 | 0.0208 |
| | FluxCube | 0.0704 | 0.0271 | 0.0711 | 0.0304 | 0.0831 | 0.0351 |
| World#3 | EcoWeb | 0.0523 | 0.0208 | 0.0626 | 0.0200 | 0.1080 | 0.0293 |
| | SMF | 0.0206 | 0.0111 | 0.0289 | 0.0160 | 0.0254 | 0.0195 |
| | DeepAR | 0.0211 | 0.0110 | 0.0275 | 0.0099 | 0.0613 | 0.0214 |
| | GRU | 0.0191 | 0.0090 | 0.0217 | 0.0096 | 0.00235 | 0.0115 |
| | Informer | 0.0223 | 0.0105 | 0.0226 | 0.0105 | 0.0214 | 0.0108 |
| | FluxCube | 0.0176 | 0.0085 | 0.0214 | 0.0096 | 0.0221 | 0.0100 |

測の RMSE を図 4 に示す。これらの結果から、どのアプローチモデルも提案モデルよりも低い予測性能となっており、FluxCube の各構成要素はオンラインデータのモデリングに有用であることが示唆される。

7 Case study

US#1: 図 5 (a-1) は提案モデルが、複数の国やキーワードの潜在的な動的パターンをうまく捉え、amazon や apple の季節性パターンや、craigslist の減少傾向を捉え、将来の値を予測できることを示す。提案モデルの反応項に基づいて算出される図 5 (a-2) は、ニューヨーク州で amazon は bestbuy や ebay と

競合し、target と macys は相互依存の関係にあることが、ネットワークグラフを通じて示される。図 5 (a-3) から、アメリカの東部が属するグループ A から他の地域に Macy's の影響が流れているなどが明らかになる。

World#2: 図 5 (b-1) では、提案モデルが、スペインでの slack や snapchat の増加傾向を、日本での facebook の減少傾向を捉えていることが示される。一方で、日本で 2018 年以降急速に普及が進んだ slack の動向を十分に捉えることができなかった結果となっている。図 5 (b-2) では、スペインで facebook は LINE、viber、twitter など複数のプラットフォームと相互依存的な関係であることが示唆される。多くのソーシャルメディアは、ネットユーザを奪い合うというよりも、共生する傾向があることがわかる。図 5 (b-3) は提案モデルが、twitter を主に利用する日本やサウジアラビアをグループ B、facebook を主に利用するアメリカやイギリスをグループ F に分類したことがわかる。地域グループ間の影響力の相互作用の結果、facebook の影響力はグループ F から他の地域グループへ流れ、イスラエルで開発された viber の影響力は中東諸国を含むグループ E から他の地域グループに流れていることがわかる。

これらのケーススタディが示すように、FluxCube はキーワードや場所の関係性を明らかにする複雑な相互作用を記述することができ、オンラインデータのモデリングや予測に適している。

8 おわりに

本研究では、Web 検索データのような時間発展する大規模オンラインデータのモデル化・予測を試み、拡散反応方程式と生態系モデルに基づく FluxCube を提案した。提案手法では、観測データの潜在的な力学的作用を抽出することで、その傾向・季節性・相互作用などを取り出すことができる。また、複数の Google Trends データに適用することで、提案モデルは潜在的ダイナミクスを捉え、高い予測性能を達成した。また、予測性能だけでなく、観測データの背後に存在する潜在的な相互作用や影響の流れを、人間が解釈可能な形で提供した。例えば、ワシントン州では costco は bestbuy や homedepot とオンラインユーザーを獲得するために競合していることが、我々の事例で示されている。このようなデータの背後にある潜在的な相互作用を発見することは、我々の意思決定を支援することにつながることが期待できる。

謝 辞

本研究の一部は JSPS 科研費、JP20H00585、JP21H03446、JP22K17896、国立研究開発法人情報通信研究機構委託研究 NICT 03501、総務省 SCOPE JP192107004、JSTAIP 加速課題 JPMJCR21U4、ERCA 環境研究総合推進費 JP-MEERF20201R02、の助成を受けたものです。

文 献

- [1] Yasuko Matsubara, Yasushi Sakurai, Christos Faloutsos, Tomoharu Iwata, and Masatoshi Yoshikawa.

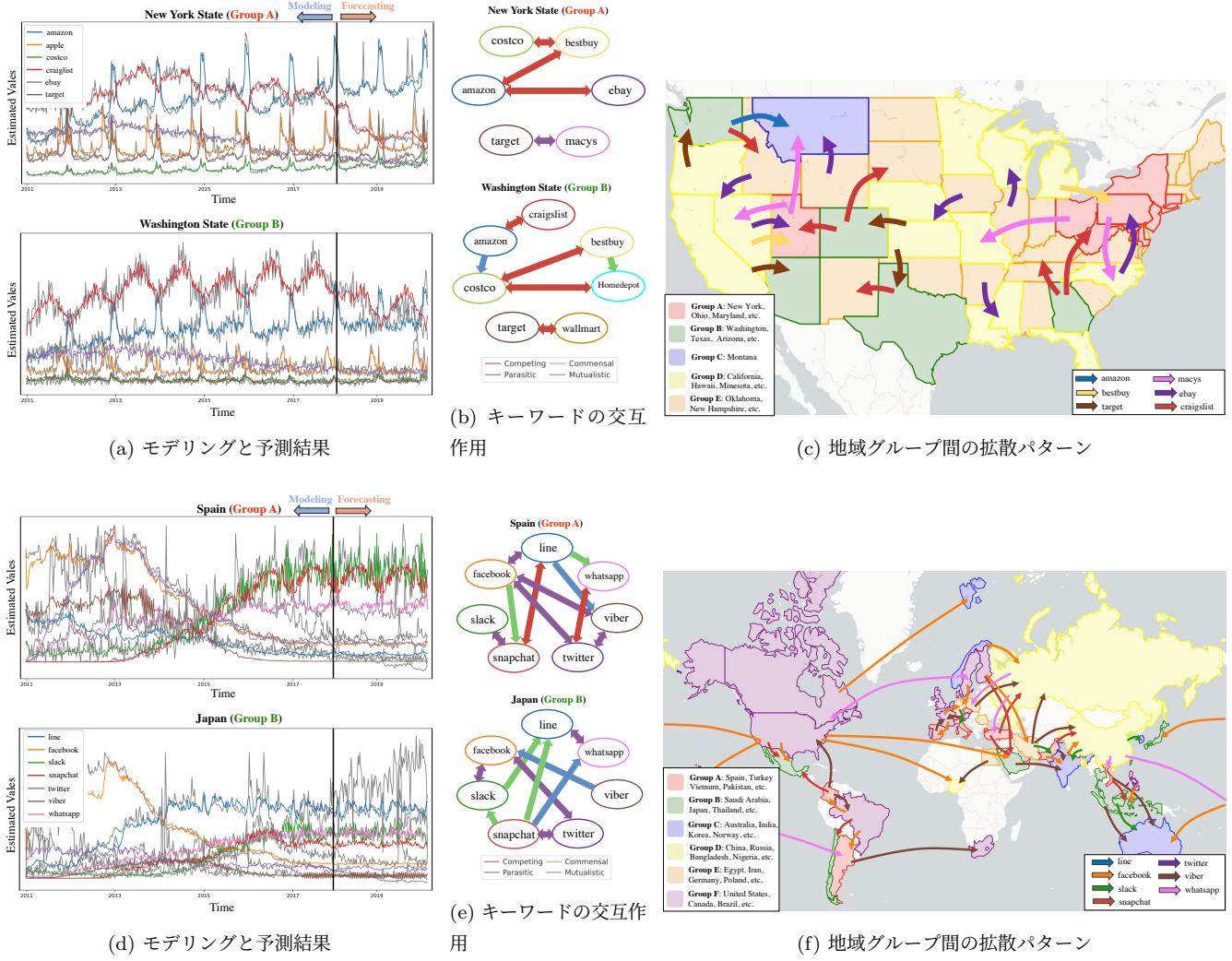


図 5: US#1 (a) と World#2 (b) データへの FluxCube のモデリング・予測結果 : (a/b-1) オンラインデータに対するモデリング・予測結果を示す。観測データをグレーで表示する。(a/b-2) 各地域におけるキーワード間の潜在的な相互作用のネットワークグラフを示す。提案モデルによって示唆された相互作用の種類を矢の色で示す。競合関係や相互依存は双方向性である一方で、共生や寄生関係は一方向の矢印で表現される。(a/b-3) 国 / 州を地域グループに分割したクラスタリング結果と、各地域グループ間の影響の流れを示す。

Fast mining and forecasting of complex time-stamped events. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 271–279, 2012.

- [2] Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2009.
- [3] Jure Leskovec, Lars Backstrom, and Jon Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 497–506, 2009.
- [4] Julia Proskurnia, Przemyslaw Grabowicz, Ryota Kobayashi, Carlos Castillo, Philippe Cudré-Mauroux, and Karl Aberer. Predicting the success of online petitions leveraging multidimensional time-series. In *Proceedings of the 26th International Conference on World Wide Web*, pages 755–764, 2017.
- [5] Yuchen Zhao, Neel Sundaresan, Zeqian Shen, and Philip S Yu. Anatomy of a web-scale resale market:

a data mining approach. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1533–1544, 2013.

- [6] Yasuko Matsubara, Yasushi Sakurai, and Christos Faloutsos. The web as a jungle: Non-linear dynamical systems for co-evolving online activities. In *Proceedings of the 24th international conference on world wide web*, pages 721–731, 2015.
- [7] Maya Okawa, Tomoharu Iwata, Yusuke Tanaka, Hiroyuki Toda, Takeshi Kurashima, and Hisashi Kashima. Dynamic hawkes processes for discovering time-evolving communities' states behind diffusion processes. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 1276–1286, 2021.
- [8] Elizabeth E Holmes, Mark A Lewis, JE Banks, and RR Veit. Partial differential equations in ecology: spatial interactions and population dynamics. *Ecology*, 75(1):17–29, 1994.
- [9] SS Antman JE Marsden, L Sirovich S Wiggins, L Glass, RV Kohn, and SS Sastry. Interdisciplinary applied

- mathematics. 2002.
- [10] James W Haefner. *Modeling Biological Systems:: Principles and Applications*. Springer Science & Business Media, 2005.
 - [11] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.
 - [12] James Durbin and Siem Jan Koopman. *Time series analysis by state space methods*, volume 38. OUP Oxford, 2012.
 - [13] Joel Janek Dabrowski, Ashfaqur Rahman, Andrew George, Stuart Arnold, and John McCulloch. State space models for forecasting water quality variables: an application in aquaculture prawn farming. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 177–185, 2018.
 - [14] Hsiang-Fu Yu, Nikhil Rao, and Inderjit S Dhillon. Temporal regularized matrix factorization for high-dimensional time series prediction. *Advances in neural information processing systems*, 29, 2016.
 - [15] Bryan Hooi, Kijung Shin, Shenghua Liu, and Christos Faloutsos. Smf: Drift-aware matrix factorization with seasonal patterns. In *Proceedings of the 2019 SIAM International Conference on Data Mining*, pages 621–629. SIAM, 2019.
 - [16] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.
 - [17] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.
 - [18] 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.
 - [19] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of Advances in neural information processing systems*, pages 5998–6008, 2017.
 - [20] 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 AAAI*, 2021.
 - [21] Le Zhang, Hengshu Zhu, Tong Xu, Chen Zhu, Chuan Qin, Hui Xiong, and Enhong Chen. Large-scale talent flow forecast with dynamic latent factor model? In *The World Wide Web Conference*, pages 2312–2322, 2019.
 - [22] Swapnil Mishra, Marian-Andrei Rizoiu, and Lexing Xie. Feature driven and point process approaches for popularity prediction. In *Proceedings of the 25th ACM international conference on information and knowledge management*, pages 1069–1078, 2016.
 - [23] Alex Beutel, B Aditya Prakash, Roni Rosenfeld, and Christos Faloutsos. Interacting viruses in networks: can both survive? In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434, 2012.
 - [24] Hemank Lamba and Neil Shah. Modeling dwell time engagement on visual multimedia. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1104–1113, 2019.
 - [25] Tianran Hu, Yinglong Xia, and Jiebo Luo. To return or to explore: Modelling human mobility and dynamics in cyberspace. In *The World Wide Web Conference*, pages 705–716, 2019.
 - [26] Shubhra Kanti Karmaker Santu, Liangda Li, Dae Hoon Park, Yi Chang, and ChengXiang Zhai. Modeling the influence of popular trending events on user search behavior. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 535–544, 2017.
 - [27] Maziar Raissi and George Em Karniadakis. Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, 357:125–141, 2018.
 - [28] Zhiwei Fang and Justin Zhan. A physics-informed neural network framework for pdes on 3d surfaces: time independent problems. *IEEE Access*, 8:26328–26335, 2019.
 - [29] Richard FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical journal*, 1(6):445–466, 1961.
 - [30] Srinjoy Ganguly, Upasana Neogi, Anindya S Chakrabarti, and Anirban Chakraborti. Reaction-diffusion equations with applications to economic systems. In *Econophysics and Sociophysics: Recent Progress and Future Directions*, pages 131–144. Springer, 2017.
 - [31] John E Pearson. Complex patterns in a simple system. *Science*, 261(5118):189–192, 1993.
 - [32] C Cosner. Reaction–diffusion equations and ecological modeling. In *Tutorials in mathematical biosciences IV*, pages 77–115. Springer, 2008.
 - [33] Shiyu Liang and R Srikanth. Why deep neural networks for function approximation? In *5th International Conference on Learning Representations, ICLR 2017*, 2017.
 - [34] Kareem Darwish, Peter Stefanov, Michaël Aupetit, and Preslav Nakov. Unsupervised user stance detection on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 141–152, 2020.
 - [35] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.
 - [36] Deepayan Chakrabarti, Spiros Papadimitriou, Dharmendra S Modha, and Christos Faloutsos. Fully automatic cross-associations. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 79–88, 2004.
 - [37] Chun Tang and Aidong Zhang. Mining multiple phenotype structures underlying gene expression profiles. In *Proceedings of the Twelfth International Conference on Information and knowledge management*, pages 418–425, 2003.
 - [38] Jorma Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.

- [39] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1):1–12, 2018.
- [40] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.