

論文の内容の要旨

論文題目 Study on Predicting Influenza Outbreaks
(インフルエンザにおける流行予測に関する研究)

氏名 張捷

インフルエンザは、毎年世界中で流行しており、人々の健康問題の上で大きな問題となっている。その予防や感染対策を的確に行うためには、その流行予測が重要な課題となっている。これまでの多くの研究は、地域ごとのごく短期的な発生の予想に留まっていた。

本論文は、インフルエンザの流行を時間・空間的に分析し、長期、多段階にわたる世界的なインフルエンザの流行を長期に渡って予測を試みているものであり、次の3段階の手順によって分析を行っている。i) 予測のための最善のモデル、タイムラグの期間を見出すこと。このために論文では、6つの異なるモデルについて一連の試行を試みその比較を行った。この結果、Long Short Term Memory (LSTM)モデルにおいて52週までのラグを使ったモデルが最善の予測結果を与えることを見出した。ii) 多段階モデルの推定のための最善のアルゴリズムを見出すこと。4つの異なる多段階モデルの比較を行い、52週のタイムラグを使ったLSTMを用いて、2,3,...,13週先までのインフルエンザデータの予測を行い、Multiple Single-Out Prediction(MSOP)が最善のものであることを見出した、iii) 空間・時間的な広がりをもつインフルエンザ流行の最善な予測を行うシステムを開発すること。まず、152の国・地域のデータを世界保健機構(World Health Organization, WHO)のデータベースであるFluNetから収集した。データに欠損値がなく22カ国のうち人口の多い6カ国を選択し、予測の対象とした。ラグが52週のLSTM、MOSPを使い、1-4週間後の予測を行いその精度の比較を行った。比較にはMean Absolute Percent Error (MAPE)を用いた。その結果、南半球に属する国(オーストラリアおよびブラジル)では、MAPEで比較した予測誤差が、他国の情報を使うモデル方が他国の情報を使わないモデルより大きくなってしまったことを見出した。北半球の国(中国、日本、英国、米国)では、他国の情報を使ったほうが2-4週間のMAPEが他の国の情報を使わない場合に比較して小さくなり、予測精度が向上することを見出した。また、1週間後の予測精度においては、英国以外の北半球において他国の情報を使わない方がMAPEが小さくなり、予測精度が良いことを見出した。

論文では、さらなる試行を行いその理由の考察を行っている。先に選択した22か国の大半は、北半球に存在する。22か国のインフルエンザのこれまでのデータを予測モデルに加えた方がインフルエンザの予測が改良されることを見出した。これは、北半球にある国で

は過去のインフルエンザのデータに高い相関関係があると考えられる。しかしながら、南半球では、他国のデータを使用した場合、南半球のインフルエンザ流行のピークが6.7.8月で北半球の流行とはまったく関係ないため、予測精度が落ちてしまうことを見出した。さらに、北半球においても、1週間後の予測 MAPE が他国のデータを使用しない方が小さくなる理由としては、インフルエンザの流行が広まるためには時間がかかり、タイムラグを生じるためと考えられる。さらに、英国では2017年、2018年において旅行者数が急激に増加している（英国人口の約2/3）ことが、英国におけるタイムラグを打ち消していることを示唆している。同一半球における関係する時間・地理的な要素を考慮することによって、予測モデルを改善し、インフルエンザ流行の世界的な多段階予測の精度の向上に繋がることを結論としている。