

情報可視化論最終レポート

216x223x 藤田啓二郎

1.はじめに

新型コロナウイルス感染症（COVID-19）の拡大は今もなお続いており，また，最近では変異株により，死亡率・重症化率が高いことが問題となっている．今回は，厚生労働省の発表している，日本国内における COVID-19 の新規陽性者数，重症者数，新規死亡者数の 1 日ごとのオープンデータ[1]（2020/02/14～2021/06/10）から，感染拡大の変遷を概観し，変異株による影響の考察を行う．

2.手法

それぞれのデータを折れ線グラフとして表示し，オプションとしてデータを新規陽性者数に対する重症化率，死亡率及び重症者数に対する死亡率が高い順にソートして表示する．その下に，元データ（ 480×3 ）を umap[2]によって 2 次元に次元削減したデータ（ 480×2 ）を散布図として表示し，クラスタを見て分析の助けとする．点の重なりがわかりやすいよう，散布図の点は半透明でプロットしている．

3.結果

折れ線グラフと散布図は図 1 のようになった．

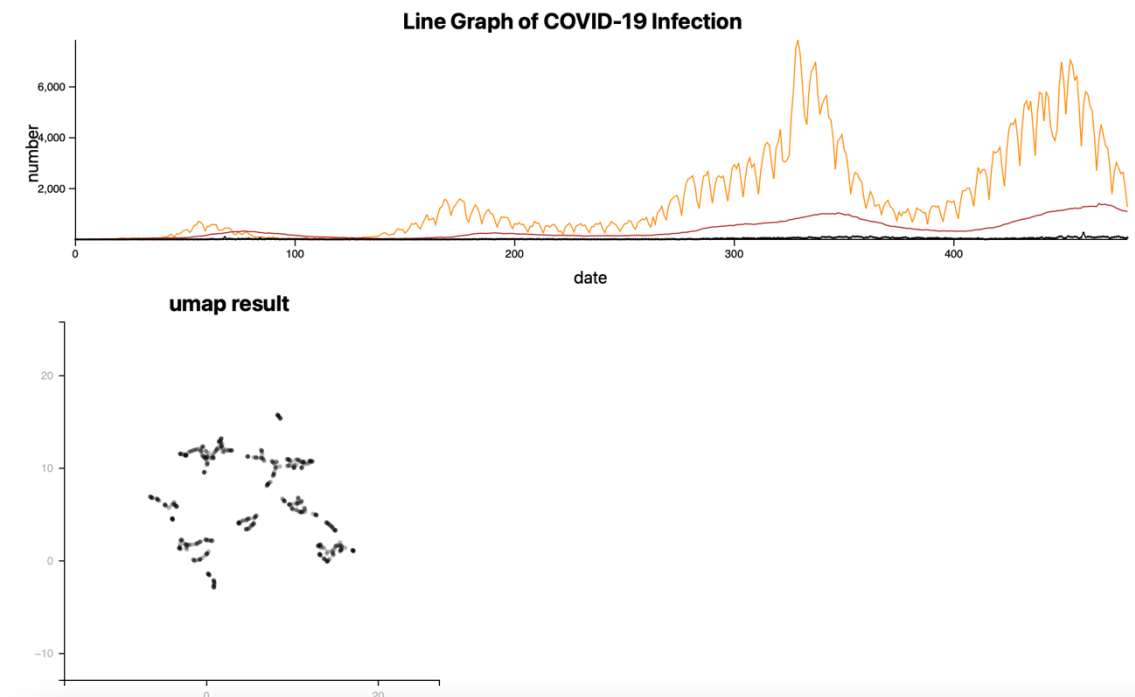


図 1：各データの推移を示す折れ線グラフ（上）と umap の結果を示す散布図（下）

黄色が新規陽性者数，赤色が重症者数，黒色が死亡者数を表す．

4.考察

図1の折れ線グラフの，2020年2月中旬から3月半ばまではどの数字も小さく，この部分は下の散布図の左下のクラスタに現れている．

上の折れ線グラフを新規陽性者数に対する死亡率が高い順にソート（sort1操作）すると，2020年5月半ばから下旬にかけて，高くなっていることがわかった．これらは，下の散布図の中央のクラスタに現れている．これは，陽性者数が非常に少なかったためと考えられる．また，2021年2月中旬から3月にかけて，この値が高くなっている日が多いことがわかった．このあたりの陽性者数はそれなりに多いので，大きな特徴を持つと考えられる．

上の折れ線グラフを重症者数に対する死亡率が高い順にソート（sort1操作）すると，2021年2月半ばから3月にかけて，この値が高くなっている日が比較的多いことがわかったが，ばらつきが多く，良い比較指標とはならなかったことがわかった．

上の折れ線グラフを新規陽性者数に対する重症化率が高い順にソート（sort3操作）すると，sort1操作と同様に5月半ばから下旬にかけて，高くなっていることがわかった．また，この操作から，2020年2月から6月にかけては，感染者数が非常に少ないために，感染者数に対する重症化率は高くなっていることがわかった．

5.結論

新規陽性者数に対する死亡率，重症化率だけで判断すると，新規養成者数が少ない2020年2月から6月にかけての方が高いことになってしまったが，それらを除外すると，2021年2月半ばから3月にかけてが，新規陽性者に対する死亡率が高くなっていることがわかった．また，このあたりは重症者数に対する死亡率が高い部分でもあることがわかった．

sort1操作やsort3操作によって，単純にそれぞれの値の大小を見るだけではわからない特徴を持つ日付クラスタがあることがわかり，また，umapによる次元削減結果の可視化結果にそのクラスタが含まれていたことから，この散布図の他のクラスタを分析することでも特徴のある日付クラスタを発見できることが期待される．

6.参考

[1] 厚生労働省新型コロナウイルスオープンデータ，
<https://www.mhlw.go.jp/stf/covid-19/open-data.html>

[2] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Grossberger. Umap: Uniform manifold approximation and projection. The Journal of Open Source Software, Vol. 3, No. 29, p. 861, 2018.