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Illusrating Changes in Time-Series Data With Data(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

### 地域の可視化

## 文章生成

## インタラクション

# ユースケース

ユースケース

# 考察

考察

# まとめ

まとめ

An Example of the Manuscript for  
the Visualization Symposium Japan (Times New Roman 16pt)

―Subtitle (Times New Roman 14pt)―

Taro KASHIKA and Hanako JOHO

##### 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. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*《150語程度の英文ABSTRACT及び5語以下のKeywordsを記入してください．文字の大きさは10pt，字体はCentury，行間は１行分 (シングルスペース) です．》

Keywords: Visualization, Digital image processing, \*\*<5語以下>\*\*, \*\*\*\*\*, \*\*\*\*\*

# 序　論

本原稿は，可視化情報シンポジウムの原稿テンプレートです．原稿執筆にはこのテンプレートをご使用ください．原稿サイズはA4版とし，ページ数は任意ですがファイルサイズは5MB以下としてください．ページ余白は，上25mm，左20mm，右20mm，下20mmとし，ヘッダーおよびフッダーは入れず，表題は3行目から始めてください．

# 文字及び改行

## 文字の大きさと改行

使用言語は日本語または英語とします．和文原稿の場合，題目，著者名，所属を和文で書き，続けて題目，著者名の英文名を記入してください．英文原稿の場合，題目，著者名の英文名を記入してください．なお，和題・英題は講演申込時と同じであることが必要です (異なる場合，プログラムや講演論文集の目次には講演申込時の題目が使用されます)．

著者名は11pt，1行に36字以内で右寄せとし，これを越える場合は改行してください．本文は9ptの2段組みとし，片側25字×50行，段間は10mmを基準としてください．章間の行間は2行分 (ダブルスペース) とし，その中央に章名を10ptで記し，節間は行を空けないようにしてください．

## 字　体

題名，章・節名はゴシック体，本文は明朝体，英文題名，著者名はTimes New Roman, 本文中の英文はCenturyとします.参考文献1) は8ptで表記ください．

* *強調*
* **とても強調**

# 図　表

本文中の図表の表記には，下記Fig.1と Table 1のように，図表には通し番号をふり，図表番号はゴシック体，図表題はCenturyで書きます．キャプションは表の場合は上部に，図の場合は下部に記し，説明文は英文で表記してください．

Table 1: Sample of table

| Table | Table |
| --- | --- |
| 1 | 2 |

表 [Sample of table] を参照のこと。



Figure 1: PacificVis PC meeting

# 原稿提出

本テンプレートで作成した講演原稿 (PDF形式) と，調査票 (PDF形式，必ず押印したものをお願いします) をシンポジウムHPの案内に従って提出して下さい．

# 数式

二点を結ぶふたつの制御点列とについて。。。

macros.tex に定義したようなTeXマクロにも対応しています。

二点を結ぶふたつの制御点列とについて。。。

以上で定義した三つの指標がいずれも一定値以上出会った場合に限り，束化圧があるものと評価する．このことは閾値 () を用いて以下のように定式化される：

# その他

## 文献の引用

この例題では、library.bib 文献データベースとし、文献引用のスタイルにシカゴスタイルを指定して文献リストと文献の引用をしている。.cslファイルを別のものに指定すれば、他の引用形式にも対応できる。文献の引用方法は LaTeX の \cite {...} コマンドとは異なる。以下を参考にすること。

(Kashika 1997; Kashika and Hanko 1999) は英語の文献で(可視化 and 情報 1998)は日本語の文献である.

# 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.

Kashika, T. 1997. “Visualization Technique.” *VSJ Press*, 1111–25.

Kashika, T., and J. Hanko. 1999. “An Example of the Manuscript for VSJ Journal.” *VSJ Journal* 10 (5): 10–15.

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/>.

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.

可視化太郎, and 情報花子. 1998. “可視化情報学会における論文の書き方.” *可視化情報* 100 (1000): 11–12.