時系列データに対する説明的可視化

Explanatory Visualization for Time Series Data

牧 修平○ (東京工業大学）

##### ABSTRACT

This template is a guide to prepare manuscript for Visualization Symposium Japan. Here, please write the abstract serving as an index and as a summary of the present paper. It should be as long as approximately 150 words.

データを理解しやすくするために様々な可視化手法が使われているが、データ構造の複雑さやデータの多さなどの要因のせいで、意図する情報を全てユーザが可視化から認識することは難しい。そこで可視化に加えてデータから文章を生成し、データに含まれる有益である可能性のある情報を、前もってユーザに提供するような研究が行われている。本論文のシステムでは、多次元の時系列データに関してデータの分布や変動の分析をサポートする可視化とともに、そのデータからテンプレートベースで生成した要約文を提供する。論文中では、時系列データの類似度の計算方法や時系列データを可視化した折れ線の形からデータを探索する方法、要約を生成するテンプレートの構成について説明する。そして最後にシステム全体をSocial Progress Indexというデータを用いたユースケースを通して説明する。

to English

# はじめに

生のデータを見るだけでは分析したり理解したりするのが難しいデータに対して、その分析や理解のために可視化を行う研究はたくさん行われてきた(Gschwandtner and Erhart 2018; Bryan, Ma, and Woodring 2016)。 しかし可視化によるデータの分析や理解に慣れていないユーザ、可視化だけでは理解が難しいデータを解析したいユーザがいるのため、データの分析や理解のために必ずしも可視化で満足させれるわけではない。 そこで文章と可視化によってデータの説明をする**説明的可視化**が使用されてる(Hosokawa, Arimoto, and Wakita 2020; Latif and Beck 2019; Ynnerman, Löwgren, and Tibell 2018)。

時系列データは日常の気温や降水量、近年では新型コロナ感謝数の推移など様々な場面で扱われているデータである。 時系列データは必ずしも同様の事柄に対して一つのデータだけではなく複数のデータが存在することがある(新型コロナ感謝数の推移など)。 このような場合、時系列データは可視化による理解が困難になるため、説明的可視化を使ってデータの説明をするのが望ましい。 しかし今まで説明的可視化を使ってこのようなデータを説明する研究は行われてきていない。 そこで本研究では[Social Progress Imperative](https://www.socialprogress.org/)が10年にわたって発表しているSocial Progress Indexのデータ(Social Progress Imperative n.d.)(以下SPIデータ)を例として、説明的可視化による時系列データの理解を手伝う手法を提案する。

2章では説明的可視化と時系列データの可視化について関連研究をまとめ、3章でシステムの設計について述べる。 4章ではシステムのユースケースを通してシステムの有効性を説明し、5章と6章で考察とまとめを行う。

# 関連研究

時系列データについての説明的可視化を行う上で、今まで研究されてきた説明的可視化と時系列データに対する可視化をみていく。

**説明的可視化** Latifたちは(Latif and Beck 2019)、二つの地理統計変数を理解するためにiMRというシステムを開発し説明的可視化を行っている。 地図の領域の色やその領域の上にに描かれる円の大きさを使って二つの地理統計変数を可視化するとともに、それらのデータを解析して文章を自動生成している。 図は各州の年間の死亡者数、ハリケーンの発生数に対して実際にiMRのシステムを使用した例である。 生成されている文章には、死亡者数の平均値や最大値・最小値など、統計的なデータをはじめ、「南部の州が他の週より死亡者数やハリケーンの発生数が多い」などデータからは読みとりづらい情報も組み込まれている。 iMRでは文章生成を行うプロセスを自身で定義し、ユーザーの要望に合わせてパラメータを変更することで柔軟な文章を生成している。 本研究ではデータからは読み取りづらい情報を文章にしてユーザーに伝えるために、この論文の文章生成プロセスを参考に文章を生成するプロセスを定める。

Hosokawaたちは(Hosokawa, Arimoto, and Wakita 2020)、多変数の地理的統計量に対して指標の階層構造や地域的な階層構造に着目して説明的可視化を行っている。 HosokawaたちはSPIデータを使用しているがある年についてのみ注目して、時系列ではない二つの階層構造が含まれるデータとしてSPIデータを扱っている。 本研究では時系列ということに注目してシステムを作る。 またHosokawaたちとはことなり、地理的階層構造についてのみに着目しており、SPIの指標の階層構造についての説明的可視化は行っていない。

**時系列データの可視化** 時系列データの情報を伝えるために可視化を使った研究はいくつか存在しており(Bryan, Ma, and Woodring 2016)、その中でJunhuaたち(Lu et al. 2020)は、二つの地理統計変数データを二変数マップと動画を用いて可視化している。 図では横軸を収入、縦軸を平均寿命とした例で、ドットの動きによって過去100年で世界各国の収入や平均寿命がどのように変化していったかが分かるようになっている。 この研究では、時系列データを事前に解析し時系列データの変化の傾向が変化した期間を特定して、その期間を下のビューでユーザーに伝えている。 本研究ではこの研究を参考に、タイムステップ数が10と、データが少ない時系列データに対して傾向の変化があるタイミングを特定し、可視化と文章に傾向の変化の情報を取り入れる。

# システムの設計

時系列データの分析のための可視化を表示しつつ、情報を手に入れたい地理的範囲や指標の選択に応じた情報を文章としてユーザに提供するためにシステムを設計した。 このときSPiデータを時系列データとして分析するときに、ユーザーは地理的な階層(東アジアの中の日本やヨーロッパの中のイギリスなど)の情報を含んでデータを分析したいと仮定している。 しかしSPIデータには地理的な階層の情報が含まれていないので、[国際連合](https://www.un.org/)が提供している、地理的な階層についてのデータ(UNSD 2003)を追加で扱う。

## 要件

実際にシステムを作る前に、以下のように要件を定める。

**R1 変化の特定** 指標の値が極端に低い国に注目するときに表示する地理的範囲が広い(表示する国数が多い)と、注目する国のデータが潰れて表示されてしまい、そのデータの分析は困難になる。 そこで10年の時系列データの中で傾向の変化があった年を特定して、データのの概要としてユーザーに提供する必要がある。

**R2 階層的な情報** 生データをそのまま分析するとき、個々の国についての情報を分析することはできないが、その階層的な見方はなかなか難しい。

**R3 地理的範囲の表示** 地理的階層の情報を伝える際にどの国がその範囲に含まれているかをユーザーが知っている必要がある。 しかし、ある範囲に含まれる国を全てユーザーが知っていることはほとんどない。 そのため、選択する地理的範囲を資格的にユーザーに提供する必要がある。

**R4 概要** 複数のデータに対し、時系列データの変化や地理情報、階層的な情報などを視覚的な表現によって理解することは困難である。 そこでその表示しているデータの概要を、文章によってユーザーに提供する必要がある。

## データの可視化

### 時系列の可視化

時系列データの可視化は折れ線やヒートマップを使って行うのが一般的である。 本論文では折れ線を用いてSPIデータの時間的変化を表現する。 しかし注目する地域的範囲が広いときに、ほとんどの折れ線では時間的な変動を認識することができない。 注目する国の概要やその国データと類似した国データをハイライトする。 注目する国に関してはデータの傾向が変化する点をドットを用いて表現しつつ(**R1**)、色を他の表現と異なるマゼンタ色を使う。 注目する国と類似する国のデータに関してカーキ色を使う。

**傾向の変化する点** データの傾向の変化する点の求める方法ははJunhuaたち(Lu et al. 2020)と同様にWin法(Truong, Oudre, and Vayatis 2020)を適用する。 時系列データをとして初めに、に対しての線形回帰直線の傾きを求め、に対しての線形回帰直線の傾きを求める。 次に、に対してを求める。 そして両方ともに対して、が注目する指標に対するSPIデータの平均値を元にした閾値以上であるとき、の点をデータの傾向が変化した点として捉える。 これは、ある点をを中心にして左右の直近のデータの傾きの差が大きい点を傾向の変化があった点としている。 は短い時間間隔でのデータの変化を捉えることができ、は少し長い時間間隔でのデータの変化を捉えることができる。

**類似する国** 類似する国を定義する。 初めに注目する国と他の国の時系列データで相関係数が大きい国を集める。 その集合の中で、注目する国と他の国データでユークリッド距離を計算する。 その距離が注目する指標に対するSPIデータの平均値を元にした閾値未満であれば、類似する国とする。 相関係数は、二つの時系列データの上昇や現象などの大まかな変動が一致していることを条件にするために使用している。 その後時系列データを次元ベクトルと考え、ユークリッド距離が近いデータを選択することで、注目する国と値が近く似た変動をする国を抽出することができる。

## 地域の可視化

選択された地域的範囲をユーザーが難なく理解できるように、世界地図を用いてその範囲を表示する(**R3**)。 選択された地理的範囲に含まれていない国々と、データが存在しない国は灰色で表される。 注目する国については、折れ線と同様の色で表現する。 これらによって、ユーザーが注目する国と地理的範囲の国々とで地理的比較が可能になっている。

## 文章生成

## インタラクション

ユーザーがより可視化からデータに関する情報を手に入れやすいようにインタラクションを導入する。

**ズーム** 折れ線の表示において、注目したい国データの変動をより詳細にみたいときのために、ボタンによって縦軸の幅を調整できるようにしている。その他にも、選択した地理的範囲がアジアや南アメリカなど世界全体ではないときに、その選択した地理的範囲のデータだけを表す縦軸にするか世界のデータの範囲を縦軸にするかを調整するボタンも準備している。 選択した地理的範囲を見ることは、世界から比べて極端に値が小さい、または大きい国しかその地理的範囲に存在しないときに、その国々のデータを確認するのに役立つ。 一方世界のデータの範囲を縦軸にすることで、世界から比べて極端に値の低い、または大きい国がその選択した地理的範囲に含まれていることなどを知ることができる。

**フィルタリング**

**クエリ**

**関連付け**

**パラメータ選択**

ぼたん 折れ線/国選択

## システム

# ユースケース

ユースケース

# 考察

考察

# まとめ

まとめ

# 参　考　文　献

Bryan, Chris, Kwan-Liu Ma, and Jonathan Woodring. 2016. “Temporal Summary Images: An Approach to Narrative Visualization via Interactive Annotation Generation and Placement.” *IEEE Transactions on Visualization and Computer Graphics* 23 (1): 511–20.

Gschwandtner, T., and Oliver Erhart. 2018. “Know Your Enemy: Identifying Quality Problems of Time Series Data.” *2018 IEEE Pacific Visualization Symposium (PacificVis)*, 205–14.

Hosokawa, Natsuki, Kohei Arimoto, and Ken Wakita. 2020. “A Scalable" Exploranation" Technique for Hierarchically Indexed Table Data.” In *Proceedings of the 13th International Symposium on Visual Information Communication and Interaction*, 1–5.

Latif, Shahid, and Fabian Beck. 2019. “Interactive Map Reports Summarizing Bivariate Geographic Data.” *Visual Informatics* 3 (1): 27–37.

Lu, Junhua, Jie Wang, Hui Ye, Yuhui Gu, Zhiyu Ding, Mingliang Xu, and Wei Chen. 2020. “Illustrating Changes in Time-Series Data with Data Video.” *IEEE Computer Graphics and Applications* 40 (2): 18–31.

Social Progress Imperative. n.d. “Social Progress Imperative.” <https://www.socialprogress.org/>. Accessed: 2020-10-10. Accessed October 10, 2020. <https://www.socialprogress.org/>.

Truong, Charles, Laurent Oudre, and Nicolas Vayatis. 2020. “Selective Review of Offline Change Point Detection Methods.” *Signal Processing* 167: 107299.

UNSD, UNdata. 2003. “United Nations Statistics Division.”

Ynnerman, A., J. Löwgren, and L. Tibell. 2018. “Exploranation: A New Science Communication Paradigm.” *IEEE Computer Graphics and Applications* 38 (3): 13–20.