採択課題番号：IDEAS202109

AST-GCN モデル・ビッグデータを用いた新型コロナウィルス感染症流行予測

安納住子*、平川翼**、安本晋也**、杉田暁**
*上智大学、**中部大学

1. はじめに

感染症の流行は、気候変動、グローバル化、高齢化や医療資源の配置など様々な問題が絡み合う問題複合体である。新型コロナウィルス感染症(以下 COVID-19)は、SARS-CoV-2 と呼ばれるコロナウィルスを保有する野生動物(自然宿主)から中間宿主を経て、飛沫・接触を介し、ヒトーヒト間で感染する人獣共通感染症である。パンデミック(世界的大流行)は、世界の健康と経済に甚大な影響を及ぼしており、対策が必要とされている。対策の１つとして、ワクチン開発が行われているが、ワクチンの安全性・有効性の懸念から、効率的な介入と予防効果の高い対策の確立が急務となっている。

本研究では、グラフ深層学習および空間的なビッグデータ(デジタルアース)を応用し、COVID-19 のパンデミックに起因する問題複合体の解題と、予防策として COVID-19 流行を高精度で予測するモデルの開発を目的とする。

2. 方法

COVID-19 のサーベイランスデータは Our World in Data(1)から取得し、COVID-19 流行に関与している建造環境要因に関するデータは空港・航空路データ(OAG)(2)、駅・鉄道路線データ(Natural Earth)(3)、国境の形状データ(国地域同士が隣接している場合、道路による接続性があると判断した)(Natural Earth)(3) から取得した。これらのデータを用いて、Graph Convolutional Neural Network(GCN)(4)(5)(6)(図 1)に入力するノードとエッジからなるグラフを構築し、2 種類の行列で表した。GCN では、隣接行列：国地域同士の隣接関係を表す行列と、特徴行列：各ノードの特徴ベクトルを表す行列の 2 つを用いた。各ノードの特徴量は、国地域ごとの COVID-19 新規感染者数 7 日間のデータを与えた。

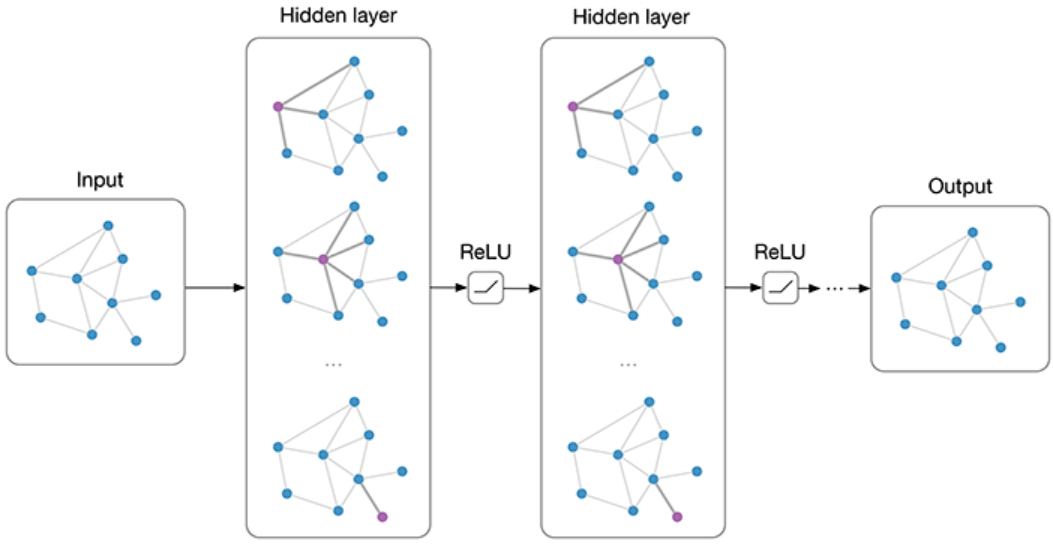

図 1. Multi-layer Graph Convolutional Network (GCN) with first-order filters. Source: https://tkipf.github.io/graph-convolutional-networks/

入力の隣接行列を単独あるいは組合せて、7 日間＋1 日先の COVID-19 新規感染者数の予測を 7 通り行い、また、Mean Squared Error(MSE)を用いた予測精度の評価を行った。さらに、GCN の有効性を確認するため、Multilayer Perceptron(MLP)をベースライン手法として、予測精度の比較を行った。データセットは、60%を

訓練用、40%をモデルの精度評価用に分割した。モデルの訓練は 500 エポック、1 エポックあたり 256 バッチで行った。この実験では、Adam optimizer を使用し、学習率は 0.01 とした。この学習モデルをテストデータセットに適用し、モデルの精度を評価した。なお、初期値の影響などから実験結果が安定しないため、GCN の実験では 5 回の実験、MLP の実験では 14 回の実験をそれぞれ行い、平均値と標準偏差をそれぞれ算出した。

3. 結果

MSE に関して、GCN および MLP ともに予測精度が高かった(表 1)。また、GCN の感染者数の平均誤差は、MLP と比較して約 10 倍の差がみられ、GCN の予測精度がベースライン手法より高かった(表 2)。さらに、入力の隣接行列を単独、特に、航空路データを使用することが、深層学習の精度向上に有効であることが確認できた。

表 1. 正規化予測値の MSE.

Model	Connection	Each experiment					Mean	SD
GCN	Airway	6.92E-05	7.00E-05	6.37E-05	5.35E-05	5.93E-05	6.31E-05	6.93E-06
	Rail	8.14E-05	7.74E-05	9.90E-05	7.61E-05	8.47E-05	8.37E-05	9.20E-06
	Road	8.86E-05	6.82E-05	7.76E-05	8.39E-05	7.27E-05	7.82E-05	8.24E-06
	Air/Ra/Ro	1.27E-04	3.67E-04	2.05E-04	1.65E-04	1.68E-04	2.06E-04	9.38E-05
	Air/Ra	8.33E-05	1.72E-04	9.01E-05	1.06E-04	7.47E-05	1.05E-04	3.90E-05
	Air/Ro	7.62E-05	7.33E-05	7.94E-05	8.09E-05	1.02E-04	8.24E-05	1.15E-05
	Ra/Ro	2.08E-04	9.81E-05	1.56E-04	1.35E-04	1.04E-04	1.40E-04	4.45E-05
MLP		7.89E-04	4.92E-04	2.93E-03	8.41E-04	7.19E-04	1.11E-03	8.67E-04
		5.95E-04	5.10E-04	1.96E-03	4.93E-04	1.96E-03		
		4.94E-04	6.50E-04	4.86E-04	2.63E-03			

表 2. 感染者数の予測値と実測値の差の MSE.

Model	Connection	Each experiment					Mean	SD
GCN	Airway	28.64	29.00	26.40	22.16	24.54	26.15	2.87
	Rail	33.71	32.04	41.01	31.53	35.06	34.67	3.81
	Road	36.69	28.22	32.12	34.75	30.13	32.38	3.41
	Air/Ra/Ro	52.50	151.82	84.91	68.22	69.45	85.38	38.87
	Air/Ra	34.48	71.16	37.31	43.87	30.93	43.55	16.15
	Air/Ro	31.55	30.35	32.87	33.52	42.38	34.13	4.77
	Ra/Ro	86.15	40.62	64.45	55.77	43.21	58.04	18.44
MLP		326.89	203.76	1213.24	348.37	297.89	460.02	358.99
		246.55	211.08	811.21	204.34	811.21		
		204.52	269.19	201.42	1090.67			

図 2 および 3 は、2022 年 1 月 20 日時点、累積感染者数および死亡者数の上位 8 地域/国における入力の隣接行列を航空路とした GCN モデルと MLP モデルによる予測値（オレンジ色）とグランドトゥルース値（青色）を示している。GCN モデルは予測値とグランドトゥルースとの間の汎化ギャップが小さく(図 2)、対して MLP モデルは予測値とグランドトゥルースとの間の汎化ギャップが大きかった(図 3)。

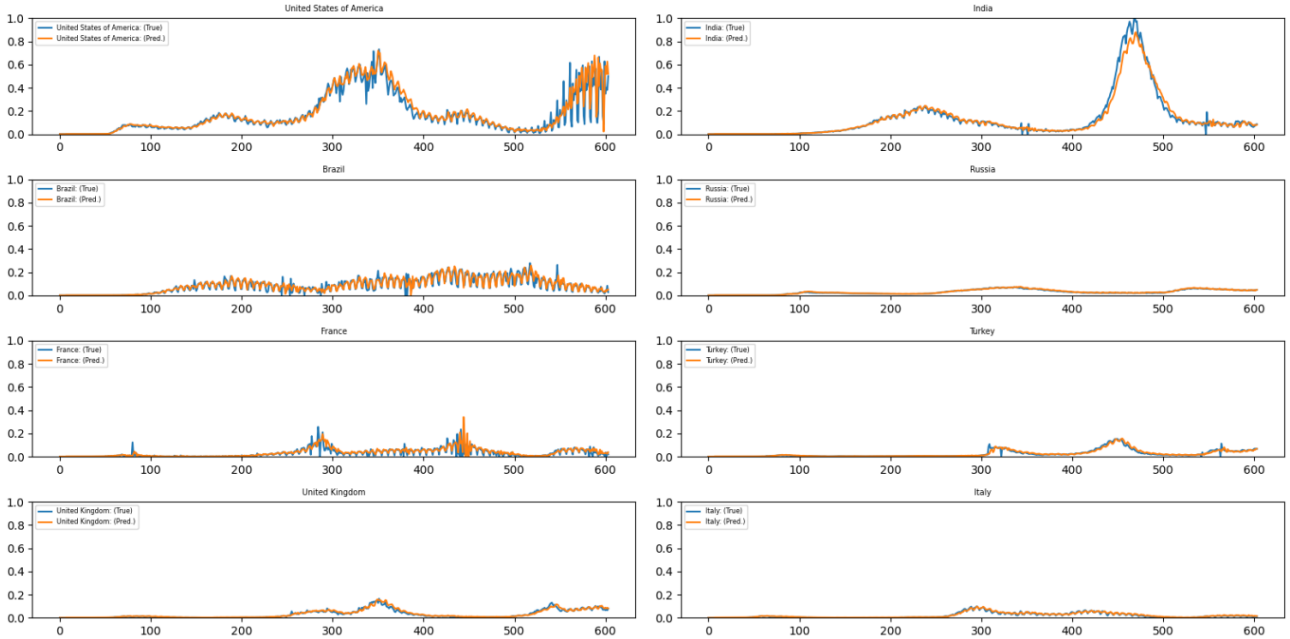

図 2.　入力の隣接行列を航空路とした GCN モデルの予測値とグランドトゥルース値.

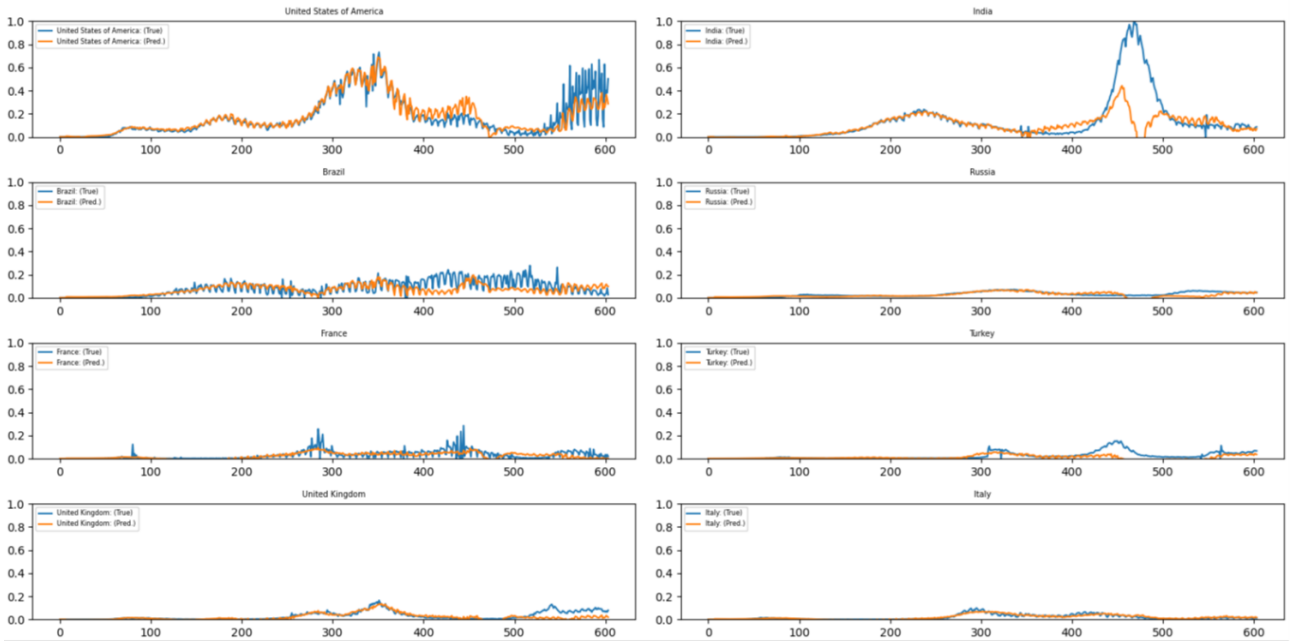

図 3.　入力の隣接行列を航空路とした MLP モデルの予測値とグランドトゥルース値.

4. 考察

公共交通機関による人流を考慮した GCN モデルの方がよい結果となった。GCN モデルの特徴である予測計算に合流先の情報を加味できたことが、精度向上につながったと考えられる。COVID-19 の発生地と考えられている武漢は、国内高速鉄道の中心地となっているほか、国際空港は主要なハブ空港となっており、世界 60 以上の空港と航空便を結んでいる。武漢は公共交通のハブであることが、パンデミックの 1 つの原因になったと考えられる。

5. まとめ

データプーリングによるグラフベースの深層学習は、COVID-19 パンデミックに対する公衆衛生対応のためのデジタルヘルス・ソリューションとして期待できる。また、早期警戒システムに深層学習技術を組み込むことで、効果的な警戒システムの実現や、感染リスクの高い場所を特定する地図の作成など、パンデミックの可能性を持つ新興・再興感染症への適切な対応を導くことも期待できる。

6. 謝辞

本研究は中部大学問題複合体を対象とするデジタルアース共同利用・共同研究 IDEAS202109 の助成を受けたものです。

参考文献・データ

1. Our World in Data, https://ourworldindata.org/covid-cases.
2. OAG, https://www.oag.com/ja/airline-schedules-data
3. Natural Earth, https://www.naturalearthdata.com/.
4. Li D, Gao J. Towards perturbation prediction of biological networks using deep learning. Sci Rep. 2019 Aug 16;9(1):11941. doi: 10.1038/s41598-019-48391-y. PMID: 31420588; PMCID: PMC6697687.
5. Kipf TN, Welling M. Semi-Supervised Classification with Graph Convolutional Networks. 2016 arXiv preprint arXiv:1609.02907.
6. Graph convolutional networks, https://tkipf.github.io/graph-convolutional-networks/.