

検索クエリを活用した複数国におけるインフルエンザ流行予測

村山 太一[†] 若宮 翔子[†] 荒牧 英治[†]

[†] 奈良先端科学技術大学院大学 〒630—0192 奈良県生駒市高山町 8916-5

E-mail: †{murayama.taichi.mk1,wakamiya,aramaki}@is.naist.jp

あらまし インフルエンザは多くの死者をもたらす可能性があり、流行予測は重要な課題である。これまでの研究では、過去のインフルエンザデータや、ソーシャルメディアや検索クエリなどの Online User-Generated Contents を活用したモデルが提案されてきた。しかし、複数国を対象にしたインフルエンザ予測モデルは存在しない。これは、予測にとって有用な適切な検索クエリを見つけないことができない、検索クエリを有効に活用したモデルが提案できていないという2つの課題が存在するからである。本研究では、複数国を対象としたインフルエンザ予測モデルの構築をマルチタスク学習でおこないつつ、上記の2つの課題に取り組む。1つ目の課題に対しては、英語から他言語への転移学習を用いて、適切な検索クエリを特定する方法を検討する。2つ目の課題に対しては、Attention 機構を活用したモデルを提案することで検索クエリの潜在的特徴を活用する。5カ国のインフルエンザ流行を予測した実験により、ベースラインと比較して、検索クエリとマルチタスク学習を活用した提案モデルの精度が大幅に向上したことを確認した。

キーワード 感染症, 時系列予測, 公衆衛生, マルチタスク学習, 検索クエリ

1 はじめに

近年、COVID-19 の世界的流行が示すように、感染症流行は患者数の増加による経済的損失や混乱を引き起こす可能性がある。これらの感染症を予測することは予防や流行の対処といった観点からも重要である。中でもインフルエンザは、毎年300–500万人の重症患者や、29–65万人の死亡を引き起こしていることが報告されており、各国の公衆衛生機関でも、インフルエンザの流行予測は重要な課題となっている。

これまで、UGC や過去のインフルエンザデータをリソースとして将来のインフルエンザを予測する手法やモデル [1], [2] が数多く提案されてきた。しかしながら、複数国を対象としたニューラルネットワークに基づくモデルは開発されていない。既存研究の多くは1地域もしくは1国の予測に焦点を当てている。我々は、各国のインフルエンザ時系列は季節性の傾向が強く類似性が高いことから、複数国を対象とした単一のインフルエンザ予測モデルの構築が可能と考える。表1に2016年29週から2019年30週の期間に週毎に報告される、人口10万人あたりのインフルエンザ罹患者数を表現する Influenza like illness rates (ILI レート) の5カ国、具体的にはアメリカ合衆国 (US)、日本 (JP)、イングランド (UK)、オーストラリア (AU)、フランス (FR)、のピアソン相関表を示す。これらの値は、文化、場所や言語の異なる5カ国のILIレートが互いに中程度の相関関係であることを表す (相関0.6以上)。また、検索クエリに関しても、特定の疾患を検索する際には異なる国でも似た検索クエリが用いられるなど、各国の健康テーマに対するユーザの検索行動が似ていることが報告 [4], [5] されている。このことから、単一のモデルでも複数国を対象としたインフルエンザ予測モデルの構築が可能と考える。複数国のデータを学習に用いることは、感染症予測モデル構築で生じやすい問題の1つである

表1: 5カ国のILIレート時系列同士のピアソン相関。AUとUSの時系列のピークを合わせるため、AUの時系列を22週先にずらして計算。

	US	JP	UK	AU	FR
US		0.793	0.614	0.840	0.745
JP			0.592	0.751	0.527
UK				0.772	0.693
AU					0.681
FR					

データ量の少なさから生じるオーバーフィットを回避し [6], [7], 各国のインフルエンザ流行の性質を捉えることもできる。

本研究では、1つのモデルによる複数国のインフルエンザ予測を実現するために、タスク間で共通の情報を活用するために複数タスクを共同で学習するマルチタスク問題として捉える。しかし、このようなモデルを実現するためには、予測に有用な検索クエリをどのようにして見つけるか?、検索クエリをモデルの内部でどのように活用するのか? という2つの課題を解決する必要がある。本論文ではこれらの問題に取り組みつつ、複数国のインフルエンザ予測モデルの構築を目指す。

1つ目の課題であるインフルエンザ予測に有効な検索クエリを見つけないことが困難であることは、対象とする国の数に関わらずインフル予測モデルの構築において解決すべき問題である。Google Flu [1] が開発されて以後、UGC を用いたインフルエンザ予測モデルが数多く提案されている。既存の研究 [8], [9] の多くは、入力時系列との相関が高い英語の検索クエリを返す “Google Correlate” を用いて適切な検索クエリの選択が行われてきた。しかし、このような検索クエリの探索方法は非英語圏などの多くの地域では利用できず、さらに2019年12月以後にはサービス自体が終了している。そこで、我々は英語以外の言語で予測モデルの構築において有用な検索クエリを探索する方

法を検討する。具体的には、英語から他言語（日本語とフランス語）への検索クエリの転移手法として、翻訳ベースの方法と word alignment と時系列相関を組み合わせた方法の 2 種類を検討する。

2 つ目の課題は、検索クエリデータなどの UGC データをどのようにしてインフルエンザ予測モデルに組み込むかである。これまでの研究では、過去のインフルエンザデータ [11], [13] と UGC データ [2], [9], [14] の 2 種類のデータが主に用いられてきた。いくつかのモデルではこれらのデータがどちらも用いられており、例えば ARGO モデル [10] では、Google の検索クエリデータと過去のインフルエンザデータを入力とした線形回帰モデルを構築している。しかし、最近の研究 [15] ではインフルエンザ予測モデルにおいて検索クエリの効果は小さく、過去のインフルエンザデータで予測においては十分に報告されている。これらの報告から、検索クエリデータを単純な入力として用いることは予測モデルの改善には不十分であり、入力データ間の相互関係を考慮するなど、データを効果的にモデルに組み込む方法について一考の余地があると考えられる。我々は、この問題を解決するために、過去のインフルデータと検索クエリデータの特性を考慮して入力を行う新たなインフルエンザ予測モデルを提案する。我々のモデルは、インフルエンザの時系列が強い季節性を持つことと、検索クエリデータが非季節性部分の予測において有用な特徴であるという 2 つの背景に基づいて構築される。具体的には、検索クエリデータがインフルエンザデータの非季節性部分の予測において有用であると仮定したモデルを構築する。まず、インフルエンザデータを季節性部分と非季節性部分に分解し、非季節性部分を対象に Attention 機構 [16] によって重み付けされた検索クエリデータを活用して将来の値を予測する。本研究では、このモデルを元に、複数国を対象としたインフルエンザ予測モデルへと拡張した。

本研究では、2 つの課題を解決することで複数国を対象としたインフルエンザ予測モデルの構築を目指す。1 つ目の課題に対し、適切な検索クエリを見つけるために、英語から他言語への転移手法について検討する。2 つ目の課題に対し、検索クエリデータを Attention 機構を用いて活用するなど、入力データの特性を考慮した新たな予測モデルを提案する。5 カ国のインフルエンザ流行予測を行う実験によって、我々は提案モデルがベースラインと比較して高い精度を達成していることを示した。

2 準備

2.1 データセット

2.1.1 ILI データ

本研究では、モデルの入力や予測の対象となるインフルエンザ罹患患者数を表す値として、人口 10 万人あたりのインフルエンザ罹患患者数を表現する ILI データを用いる。実験の対象となる 5 カ国の ILI レートを取得するため、各国の公衆衛生機関のサ

イト (アメリカ合衆国¹, 日本², オーストラリア³, イングランド⁴, フランス⁵) からデータを収集した。週毎に報告される ILI レートを、イングランドについては 2013 年 41 週目–2020 年 29 週目まで、それ以外の国は 2013 年 26 週目–2020 年 29 週目までの期間の値を収集した。以後、国名については国名コード (US, JP, UK, AU, FR) を略称として用いる。

2.1.2 検索クエリ

本研究では入力した UGC データとして検索クエリを用いる。これは、特定の地域内での検索クエリの検索活動を表したもので、週毎に報告される頻度が入力となる。我々はこれらのデータを Google Trends⁶ から取得した。取得した期間は ILI レートと同じであり、検索クエリの時系列は最小値が 0、最大値が 1 となるように正規化した (min-max 正規化)。検索クエリの選択方法として次章で記載する 2 つの手法について検討する。

2.2 検索クエリの選択方法

これまで用いられてきた “Google Correlate” が利用できないことから、インフルエンザ予測のための検索クエリを探索する手法を再考する必要がある。そこで、先行研究 [14] で作成された英語の検索クエリリスト⁷ を用いて、翻訳ベースの方法 (“Translation-based” と表記) と、単語アライメントと時間相関に基づいた方法 (word alignment + Time series correlation から “WT-based” と表記) の 2 手法による多言語への転移手法を検討する。本研究では、US, UK, AU については英語の検索クエリリストの上位 L 個を選択し、JP, FR については 2 つのクエリ転移手法を用いて、英語検索クエリに対応するものを選択する。英語から他言語への転移手法の有用性については、[9], [17], [18] に記述されている。これらの研究は、特定の健康トピックにおいて、ユーザの検索行動は異なる国であっても似ているという点を主張している。

2.2.1 Translation-based

Translation-based は、英語のクエリを他言語に翻訳する単純な手法である。翻訳には Google Translate⁸ を用いる。検索クエリとして文章を入力することは不適切であるが、日本語は翻訳により文章となることがある。このため、日本語は形態素ごとに分割し空白で区切る。

2.2.2 WT-based

翻訳を用いた Translation-based の手法では、適切なクエリを探索できない可能性がある。例えば、英語でインフルエンザの略語を表す “flu” は “インフルエンザ” と訳され、“インフル” という訳を捉えられない。また、翻訳では漢字、ひらがな、カタカナといった字体のヴァリエーションから適切なものを選択できない可能性がある。これらの問題に対して、英語の

1 : <https://www.cdc.gov/>

2 : <https://www.niid.go.jp/niid/ja/>

3 : <https://aspren.dmac.adelaide.edu.au/>

4 : <https://www.gov.uk/government/organisations/public-health-england>

5 : <https://www.sentiweb.fr/>

6 : <https://trends.google.com>

7 : <https://github.com/binzou-ucl/google-flu-mtl>

8 : <https://translate.google.com>

検索クエリと意味的類似性を考慮するための手法である word alignment と過去の ILI データとの時間相関である time series correlation の 2 つを考慮した手法 (WT-based) を検討する。

Word alignment は異なる言語間における単語の類似度を測定するために、横断的な単語埋め込み表現を作成する方法の 1 つであり、単一言語テキストと言語間のアライメントコーパスを用いて学習される [19]。このアプローチによって、他言語の単語から意味的に近い単語を探索できる。Word alignment は、Zhou ら [20] によって提案された手法を用いた。本研究では、英語と日本語、英語とフランス語の言語横断的な埋め込み表現を作成するために、Wikipedia コーパス [21] から学習された fasttext [22] による単語埋め込み表現をリソースとして用いて、word alignment の手法を適用した。意味が似ている語を探索するために、作成した言語横断的な共通の単語埋め込み表現を用いて、コサイン類似度の尺度に基づき、前置詞と冠詞を除いた検索クエリの各単語に類似している k 個の単語を列挙した。コサイン類似度の値は、 Θ_w で表される。

加えて、検索クエリの時系列と学習データの ILI データとの類似度の観点からも検索クエリの探索を行う。具体的には、word alignment により示された候補の検索クエリの時系列と各国の過去の ILI データの時系列とのピアソン相関 Θ_t を用いる。

WT-based は word alignment と time series correlation を組み合わせた手法であり、英語の検索クエリに対応した、 $\Theta_w + \Theta_t$ のスコアが最も高い検索クエリを選択する。この手法はインフルエンザ予測の Transfer モデルを作成するために最適な検索クエリを探索した先行研究 [9] に触発されたものであり、本研究では他言語においてインフルエンザ予測において最適な検索クエリを見つけるための手法として検討する。

3 手 法

まず、3.1 節で本研究で取り組む問題を定式化する。3.2 節では、1 国のインフルエンザ予測を行う提案モデルを紹介し、3.3 節では提案モデルを拡張した複数国を対象としたインフルエンザ予測のためのモデルについて述べる。

3.1 問題設定

本実験の目的は、複数国の将来の ILI レートを予測することであり、この問題を教師ありの機械学習のタスクとして捉える。入力データの 1 つである、 N 週分の過去の ILI データ時系列を $\mathbf{X} = \{x_{t-N+1}, \dots, x_{t-1}, x_t\} \in \mathbb{R}^N$ と表現する。別の入力データである、 N 週、 L 個の検索クエリの時系列を $\mathbf{Q} = \{q_{t-N+1}, \dots, q_{t-1}, q_t\} \in \mathbb{R}^{N \times L}$ と表現する。これらのデータを入力として ILI データの S ステップ先までの値 $\mathbf{Y} = \{y_{t+1}, \dots, y_{t+S}\} \in \mathbb{R}^S$ を予測する。つまり、各国の予測性能を最大化することを目的として、関数 $f: \{\mathbf{X}, \mathbf{Q}\} \rightarrow \mathbf{Y}$ を学習する。

3.2 提案モデル

提案モデルは、検索クエリは ILI データの非季節性部分の予測において有用であるという仮説の元で設計した。これは過去

の ILI データで予測する部分と検索クエリで予測する部分に分けることで将来の ILI レートの予測精度が向上したという先行研究 [23] に基づいて考えられた。

モデルの全体像は図 1 に示す。まず、過去の ILI データを季節性と非季節性成分に分解する。そして、季節成分が将来も一定であるという仮定のもと、非季節性成分の将来の値を予測する。予測のためのモデルとして、検索クエリを Attention 機構 [16] で考慮しエンコーダ・デコーダモデルで予測する。

3.2.1 ILI データの季節性と非季節性への分解

はじめに、入力とする過去の ILI データを seasonal-trend decomposition using LOESS (STL) [24] を用いて成分ごとに分割する。 y_t が時間 t の ILI データとした時、トレンド成分 τ_t 、季節性成分 s_t 、残差成分 r_t と以下の式のように分割される。

$$y_t = \tau_t + s_t + r_t, \quad t = 1, 2, \dots, N, \quad (1)$$

季節成分は、 T の長さで繰り返される一定のパターンである。一方で、トレンド成分は増加や減少などの傾向を示す。詳細なアルゴリズムは [24] に記載される。

提案モデルでは、過去の ILI データを季節性と非季節性成分に分割する。非季節性成分 \mathbf{X}^τ とは、過去の ILI データ \mathbf{X} から季節性成分 \mathbf{X}^s を引いたものである ($\mathbf{X}^\tau = \mathbf{X} - \mathbf{X}^s$)。提案モデルのアーキテクチャでは、季節性成分 \mathbf{X}^s が予測先においても一定のパターンであると仮定し、非季節性成分の将来の値を予測するように開発した。提案モデルの非季節性成分の予測値と、一定のパターンと考えられる季節性成分を足し合わせることで、将来の ILI データの予測値 \mathbf{Y} として出力する。

3.2.2 非季節性成分のエンコーダ・デコーダモデル

提案モデルは、recurrent neural network(RNN) の一種である gated recurrent units (GRUs) [25] と、重要なデータの一部を着目する Attention 機構を組み合わせて構築しており、2 週先以上の予測をするためにエンコーダ・デコーダモデルを適用する。GRU は、過去の ILI データの非季節性成分 \mathbf{X}^τ と検索クエリ \mathbf{Q} の潜在表現を捉えるために用いられ、モデルの解釈性の助けになる Attention 機構は ILI データの予測における各クエリの重要度を算出するために用いられる。

過去の ILI データの非季節性成分 \mathbf{X}^τ の潜在表現をエンコーダ機構で獲得するために、入力データ \mathbf{X}_i^τ と 1 期前の潜在表現 \mathbf{H}_{t-1} を GRU へ以下のように用いる：

$$\begin{aligned} \mathbf{r}_t &= \sigma(U_r \mathbf{X}_i^\tau + W_r \mathbf{H}_{t-1}), & \mathbf{f}_t &= \tanh(U_h \mathbf{X}_i^\tau + \mathbf{H}_{t-1} \odot W_h \mathbf{r}_t), \\ \mathbf{z}_t &= \sigma(U_z \mathbf{X}_i^\tau + W_z \mathbf{H}_{t-1}), & \mathbf{H}_t &= (1 - \mathbf{z}_t) \odot \mathbf{H}_{t-1} + \mathbf{z}_t \odot \mathbf{f}_t, \end{aligned} \quad (2)$$

\mathbf{z}_t と \mathbf{r}_t はそれぞれ reset gate と update gate を表す。加えて、 $U_z, U_r, U_h \in \mathbb{R}^{1 \times M}$ と $W_z, W_r, W_h \in \mathbb{R}^{M \times M}$ はそれぞれの gate のパラメータを表し、 M は GRU の出力次元とする。式 (2) を以下のように表す。

$$\mathbf{H}_i^\tau = GRU(\mathbf{X}_i^\tau), \quad i \in \{t - N + 1, \dots, t\}, \quad (3)$$

$\mathbf{H}_i^\tau \in \mathbb{R}^{1 \times M}$ は GRU の潜在表現を表す。検索クエリも、過去の ILI データと同様に GRU を用いて以下のように潜在表現を

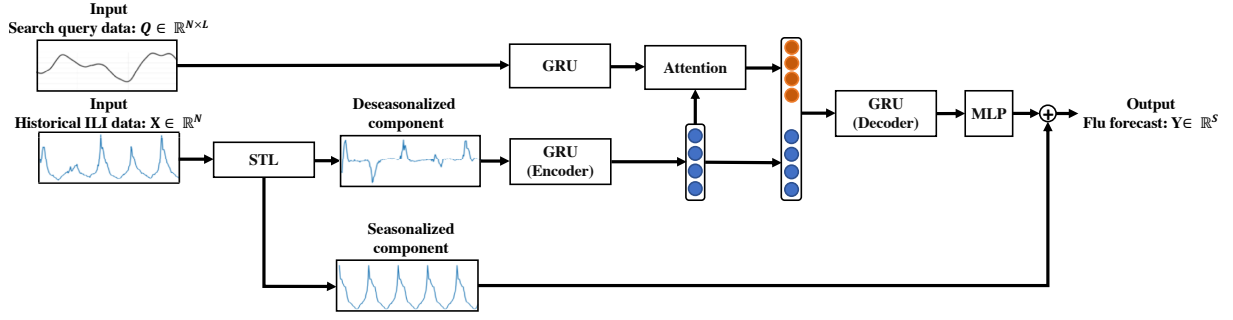


図 1: 提案モデル. 入力とする過去の ILI データを季節性と非季節性に分割し, 非季節性成分の予測に着目する. 非季節性成分の将来の値を予測するために, 検索クエリの Attention 機構を組み込んだエンコーダ・デコーダモデルを適用.

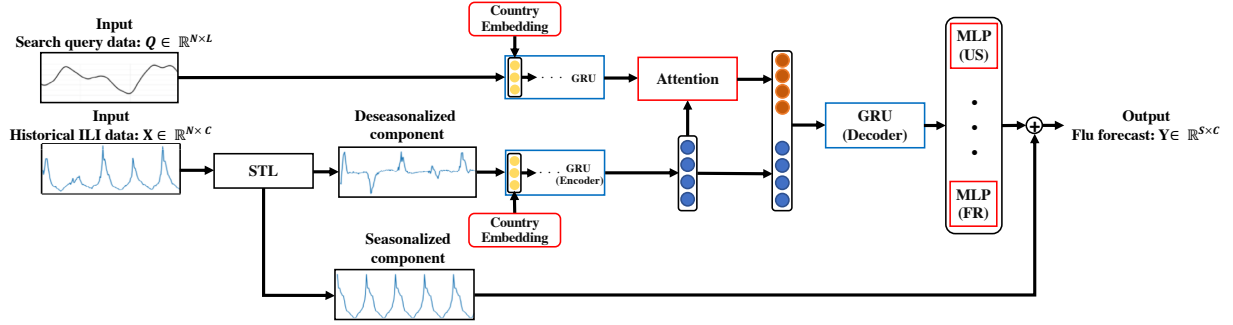


図 2: マルチタスク学習に拡張した提案モデル. 赤枠で囲んだ部分は対象とする国に応じて学習パラメータを変更. 一方で, 青枠で囲んだ部分は共有パラメータを用いる. 更に, 国埋め込みを GRU の初期潜在状態として導入.

獲得する.

$$\mathbf{H}^{enc} = \text{MLP}([\mathbf{H}_t^T \cdot \mathbf{H}^{\tau q}]), \quad (7)$$

$$\mathbf{H}_{i,j}^q = \text{GRU}(\mathbf{Q}_{i,j}), \quad (4)$$

$$i \in \{t - N + 1, \dots, t\}, \quad j \in \{1, \dots, L\},$$

L は検索クエリ数, $\mathbf{H}_t^q \in \mathbb{R}^{L \times M}$ は GRU の最後の潜在表現を表す.

一般的に, Attention 機構はキーとバリューのセット $\{k, v\}$ から Attention の重みを算出し, クエリ q と組み合わせることで, 出力 o にマッピングするように定義する. 各位置 i ごとに, クエリ q_i とキー k_i を内積することで Attention 重みを計算する. 提案モデルでは, 過去の ILI データの非季節性成分の潜在表現 \mathbf{H}_t^T をクエリとして扱い, 検索クエリの潜在表現 \mathbf{H}_i^q をキーとバリューとして扱う. 位置 i は各検索クエリ表現 ($i \in \{1, \dots, L\}$) を指す. クエリ, キー, バリューはそれぞれ以下のように線形変換を通して計算される.

$$\mathbf{S}_q = \mathbf{W}^q \mathbf{H}_t^T, \quad \mathbf{S}_k = \mathbf{W}^k \mathbf{H}_i^q, \quad \mathbf{S}_v = \mathbf{W}^v \mathbf{H}_i^q, \quad (5)$$

$\mathbf{S}_q \in \mathbb{R}^{1 \times M}$ はクエリ表現, $\mathbf{S}_k, \mathbf{S}_v \in \mathbb{R}^{L \times M}$ はキーとバリューの表現となる. これらの表現に基づいて内積計算を用いて, 各クエリ表現の重要度と Attention 表現 $\mathbf{H}^{\tau q}$ を以下のように求める.

$$\mathbf{H}^{\tau q} = \text{Softmax}(\mathbf{S}_q \mathbf{S}_k) \mathbf{S}_v, \quad (6)$$

ここで, $\text{Softmax}(\mathbf{S}_q \mathbf{S}_k)$ は各クエリの重要度を, M は $\mathbf{H}^{\tau q}$ の次元数を示す. その後, Attention 表現である $\mathbf{H}^{\tau q}$ と ILI データの非季節性表現である \mathbf{H}_t^T を結合した特徴を以下のように multi-layer perceptron (MLP) に適用する.

M は \mathbf{H}^{enc} の次元数である.

ILI データの将来の期間 $\{t + 1, \dots, t + S\}$ の値を予測するために, エンコーダの潜在表現 \mathbf{H}^{enc} をデコーダである GRU と 2 層の MLP に適用する.

$$\begin{cases} \mathbf{H}_i^{dec} = \text{GRU}(\mathbf{X}_i^T, \mathbf{H}^{enc}), & i = t + 1 \\ \mathbf{H}_i^{dec} = \text{GRU}(\mathbf{O}_{i-1}^{schedule}), & i \in \{t + 2, \dots, t + S\} \end{cases} \quad (8)$$

$$\hat{\mathbf{O}}_i = \text{MLP}(\mathbf{H}_i^{dec}), \quad i \in \{t + 1, \dots, t + S\}, \quad (9)$$

デコーダの出力である $\hat{\mathbf{O}}_i$ は時間 i における非季節性の値の予測となる. $\mathbf{O}_{i-1}^{schedule}$ は ϵ の確率で真の値を $1 - \epsilon$ の確率でモデルの出力値を入力とするスケジュールサンプリング [26] を適用した値となる. スケジュールサンプリングは, 訓練の出力と真の値の分布の不一致によって引き起こされるオーバーフィッティングで予測性能が低下する問題を緩和する手法である.

将来の ILI レートの予測値は, 提案モデルによる非季節性の予測値 $\hat{\mathbf{O}}_i$ と季節性の値 $\hat{\mathbf{Y}}_i$ を単純に足し合わせることで計算される.

$$\hat{\mathbf{Y}}_i = \hat{\mathbf{O}}_i + \mathbf{X}_i^s, \quad i \in \{t + 1, \dots, t + S\} \quad (10)$$

3.2.3 学習方法

モデルの学習のために, 正解データである非季節成分の真の値 \mathbf{O}_i を決める必要がある. これは, 将来の真の ILI レート \mathbf{Y}_i から, 一定と考えられる季節性成分 \mathbf{X}_i^s を引くことで真の非季節性成分の値が算出される. この真の非季節性成分の値 \mathbf{O} とモデル出力値 $\hat{\mathbf{O}}$ の平均二乗誤差 (MSE) を用いてモデル学習が行われる.

3.3 マルチタスクモデルへの拡張

共通の潜在表現を獲得し表現力の向上を目指すマルチタスク学習を適用することで、提案モデルを各国のインフルエンザ予測が可能なモデルへと拡張する。マルチタスクモデルの全体像を図2に示す。青で囲んだ部分には共通のパラメータを、赤で囲んだ部分では国ごとのパラメータを設定する。具体的には、GRUsに関する式(2), (3), (4), (7), (8)では各国の予測に対して共通のパラメータを設定し、一方で Attention 機構と最終出力にあたる MLP の式(5), (6), (9)では予測対象とする各国特有のパラメータを設定する。

さらに、インフルエンザ流行予測をマルチタスク学習で行うために、非季節性成分の ILI データと検索クエリの2つの GRUs の初期潜在表現を “country embedding” として提案する。これは、より柔軟なモデリングを可能にするために、予測対象の国に応じて初期の潜在表現を変化させることを目的としている、国の埋め込みは以下のように計算する。

$$\mathbf{H}^{\text{country}} = \text{MLP}(\text{Country_id}), \quad (11)$$

$\mathbf{H}^{\text{country}} \in \mathbb{R}^M$ は式(3), (4)の入力として GRU の初期表現を示す。“Country_id” は US は 1, JP は 2 という形で国によって一意の値を割り当てる。各ステップで、国をランダムに選択しその国のパッチデータ $\{\mathbf{X}^{\text{country_id}}, \mathbf{Q}^{\text{country_id}}, \mathbf{Y}^{\text{country_id}}\}$ を作成し学習する。

4 実験

4.1 実験設計

提案モデルの評価として、5カ国 (US, JP, AU, FR, UK) に対して将来の ILI レートを予測する。提案モデルの性能を検証するために、1週目から5週目先までの ILI レートの予測性能を3つのテスト期間 (2017/30週–2018年29週目, 2018年30週–2019年29週目, 2019年30週–2020年29週目) で他モデルとの比較した。テスト期間前の52週間 (1年間) を検証期間とし、ILI データの初週から検証期間の前までの3年以上の期間を訓練期間として設定した。全てのモデルでの検索クエリの選択方法は WT-based に統一し、WT-based と Translation-based の比較については5.3節で述べる。入力の長さとなる N を52週、予測先の長さ S を5週を設定した。検索クエリの数 L を10に WT-based のパラメータ k を100に設定した。提案モデルのハイパーパラメータについては、learning rate は 0.001, 0.01, 0.1, 1.0 から、GRU の潜在次元数 M は 8, 16, 32, 64 の中から検証期間での予測精度を通じて決定した。実験に用いる全てのモデルの学習は勾配ベースの手法である Adam を用いて行い、訓練エポック数は300回 (early stopping による途中停止あり) に設定した。実験では、提案モデルの4つのバリエーションについて検証した。

- **Proposed_single:** 予測対象となる1カ国のデータを用いて学習したモデル (3.2節と同様)。
- **Proposed w/o sq:** 1カ国を対象として、検索クエリを用いず過去の ILI データのみを用いて学習したモデル。

- **Proposed_multi2:** US と JP の2カ国のデータを用いて、マルチタスクの枠組みで学習したモデル (3.3節と同様)。

- **Proposed_multi5:** 5国のデータを用いて、マルチタスクの枠組みで学習したモデル。

US の実験結果を表2に示し、それ以外の国の実験結果を表3に示す。

4.2 評価指標

決定係数 R^2 と、平均絶対二乗誤差 (RMSE) の2つの指標を用いて予測性能を検証した。決定係数 R^2 は予測値が正解値をどの程度説明できているかを表す指標で、値が大きければ良い。一方で、平均平方二乗誤差 $RMSE$ は予測値と正解値の差の大きさを示した指標で、値が小さければ良い。

4.3 比較モデル

GRU: 時間依存性を捉える RNN モデルはインフルエンザ予測 [27] などにも用いられる。本実験の比較モデルとして、RNN の1つである式(2)の GRU モデルを用いる。GRU モデルにエンコーダ・デコーダを適用して、2週間以上先の予測が可能になるように設定した。比較モデルとして、検索クエリと過去の ILI データを入力とする “GRU” だけでなく、過去の ILI データのみを入力とする “GRU w/o sq” を用いた。

Transformer: Transformer は自然言語処理分野で最も使われているモデルの1つである。RNN などのシーケンスモデルと異なり、Transformer モデルは入力の順番を明示的に考慮せず、代わりにシーケンス全体を考慮して、Attention 機構を用いてシーケンスの要素の依存関係を考慮するモデルである。また、Transformer に基づいたインフルエンザ予測モデルによって、高い精度が報告されている [9]。

Multi-task Elastic Net (MTEN): MTEN [14] は検索クエリを入力とするモデルで、アメリカの地域ごとのインフルエンザを予測するために提案された。このモデルは Elastic Net をマルチタスクへと拡張したモデルであり、本実験では US と JP の国を対象としたモデルを作成するために用いる。このモデルは、1週先の予測しかできない構造であり、パラメータと入力データは提案モデルと同様のものを用いる。

4.4 結果

US の実験結果を表2に示す。この結果はマルチタスクモデルなどの我々の提案モデルが他のベースラインと比較して高い精度を達成しており、モデル構造とマルチタスク学習の有用性について確認できた。RNN アーキテクチャによって時間依存性を捉えられる **GRU w/o sq** や **GRU** では、1週間先の予測において R^2 で 0.9 前後の値を達成しており、競争力の高いベースラインの1つとなっている。最高精度を達成したインフルエンザ予測モデルの1つである **Transformer** では近い先の予測 (1週先から3週先) では GRU モデルよりも高い精度を達成しているが、遠い先の予測 (4週先, 5週先) では変わらない精度となっている。これらの結果は、遠い先の予測で精度向上が難しいことを指し示している。統計モデルとマルチタスク学習に基づいた **MTEN** も同様の特徴を有しており、入力が検索クエ

表 2: US におけるモデル予測性能

Term	Model	Input			1-week		2-week		3-week		4-week		5-week	
		Multi	Historical	Query	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²
2017/30th -2018/29th	GRU w/o sq		✓		0.797	0.841	0.925	0.787	1.033	0.734	1.103	0.697	1.150	0.671
	Transformer		✓		0.509	0.917	0.673	0.860	0.903	0.811	1.005	0.744	1.221	0.641
	*Proposed w/o sq		✓		0.392	0.961	0.599	0.905	0.819	0.832	0.984	0.758	1.109	0.695
	GRU		✓	✓	0.783	0.849	0.905	0.791	1.025	0.741	1.097	0.705	1.138	0.654
	*Proposed_single		✓	✓	0.323	0.973	0.558	0.922	0.770	0.849	0.947	0.776	1.078	0.711
	MTEN	✓		✓	0.450	0.934	—	—	—	—	—	—	—	—
	*Proposed_multi2	✓	✓	✓	0.276	0.981	0.550	0.924	0.768	0.853	0.925	0.787	1.038	0.732
	*Proposed_multi5	✓	✓	✓	0.237	0.986	0.498	0.941	0.692	0.837	0.805	0.832	0.942	0.770
2018/30th -2019/29th	GRU w/o sq		✓		0.305	0.945	0.400	0.906	0.451	0.880	0.496	0.855	0.546	0.811
	Transformer		✓		0.263	0.942	0.359	0.917	0.403	0.904	0.451	0.873	0.511	0.843
	*Proposed w/o sq		✓		0.248	0.961	0.323	0.937	0.391	0.909	0.454	0.878	0.525	0.838
	GRU		✓	✓	0.283	0.940	0.371	0.915	0.439	0.887	0.472	0.865	0.528	0.838
	*Proposed_single		✓	✓	0.201	0.976	0.302	0.946	0.378	0.916	0.439	0.886	0.494	0.855
	MTEN	✓		✓	0.429	0.915	—	—	—	—	—	—	—	—
	*Proposed_multi2	✓	✓	✓	0.232	0.968	0.268	0.957	0.323	0.938	0.388	0.888	0.454	0.856
	*Proposed_multi5	✓	✓	✓	0.255	0.963	0.296	0.941	0.369	0.915	0.404	0.877	0.499	0.843
2019/30th -2020/29th	GRU w/o sq		✓		0.698	0.882	0.910	0.807	1.096	0.713	1.153	0.683	1.167	0.648
	Transformer		✓		0.659	0.892	0.919	0.807	1.099	0.712	1.154	0.680	1.218	0.652
	*Proposed w/o sq		✓		0.538	0.932	0.838	0.837	1.084	0.725	1.171	0.683	1.241	0.605
	GRU		✓	✓	0.705	0.870	0.925	0.791	1.108	0.702	1.165	0.681	1.176	0.620
	*Proposed_single		✓	✓	0.469	0.949	0.694	0.881	0.809	0.846	0.863	0.824	0.918	0.799
	MTEN	✓		✓	0.992	0.760	—	—	—	—	—	—	—	—
	*Proposed_multi2	✓	✓	✓	0.409	0.961	0.641	0.904	0.770	0.861	0.840	0.833	0.910	0.802
	*Proposed_multi5	✓	✓	✓	0.370	0.971	0.605	0.920	0.696	0.878	0.787	0.853	0.831	0.841

*は提案モデルのバリエーションを指し示す。太字は各メトリック、各期間における最高精度を示す。

リデータのみだったことから、他のモデルに比べて精度が低い傾向であった。これらの結果からも、深層学習に基づくモデルは計算コストが高いという欠点があるものの、統計的モデルよりも遠い先の予測において高い精度を達成しており、インフルエンザの予測などにおいて有用であるといえる。

これらのベースラインモデルと比較して、提案モデルは各期間、各メトリックにおいて最も高い精度を達成した。この結果は、提案モデルのアーキテクチャがインフルエンザ予測において有用であることを示す。**Proposed_single** はマルチタスク学習なしのモデルの中では最も高い精度を達成している。**Proposed_multi2** と **Proposed_multi5** は **Proposed_single** と比較して精度が向上しており、マルチタスク学習が有用であることを示す。近い先の予測において、マルチタスク学習の効果は時々見られる程度であるが、遠い先の予測においてこれらのモデルの精度は大幅に向上している。例えば、2017年から2018年の期間において **Proposed_multi5** による5週先の予測では、**Proposed_single** の結果と比較して、RMSE では0.136ポイント、R² では0.059ポイントという大幅な向上が見られた。学習で複数国のデータを用いることで、提案モデルはILIデータの潜在的な特徴を獲得し、その結果、予測性能が向上した。一方で、2カ国のデータを用いて学習したモデル (**Proposed_multi2**) と5カ国で学習したモデル (**Proposed_multi5**) による結果はそこまで大きくなかった。実験結果は、期間によってマルチタスクモデルの中でも有用性が異なることを示したものである:2つの

期間においては **Proposed_multi5** が比較的良い精度であった一方で、1つの期間では **Proposed_multi2** が良い結果となった。

5 議論

5.1 複数国に対するマルチタスクモデルの予測性能

表3に4カ国 (JP, UK, AU, FR), 2017年30週目から2018年29週目の期間での3つのモデル (GRU, **Proposed_single**, **Proposed_multi5**) によるインフルエンザ予測性能を示す。この結果は、USの実験結果 (表2) と同様に、ほぼすべての期間でマルチタスク学習モデルである **Proposed_multi5** が最も高い精度したことを示している。これらの実験の結果によって、インフルエンザ予測に対してマルチタスク学習を適用することは、国の数に関わらず言語や環境の異なる複数国への適用が可能であり、予測精度の向上に繋がることを明らかにした。

5.2 検索クエリに関する検証

先行研究 [15] では、インフルエンザの予測における検索クエリの効果は小さいと言われる。本研究では、この問題を解決するために、検索クエリを効率的に考慮する Attention 機構を持ったモデルを提案している。検索クエリを効率的に捉えているかを検証するために、GRUベースのモデルと提案モデルで検索クエリを用いない場合 (GRU w/o sq, **Proposed w/o sq**) から用いた場合 (GRU, **Proposed_single**) に変更したときの改善度を検証した。

表 3: 2017 年 30 週目–2018 年 29 週目の期間における, JP, UK, AU, FR の各モデルの予測精度.

	Model	1-week		2-week		3-week		4-week		5-week	
		RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²
JP	GRU	3.412	0.939	4.019	0.923	5.223	0.915	5.982	0.826	6.164	0.813
	Proposed_single	2.517	0.964	3.218	0.944	3.688	0.934	4.898	0.884	5.822	0.836
	Proposed_multi5	2.429	0.970	2.878	0.951	3.411	0.941	4.057	0.920	4.423	0.905
UK	GRU	1.900	0.910	2.639	0.809	2.738	0.794	2.783	0.787	3.185	0.722
	Proposed_single	1.794	0.912	2.591	0.816	2.959	0.770	2.901	0.741	3.100	0.729
	Proposed_multi5	1.510	0.935	2.199	0.873	2.675	0.808	2.709	0.783	2.992	0.745
AU	GRU	1.754	0.939	2.085	0.914	2.430	0.884	2.739	0.852	3.122	0.807
	Proposed_single	1.764	0.938	2.131	0.922	2.480	0.883	2.683	0.859	3.058	0.816
	Proposed_multi5	1.650	0.942	1.999	0.928	2.391	0.899	2.592	0.879	2.794	0.849
FR	GRU	0.283	0.868	0.427	0.675	0.521	0.517	0.565	0.434	0.587	0.391
	Proposed_single	0.266	0.874	0.413	0.696	0.507	0.542	0.551	0.461	0.560	0.443
	Proposed_multi5	0.234	0.904	0.375	0.751	0.452	0.611	0.527	0.511	0.552	0.466

表 4: JP と FR における Proposed_single モデルでのクエリ選択法の Translation-based と WT-based の比較.

	Method	1-week		2-week		3-week		4-week		5-week	
		RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²
JP	Translation	2.492	0.964	3.307	0.939	3.770	0.929	4.800	0.880	5.976	0.828
	WT	2.517	0.964	3.218	0.944	3.688	0.934	4.898	0.884	5.822	0.836
FR	Translation	0.278	0.856	0.432	0.661	0.531	0.491	0.592	0.407	0.594	0.405
	WT	0.266	0.874	0.413	0.696	0.507	0.542	0.551	0.461	0.560	0.443

表 2 に示すように, GRU w/o sq から GRU に変更することによって, RMSE で平均 0.007 ポイント R² で平均 0.001 ポイントの向上で大幅な改善が見られなかった. 一方で, Proposed w/o sq から Proposed_single の変更では, RMSE で平均 0.091 ポイント R² で平均 0.017 ポイントの大幅な向上がみられた. これらの結果から, インフルエンザ予測において検索クエリを単純に入力するのでは精度向上への寄与が小さい一方で, Attention 機構の導入などによって検索クエリが精度向上に大幅に寄与することを明らかにした.

5.3 検索クエリ探索手法の検討

Translation-based と WT-based の 2 つの検索クエリ探索手法の検討のために, FR と JP における, 2017 年 30 週目から 2018 年 29 週目の期間を対象に比較する. これらの結果を表 4 に示す. FR の全ての評価指標と JP のほとんどの評価指標において, WT-based が Translation-based よりも高い精度を達成することが明らかになった. しかし, 精度の向上の程度は大きくない. 例えば, R² の評価指標において, JP の 2 週先の予測において 0.005 ポイントの向上が, FR においては 0.035 ポイントの向上が見られた. ニューラルネットワークに基づいた提案モデルでは, 既存のマルチタスクモデル [14] と異なり, 効率的な計算のために少量の検索クエリのみを考慮する. 提案モデルのように検索クエリをほとんど考慮しない構造では, 検索クエリの選択の影響が大きく影響を受けないと考えられる. 実

験の結果では, WT-based がインフルエンザ予測のクエリ探索手法として優れていることが示されたが, 用いる検索クエリ数が多いモデルにおいてはどのような探索手法が良いのかなど, 検討の余地があると考えられる.

5.4 Attention による検索クエリの有用性の可視化

提案モデルの Attention 機構は, 検索クエリをうまく組み合わせるだけでなく, どのクエリが予測に有用かを理解することができる. 図 3 は, 2017 年 30 週目から 2018 年 29 週目までの US の予測において, 各検索クエリの注目度を可視化したものを示す. インフルエンザが流行期となっている 2018 年 8 週目頃に注目度に大きな変化が見られた. このことは, 検索クエリが情報が流行期のインフルエンザの予測に有用であったことを示している. 具体的には, “flu and fever” や “symptoms of flu” といった語が大きな重みを持っており, 予測に有用であると考えられる. 一方で, “flu and fever” と同じ意味を持つ “fever flu” には小さい重みが与えられていることは興味深い. Attention を利用して入力の有用性を可視化することは, 有用な検索クエリを決定するのにも役立つ可能性がある.

6 おわりに

本研究では, 複数国を対象としたインフルエンザ予測モデルの構築を試みた. そのために, 英語以外の言語での検索クエリの探索と, 予測モデルの入力として検索クエリデータをどのよ

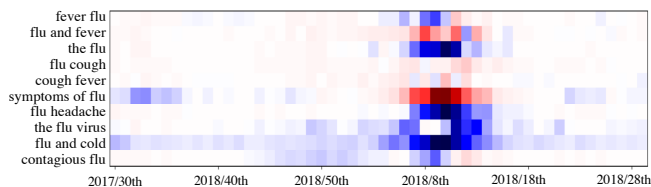


図 3: 2017 年 30 週目から 2018 年 29 週目における検索クエリへの Attention 重みの可視化。赤は大きな重みを、青は小さな重みを示す。

うに活用するかという 2 つの課題を取り組んだ。その結果、英語から他言語への転移手法として WT-based の手法が優れていることを明らかにした。入力データである過去の ILI データと検索クエリデータの特性を考慮した新しい予測モデルを提案し、その有用性を実証した。更に、複数国のインフルエンザ予測実験では、問題設定をマルチタスクとして扱うことで高い性能を達成した。

本実験で、複数国を対象としたインフルエンザ予測モデル構築の可能性と、提案モデルにおける検索クエリデータの有用性を実証された。しかし、適切な検索クエリを探索する方法は未だに大きな課題である。本研究では、英語の検索クエリリストを利用しているが、リソースが存在しない状態から適切な検索クエリを特定する方法が必要である。今後は、本研究で提案した予測モデルを COVID-19 のような新たな感染症に適用し、パンデミックに対応する方法を検討する。

謝 辞

本研究の一部は、JSPS 科研費 JP19K20279 および厚生労働省科学研究費補助金（課題番号：H30-新興行政-指定-004）の支援を受けたものです。

文 献

- [1] Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. (2009) Detecting influenza epidemics using search engine query data. *Nature* 457.7232: 1012-1014.
- [2] Aramaki E, Maskawa S, Morita M. (2011) Twitter catches the flu: detecting influenza epidemics using Twitter. In *Proc. of EMNLP 2011* p.1568-1576.
- [3] Wu N, Green B, Ben X, O'Banion S. (2020) Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case. *arXiv preprint arXiv:2001.08317*.
- [4] Ybarra M, Suman M. (2008) Reasons, assessments and actions taken: sex and age differences in uses of Internet health information. *Health education research* 23.3: 512-521.
- [5] Mavragani A, Ochoa G, Tsagarakis KP. (2018) Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. *J Med Internet Res* 20.11:e270
- [6] Lazer D, Kennedy R, King G, Vespignani A. (2014) The parable of Google Flu: traps in big data analysis. *Science* 343.6176: 1203-1205.
- [7] Lamos V, Zou B, Cox IJ. (2017) Enhancing feature selection using word embeddings: The case of flu surveillance. In *Proc. of WWW 2017*, p.695-704.
- [8] Dalum Hansen N, Mølbak K, Cox IJ, Lioma C. (2017) Seasonal Web Search Query Selection for Influenza-Like Illness (ILI) Estimation. In *Proc. of SIGIR 2017*, p. 1197-1200
- [9] Zou B, Lamos V, Cox IJ. (2019) Transfer Learning for Un-

- supervised Influenza-like Illness Models from Online Search Data. In *Proc. of WWW 2019*, p. 2505-2516.
- [10] Yang S, Santillana M, Kou SC. (2015) Accurate estimation of influenza epidemics using Google search data via ARGO. In *PNAS*. 112.47: 14473-14478.
- [11] Venna SR, Tavaneai A, Gottumukkala RN, Raghavan VV, Maida AS, Nichols S. (2018) A novel data-driven model for real-time influenza forecasting. *IEEE Access* 7: 7691-7701.
- [12] Wu Y, Yang Y, Nishiura H, Saitoh M. (2018) Deep learning for epidemiological predictions. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018*, p. 1085-1088.
- [13] Wang L, Chen J, Marathe M. (2019) DEFISI: Deep learning based epidemic forecasting with synthetic information. In *Proc. of AAAI 2019*, p. 9607-9612.
- [14] 2, Lamos V, Cox IJ. (2018) Multi-task learning improves disease models from web search. In *Proc. of WWW 2018*, p. 87-96.
- [15] Aiken EL, Nguyen AT, Santillana M. (2019) Towards the Use of Neural Networks for Influenza Prediction at Multiple Spatial Resolutions. *arXiv preprint arXiv:1911.02673*.
- [16] Bahdanau D, Cho K, Bengio Y. (2014) Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- [17] Ybarra M, Suman M. (2008). Reasons, assessments and actions taken: sex and age differences in uses of Internet health information. *Health education research* 23.3: 512-521.
- [18] Mavragani A, Ochoa G, Tsagarakis KP. (2018) Assessing the methods, tools, and statistical approaches in Google Trends research: systematic review. *Journal of medical Internet research*. 20.11: e270.
- [19] Mogadala A, Rettinger A. (2016) Bilingual word embeddings from parallel and non-parallel corpora for cross-language text classification. In *Proc. of NAACL-HLT 2016*, p. 692-702.
- [20] Zhou C, Ma X, Wang D, Neubig G. (2019) Density matching for bilingual word embedding. In *Proc. of NAACL-HLT 2019*, p.1588-1598.
- [21] Bojanowski P, Grave E, Joulin A, Mikolov T. (2017) Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*. 5: 135-146.
- [22] Grave É, Bojanowski P, Gupta P, Joulin A, Mikolov T. (2018) Learning Word Vectors for 157 Languages. In *Proc. of LREC 2018*.
- [23] Murayama T, Shimizu N, Fujita S, Wakamiya S, Aramaki E (2020) Robust two-stage influenza prediction model considering regular and irregular trends. *PLoS ONE* 15.5: e0233126.
- [24] Cleveland RB, Cleveland WS, McRae JE, Terpenning I. (1990) STL: A seasonal-trend decomposition. *Journal of official statistics* 6.1: 3-73.
- [25] Chung J, Gulcehre C, Cho K, Bengio Y. (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
- [26] Bengio S, Vinyals O, Jaitly N, Shazeer N. (2015) Scheduled sampling for sequence prediction with recurrent neural networks. In *Proc. of NIPS 2015*, p. 1171-1179.
- [27] Liu L, Han M, Zhou Y, Wang Y. (2018) Lstm recurrent neural networks for influenza trends prediction. In *International Symposium on Bioinformatics Research and Applications 2018*, p. 259-264.
- [28] Lu FS, Hattab MW, Clemente CL, Biggerstaff M, Santillana M. (2019) Improved state-level influenza nowcasting in the United States leveraging Internet-based data and network approaches. *Nature communications*. 10.1: 1-10.
- [29] Ning S, Yang S, Kou SC. (2019) Accurate regional influenza epidemics tracking using Internet search data. *Scientific reports*. 9.1:1-8.