

令和 4 年 5 月 6 日現在

機関番号：12301

研究種目：基盤研究(C)（一般）

研究期間：2019～2021

課題番号：19K10614

研究課題名（和文）Deep Learningによる感染症流行予測モデルの構築に関する研究

研究課題名（英文）Construction of an infectious disease epidemic prediction model by deep learning

研究代表者

内田 満夫（Uchida, Mitsuo）

群馬大学・大学院医学系研究科・准教授

研究者番号：00377251

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：本研究は、特定の地域の過去の感染症データを用い、学習型AIの技術により流行を予測するモデルを作成することを目的とした。感染症のデータは、群馬県衛生環境研究所において2009年から2019年の過去11年分、県に届け出られたインフルエンザ、RSウイルス、咽頭結膜熱、A群溶結連鎖球菌、感染性胃腸炎、水痘、手足口病、伝染性紅斑、突発性発疹、ヘルパンギーナ、流行性耳下腺炎、流行性角結膜炎、マイコプラズマの情報を利用した。最終年をテストデータとし、再帰型ニューラルネットワークの変法のLSTMを用いてモデルを構築した結果、報告数の多いインフルエンザやRSウイルスは高い精度の予測モデルを作成することができた。

研究成果の学術的意義や社会的意義

本研究は、毎年周期的に流行を引き起こす感染症に対し、“その年の流行を予測できれば医療資源の準備や病床の確保を行うための参考情報にできるのではないか？”という発想の下で行われた。本研究の成果より、毎年多数報告されるインフルエンザやRSウイルスは予測精度の高いモデルを構築することができたが、他方、報告数の多くない感染症の予測精度は高くなかった。現在の学習型AIは、学習のために多数のサンプル数を必要とするため、報告数の少ない感染症への対応に課題が残された。この課題は、今後の研究により解決することが望まれる。

研究成果の概要（英文）：The purpose of this study was to create a model to predict future epidemics of infectious diseases using AI technology with past epidemiological data. Data were collected at the Gunma Prefectural Institute of Public Health and Environment for the past 11 years from 2009 to 2019, including influenza, RS virus, pharyngoconjunctival fever, group A streptococcus, infectious gastroenteritis, varicella, hand-foot-and-mouth disease, erythema infectiosum, sudden rash, herpangina, epidemic mumps, epidemic keratitis conjunctivitis, and Mycoplasma. As a result of constructing a LSTM model, we were able to create a highly accurate prediction model for influenza and RS virus which were with enough cases.

研究分野：感染症の疫学

キーワード：感染症 AI RNN LSTM 数理モデル 流行予測

1. 研究開始当初の背景

感染症対策は、一般集団全てが対象となることより、公衆衛生においては重要度の高い分野である。現在の感染症対策は、感染症法に基づく行政的対応に重きが置かれ、定点観測や全数届け出により定量化をおこない、ある閾値を超えた際に注意報や警報を発するという、発生後の迅速な対応が中心となっている。しかし2003年のSARSや2009年の新型インフルエンザの際には、感染症が世界規模の災害となり、資源の確保、病院の体制、事業継続計画（BCP）、封じ込め、感染伝播の予防などの方向性を見立てて対策を立てるといった観点より、迅速な事後対応だけでなく、将来予測の重要性が極めて高まった。

この将来予測の手法を開発するため、近年では世界的に数理モデル研究が推し進められている。この数理モデル研究は、流行早期のデータで将来予測モデルを構築してその後をシミュレートするため、政策などの意思決定への参考情報にもなる。近年では、2014年のエボラ熱の流行ではWHOが数理モデルチームを立ち上げ、その後の予測シナリオを提示し、世界規模の流行を予防したことは記憶に新しい。また動物媒介のZikaウイルス感染症や、デング熱の流行対策にも数理モデルチームが流行の予防に貢献した。しかしこの数理モデル研究は採取したサンプル数などに影響を受け、頑健性や汎化性能においてはまだ弱点があり、さらなる感染症の予測手法の改善が求められている。

他方で、近年の第3次人工知能（AI）ブームにより、多くの事象の評価にディープラーニング（深層学習）が活用されている。深層学習の原理は、蓄積された膨大なデータを用いて、推定関数と観察データの誤差の少ないところを機械学習で探し出す手法である。そこでは、ヒトの予測を超える精度を得ることがあり、近年ではAlphaGoというAI囲碁ソフトが名人に勝利するなど、この分野の躍進が目覚ましい。その技術の中で、再帰型ニューラルネットワーク（Recurrent Neural Network: RNN）は時系列データと相性がよく、これまで気象データの将来予測などで用いられてきた。

以上より、本研究では以下の3点について研究仮説を立てることとした。これまで感染症の将来予測は、論理に基づく数理モデルを構築して最適解を見つける努力をしてきた。しかし深層学習は蓄積データで訓練して得られたモデルから将来予測を行うため、既存研究とアプローチ方法が異なる。AIはヒトが考える論理的なモデルよりも精度の高い予測をおこなうことができるか。これまでの数理モデルでは、複数種類の感染症の互いの影響や、地域や場所という多様性を反映させることが難しく、またサンプル数が少ない場合はその精度が不安定であった。RNNはこの多様性を包含し、サンプル数に影響を受けにくいモデルとなるか。深層学習は過去のデータを用いて訓練する。訓練データが多いほどモデルが安定するが、訓練データに必要とされる最小量は明確な基準が存在しない。それぞれの感染症はどのくらいの蓄積データがあれば頑健な予測精度を示すことが可能か。

2. 研究の目的

本研究では、AIの技術を感染症研究に融合することにより、これまでの伝統的手法では表現することが難しかった感染症の多様性を考慮し、より現実を反映した予測が可能か評価することを目的とした。本研究結果は、感染症の流行動態を忠実に予測するには不可欠であり、また地域における医療資源の配分や、集団の規模に応じた対策を立てるための情報になるなど、今後の感染症対策に寄与する情報になると考えた。

3. 研究の方法

RNNは、ヒトの神経組織ネットワークを模倣して開発された“ニューラルネットワーク”という概念に時間の要素を取り入れたモデルである。このネットワークは、現在のデータ予測に過去の状態を変数として組み込み、さらに現在の状態を次の将来予測の変数として用いる手法である。このRNNは、時系列の並びに規則性やパターンがあると仮定できるデータを訓練により学習し、未知の時系列データが与えられた場合にその未来を予測する。本研究では感染症の流行が規則的な変動を示すことに着目し、このRNNを活用することを計画した。しかしRNNは長期の時系列データの場合、その期間の影響を受けやすいため、本研究ではその変法であるLSTM(Long-Short term memory)のモデルを用いて評価することとした。

群馬県内の衛生環境研究所より2009年から2019年までの感染症届け出のデータを収集した。そのデータをもとにLSTMを構築するため、以下の3ステップで評価した。

1. 群馬県の衛生環境研究所より、研究の実施の同意を得た。本研究では、ヒトの感染症を研究対象としている側面もあるため、群馬大学倫理委員会で承認を得た。
2. 衛生環境研究所から過去の感染症週報データを収集した。感染症の種類は、感染症法に基づく届け出感染症全ての種類とした。対象の感染症は、インフルエンザ、RSウイルス、咽頭結膜熱、A群溶結連鎖球菌、感染性胃腸炎、水痘、手足口病、伝染性紅斑、突発性発疹、ヘルパンギーナ、流行性耳下腺炎、流行性角結膜炎、マイコプラズマとした。
3. 収集したデータから、群馬県の感染症データのデータベース化をおこない、1週ごとに感染症

の種類に応じて流行動態を表現できるようにした。訓練データとして先の10年分のデータを使用し、最終年の1年をテストデータとした。

本研究では、群馬県の感染症疫学データを収集し、感染症毎に予測モデルを構築した。LSTMモデルはPython3とTensorFlowにより構築した。

4. 研究成果

LSTMのモデルは、中間層は12層、学習は10,000回の訓練を行った。その結果、損失関数の値は0.78から0.27まで低下し、このモデルの精度を高めることができた。それぞれの感染症の2019年の流行の観察データと予測データの一致度合いを確認すると、報告数が3桁を超えるインフルエンザやRSウイルスは一致割合が高かったが、報告が2桁の流行性耳下腺炎やマイコプラズマの一致割合は高くなかった。以上より、報告数の多い感染症は予測することが可能であると考えられた。

以下に、感染症の種類ごとにモデルの予測データ（青線）と実際の観察データ（黄線）をグラフにプロットしたものを示す。

RSウイルス



咽頭結膜熱



A群溶血連鎖球菌



感染性胃腸炎



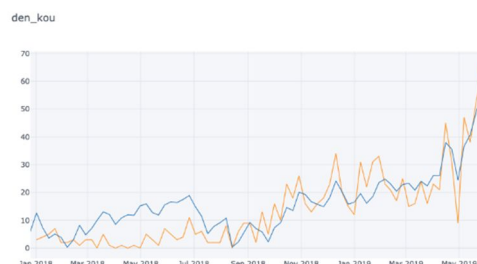
水痘



手足口病



伝染性紅斑



突発性発疹



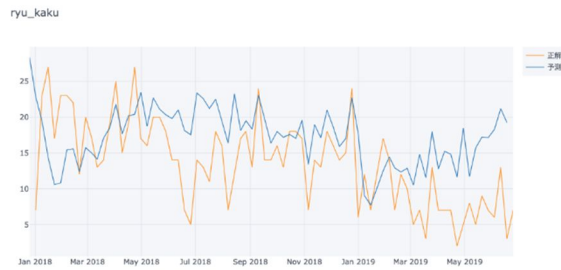
ヘルパンギーナ



流行性耳下腺炎



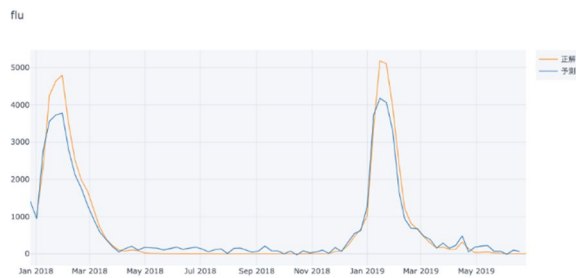
流行性角結膜炎



マイコプラズマ



インフルエンザ



以上、本研究ではAIを用いた感染症の流行予測モデルを構築した。報告ケースが多かった、インフルエンザやRSウイルス感染症は比較的予測データと観察データが一致したが、報告ケースの少なかった流行性耳下腺炎やマイコプラズマは、予測精度は高くなかった。研究背景に示した、少数サンプルでも精度の高い予測ができるかという疑問は、本研究結果からは高い精度を得ることができなかった。今後は、本研究結果を社会に還元してリアルタイムで流行を予測する方法を確立すること、また報告ケース数の少ない感染症の予測精度を高める方法を検討することが必要である。

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 3件）

1. 著者名 内田満夫, 野下浩司, 浅尾高行, 小山洋	4. 巻 3
2. 論文標題 社会医学研究におけるAIの活用例	5. 発行年 2020年
3. 雑誌名 Precision Medicine	6. 最初と最後の頁 74-78
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 内田満夫, 野下浩司, 浅尾高行, 小山洋	4. 巻 2
2. 論文標題 社会医学領域において進めるAI研究	5. 発行年 2019年
3. 雑誌名 Precision Medicine	6. 最初と最後の頁 72-76
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Uchida M, Takeuchi S, Saito MM, Koyama H.	4. 巻 6
2. 論文標題 Effects of influenza vaccination on seasonal influenza symptoms: a prospective observational study in elementary schoolchildren in Japan	5. 発行年 2020年
3. 雑誌名 Heliyon	6. 最初と最後の頁 e03385
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.heliyon.2020.e03385	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Uchida Mitsuo	4. 巻 74
2. 論文標題 Changes in Numbers of COVID-19 Cases among Residents of Tourist Resort Areas before and during the "Go To Travel" Campaign: Descriptive Epidemiology in Gunma Prefecture	5. 発行年 2021年
3. 雑誌名 Japanese Journal of Infectious Diseases	6. 最初と最後の頁 554 ~ 559
掲載論文のDOI (デジタルオブジェクト識別子) 10.7883/yoken.JJID.2021.122	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Uchida Mitsuo	4. 巻 12
2. 論文標題 How has the COVID-19 Pandemic Affected Travelers and Tourist Destinations?	5. 発行年 2021年
3. 雑誌名 Journal of Vaccines & Vaccination	6. 最初と最後の頁 1-2
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 内田満夫, 野下浩司, 浅尾高行, 小山洋	4. 巻 36
2. 論文標題 社会医学研究におけるAIの活用例	5. 発行年 2021年
3. 雑誌名 BIO Clinica	6. 最初と最後の頁 58-62
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

[学会発表] 計3件 (うち招待講演 0件 / うち国際学会 1件)

1. 発表者名 内田満夫
2. 発表標題 群馬県における新型コロナウイルス感染症の流行とその特徴
3. 学会等名 日本公衆衛生学会総会
4. 発表年 2020年

1. 発表者名 Uchida M.
2. 発表標題 Estimation of infectious disease dynamics: Application of a recurrent neural network
3. 学会等名 American Public Health Association annual meeting (国際学会)
4. 発表年 2019年

1. 発表者名 内田満夫
2. 発表標題 群馬県におけるCOVID-19に対する検査陽性率の推移に関する考察
3. 学会等名 日本公衆衛生学会総会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------