

Connected Scatterplot を用いた新型コロナウイルス感染状況の可視化

窪田 翔[†] 北 直樹[‡] 斎藤 隆文[‡]

東京農工大学 工学部情報工学科^{†‡}

1. 背景と目的

新型コロナウイルスの感染状況の可視化は多くの人によって行われているが、ほとんどの可視化は一つの時系列データのみを対象としており、二つの時系列データを同時に可視化しているものは少ない。あえて二つの時系列データを同時に可視化する場合には二つの時系列データにそれぞれ y 軸を用意する Dual axis line graph (以下 DALC と呼ぶ) が用いられる。だが、この可視化手法はそれぞれの y 軸のスケールを変化させるとグラフの形状が変化するため、ユーザに誤解を招く恐れがある。そこで本研究では新型コロナウイルスの感染状況の二つの時系列データの関係性を Connected Scatterplot (以下 CS と呼ぶ) [1] を用いて可視化する。この可視化手法は、二つの時系列データをそれぞれ x 軸、y 軸に割り当てて、各点を時系列順に繋ぐことで得られる。DALC と CS の例を図 1 に示す。

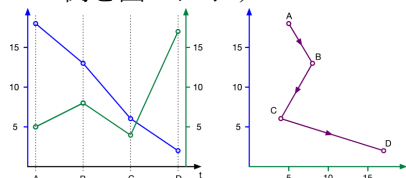


図1 DALC (左) と CS (右) の例 (文献[1]から引用)

CS は DALC と比較して線が一本であるため、重ねたときの読み取りづらさが軽減する。その点に着目して CS の重畳表示を試みた。この手法によって、新型コロナウイルスの感染状況の地域間のフェーズの違いを可視化できると考えている。

2. Connected Scatterplot の特徴

ここでは、CS の特徴について説明する。

2.1. 線の角度

x, y に割り当てた時系列データのどちらが速く変化しているかによって、線の角度が変わる。 Δ_x を x 軸、 Δ_y を y 軸の単位時間あたりの時系列データの変化とすると、その対応は図 2 のようになる。

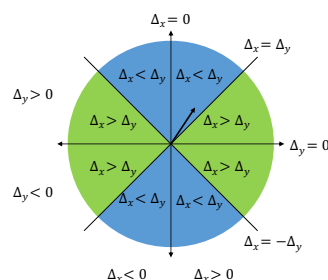


図2 線の方角と時系列データの変化の関係

例として、図 2 の黒い矢印だと $\Delta_x < \Delta_y$ であるから、y 軸の時系列データの変化の方が大きいことが読み取れる。

2.2. ループ

CS の形状としてループがある。CS がループを形成する時の CS と DALC の例を図 3 に示す。ループは時系列データのシフトを示している。つまり、極大値と極小値の組がそれぞれの時系列データで異なった時間で見られるときにループが形成される。また、ループの回転の向きによって、極大値と極小値の組がそれぞれの時系列データで見られる順番が異なる。

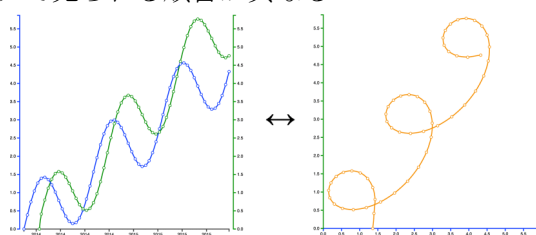


図3 ループ形成時の DALC (左) と CS (右) (文献[1]から引用)

Visualization of COVID-19 Using Connected Scatterplot

[†]Kakeru KUBOTA, Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology.

[‡]Naoki KITA, Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology.

[‡]Takafumi SAITO, Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology.

3. 提案手法

CS の重畳表示を見やすくするために、以下の手法を施した。

3.1. 進行方向の明示

CS は進む方向が決まっていなかったためユーザに進行方向を示す必要がある。図 1 の右の CS は進行方向を示すために矢印を使用している。だが、重畳表示にすると分かりづらくなってしまった。そこで、色のグラデーションを利用して、進行方向に対して色が濃くなっていくようにした。

3.2. 移動平均

日々のデータは変化の上下が大きいので、そのまま重畳表示すると読み取りづらくなってしまった。そのため、移動平均を利用して平滑化した。

3.3. 対数スケール

複数の CS を集約させて見やすくするために、x, y 軸ともに対数スケールにした。

3.4. プロット

全てのデータをプロットしてしまうと、乱雑になって見づらくなってしまった。そこで一週間ごとにプロットすることにした。またこれによって多少ではあるがどの期間であるかを提示することができる。

4. 可視化

ここでは、厚生労働省の新型コロナウイルス感染状況のオープンデータ[2]を用いて、県ごとの日々の感染者数と重症者数を CS の重畳表示と DALC の重畳表示によって可視化して比較した。表示する期間は短くして、複雑に重なり合っている読み取りづらくなるのを避けた。可視化の例を図 4 に示す。

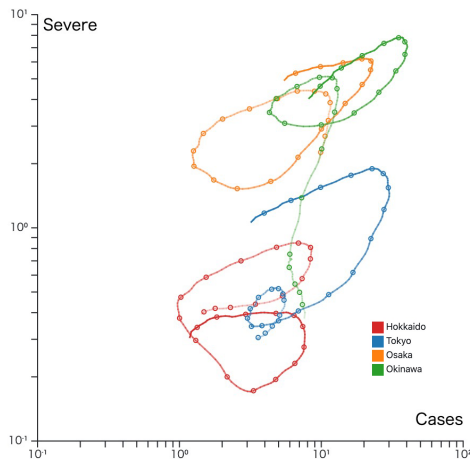


図 4 CS の重畳表示

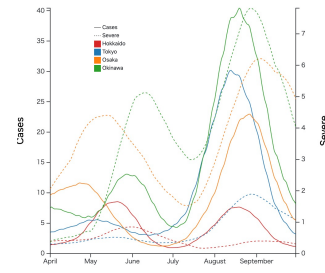


図 5 DALC の重畳表示

データは北海道、東京、大阪、沖縄の 2021/4/1～2021/9/30 までの人口 10 万人あたりの感染者数、死亡者数を扱う。CS の x 軸が感染者数、y 軸が重症者数で、DALC の左の y 軸と実線の折れ線が感染者数、右の y 軸と点線の折れ線が重症者数で、赤が北海道、青が東京、オレンジが大阪、緑が沖縄を表している。移動平均は 30 日とっている。図 4 から、CS の重畳表示を見ると、どの県も感染者数の推移に差はほとんど見られないが、重症者数は大阪と沖縄が高いところで推移しているのに対して、北海道と東京は低いところで推移している。このように相対的な位置関係からそれぞれの県がどのぐらいの規模で推移しているかが読み取れる。また、どの県も二つのループを形成しているが、それらを比較すると北海道のみ一つ目に対して二つ目のループの重症者数が低いところで推移していることがわかる。そして、線の角度を見ると前半の期間で沖縄の重症者数が感染者数に対して急激に上昇していることが読み取れる。これらの情報は図 5 の DALC の重畳表示では読み取りづらい。

5. おわりに

Connected Scatterplot の重畳表示を試みた。感染状況の地域間のフェーズの違いまでは読み取ることができなかったが、既存の手法では読み取りづらい情報を素早く読み取ることができたことがわかった。この結果はあくまで筆者自身の主観であり、ユーザスタディを行って定量的に比較する必要がある。

参考文献

- [1] Steve Haroz, Robert Kosara, Steven L. Franconeri, IEEE Transactions on Visualization and Computer Graphics Vol. 22, No. 9, pp. 2174–2186 (2016)
- [2] オープンデータ, (最終閲覧日: 2022/1/7), <https://www.mhlw.go.jp/stf/covid-19/open-data.html>