

# 地理的な関連性を持つ時系列データの探索的分析を支援する 可視化システムに関する研究

## Visualization System for Exploratory Analysis with Geographical Time-series Data

内藤 峻<sup>1\*</sup>      岩崎 有基<sup>2</sup>      松下 光範<sup>2</sup>  
Shun Naito<sup>1</sup>      Yuki Iwasaki<sup>2</sup>      Mitsunori Matsushita<sup>2</sup>

<sup>1</sup> 関西大学大学院 総合情報学研究科  
Graduate School of Informatics, Kansai University

<sup>2</sup> 関西大学 総合情報学部  
Faculty of Informatics, Kansai University

**Abstract:** The goal of our study is to support a user's analysis of time-series data in an exploratory manner. Such exploratory analysis requires repeated access to various types of information related to the user's interests such as texts and numerical data. To support such the user's analysis, we have proposed a system that visualizes temporal changes in time-series data and presents the causes of those changes with the data. In this paper, we improve the system by (1) strengthening the linkage between graphs and articles (2) adding graph interactions. In addition, we conduct user experiments and examine the point of arrival of the system and the direction of future expansion.

## 1 はじめに

時系列データとは、熊本地震の被害者数や気温など時間の経過に伴って変化するデータである。このような時系列データは意思決定や問題解決の場面で役立てられている。意思決定や問題解決の場面では、時系列データの値の変化やその変化の要因を分析することで、有益な情報や新たな知見を得ることが重要である。しかし、このような時系列データの分析は仮説の生成や検証を探索的に繰り返す負荷の高い作業であるため、ユーザがこのような探索行為を円滑に行うことが難しいという問題がある。そこで本研究では、ユーザの興味や関心に応じて様々なモダリティの情報へのアクセスを繰り返しつつ時系列データを分析するための支援システムの実現を目指している。その端緒として、著者はこれまでに、新聞記事と地図、統計データを対象に、ユーザが時系列データの経時的変化とその変化の要因の把握を可能にする可視化インタフェースを提案してきた [1]。先行インタフェースでは、(1) 容易に興味を持った時点の記事を参照することができない、(2) データの推移を把握しづらい、という問題があったため、本稿では、(1) グラフと記事との連携の強化、(2) グラ

フのインタラクションの追加、によりユーザの探索行為の円滑化を図る。加えて、ユーザ実験を行い、システムの到達点と今後の拡張の方向性について検討する。

## 2 関連研究

### 2.1 言語情報を用いた時系列データの探索的分析の支援

松下らは、動向情報テキストから統計量を抽出し、グラフとして可視化する手法を提案している [2]。動向情報テキストから単に統計量を抽出し、グラフの描画を行った場合、とびとびの値のみを抽出することしかできないため、ユーザの役に立つ可視化表現を得ることができない。この問題を解決するために、動向情報テキストの視覚的な情報要約を行うことを目的としている。提案手法では、比較表現と背景知識を用いて統計量を補填したり、テキスト中の定性表現を解釈し、グラフ上にその表現に当たる図形を描画したりしている。実験では、提案手法の補填によって多くの統計量を抽出できることと、提案手法によって生成された可視化表現がユーザの理解を高めることを明らかにしている。

山本らは、動向情報の変化とその変化の要因を視覚的に表示するシステムを研究している [3]。システムは、

\*連絡先： 関西大学大学院総合情報学研究科知識情報学専攻  
〒 569-1095 大阪府高槻市霊仙寺町 2-1-1  
E-mail: mat@res.kutc.kansai-u.ac.jp

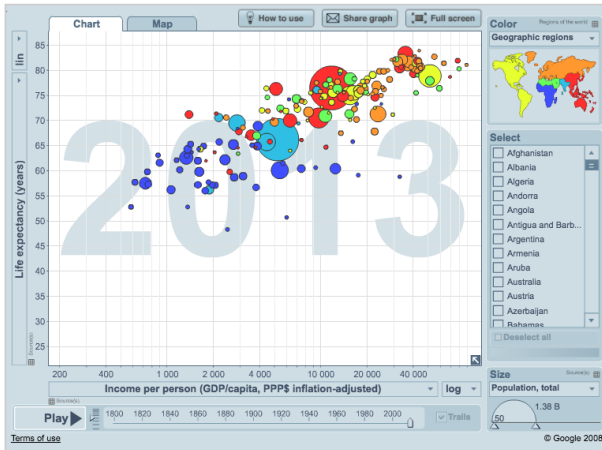


図 1: Gapminder (文献 [4] より図引用)

内閣支持率に関する新聞記事を入力することでユーザの興味を考慮し、変化の根拠となる要因をグラフ上に配置する。また要因の抽出手法では、動向情報が記載された新聞記事とコサイン距離が近い記事を要因としている。さらに、表示方法に関してユーザ実験を行い、ユーザが動向情報の要因を知りたい部分はグラフのどの部分なのか、要因をどのように表示すれば良いかを明らかにしている。

## 2.2 地理情報を用いた時系列データの探索的分析の支援

ハンス・ロスリングらは Gapminder を提案している [4]。Gapminder はユーザが瞬時の視覚分析に用いるデータを簡単にアクセスしやすくすることを手助けする。Gapminder は、国ごとの石油生産や国民一人あたりの収入、平均余命など様々な統計データを保持している。図 1 にシステムの外観を示す。ユーザはマッピングする統計データをグラフもしくは世界地図に切り替えることができる。グラフではユーザはこれらのデータの値の変化を 5 次元で一度に閲覧し、考察できる。5 次元は、x 軸、y 軸、時間、プロットサイズ、色で表現されている。ユーザはこのシステムを使うことで、国民一人あたりの収入と平均余命がどのような関係にあるのか調べることができる。

井上らは、感染症の調査データをよりわかりやすくするために WEB ベースのデータ可視化システムを提案している [5]。日本の感染症のサーベイランスレポートでは、全国の感染症の罹患率の傾向や周辺地域間の伝染を把握することが困難であることを解決するために、このシステムを開発している。システムは医療従事者が感染症の罹患率の傾向を素早く突き止め、一般住民に迫り来る病気の発生を警告することを手助けす

る。システムのデータベースは、2000 年から 2012 年までの 47 都道府県における 11 の小児科の感染症とインフルエンザの週報を含んでいる。それは約 300000 件に達する。

## 2.3 本研究の位置付け

時系列データの経時的変化と地理的な影響や規模の把握を容易にする可視化システムは提案されている [4] [5] が、時系列データの変化の理由や背景を知る用途には適していない。また、時系列データの経時的変化とその変化の理由や背景の把握を容易にする可視化システムは提案されている [3] が、地理的な影響や規模を知る用途には適していない。

本研究ではこれらのことを踏まえて、時系列データの経時的変化の把握を容易にするために折れ線グラフを、変化の理由や背景を知るために新聞記事を、地理的な関係性を把握するために地図を各々用いて、これらの異なるモダリティの情報にアクセスする方法を検討する。

## 3 システムの全体像とこれまでの取り組み

図 2 に本システムの目指す構成を示す。現状のシステムは新聞記事 DB、統計 DB、地図 DB を人手で作成している。しかし、人手での作成は、効率性や網羅性、リアルタイム性の点で問題がある。そのため、システムに用いるデータは WEB から抽出することを考えている。新聞記事 DB には、クローラを用いてあるトピックに関する記事を収集し、スクレイピング技術を用いて本文や見出しを抽出する。また、抽出された見出しや本文から自然言語処理技術を用いて日付や国名、出来事に関する文を抽出する。統計 DB や地図 DB は、オープンデータとの連携を検討している。

### 3.1 システムの全体像

図 3 に理想のシステムの全体像を示す。システムは、トピックを選択する選択ボックス (図 3-I)、地図を表示する地図ペイン (図 3-A) とグラフを表示するグラフペイン (図 3-B)、記事を表示する記事ペイン (図 3-C) で構成されている。選択ボックスには、「エボラ出血熱」や「台風」といったトピックがプルダウン形式で表示される。ユーザがトピックを選択すると、そのトピックに関連する統計量がグラフとしてグラフペインに描画され、同時に地図ペインには関連する地図を、記事ペインには関連する記事が表示される。また、地図とグ

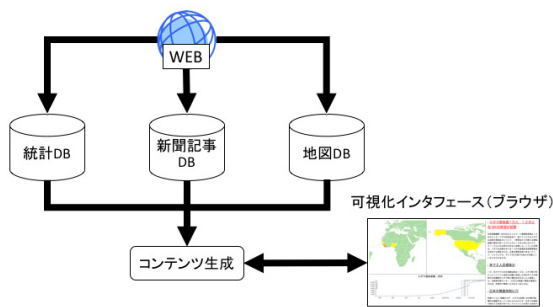


図 2: 実現を目指すシステムの構成 (文献 [6] より図引用)

ラフ上には、記事の有無を表すアノテーションとしてアイコンが付与されている。アイコンは文献 [7] を参考に、5 種類 (理由、背景、状況、出典、その他) を考えている。ユーザがアイコンをクリックすると、そのアイコンに対応する記事が表示される。さらに、グラフペインには、複数の統計グラフと凡例が提示されるようになっている。ユーザは凡例の左にあるチェックボックスに比較したい国を選択することでグラフペインに複数の統計グラフを描画することができる。これにより、ユーザは興味を持った国同士の統計量の変化を比較することで他国からの影響や規模を詳しく知ることができる。また、下のグラフを操作することで、上のグラフの表示する期間を変更したり、拡大や縮小を行うことができる。

現状のシステムには、地図上のアイコンや複数の統計グラフを描画する機能は実装されていないが、将来的にはこれらの機能を全て実装するつもりである。

### 3.2 これまでの取り組み

著者らは時系列データの経時的变化とその変化の要因の把握を可能にする可視化インタフェースを提案している [1]。グラフペインには、ユーザによって選択された日付を表示する青い線 (図 3-①) と記事の有無を表すアノテーションが表示されている。ユーザは青い線を左右にドラッグして日付を選択することができる。日付を選択すると、その時点の統計量が地図にマッピングされる。また、青い線をアノテーションに重ねると、その時点で起こった出来事について記述された記事がハイライトされる。さらに、地図にマッピングされた国をクリックすることができる。国をクリックすると、その国の統計量がグラフとしてグラフペインに描画される。同時に、その国名が含まれる記事がハイライトされる。加えて、記事をクリックすると、記事に含まれる日付がグラフ上に表示される。同時に、記事の本文に含まれている国がハイライトされる。

## 4 デザイン指針

### 4.1 対象とするデータの特徴

新聞記事は、ある時期における出来事やその出来事が起こった原因、場所、統計量、その統計量の具体的な値への言及や予測、記者の意見などが書かれている。そのため、出来事が起こった理由や背景を理解する上で有用である。しかし、新聞記事に書かれている統計量の値は近似値が用いられているため正確でなかったり、記者の観点で纏められていたりするため、客観性に欠ける。

統計データは、ある観測された場所において、ある時点の事象について測定された値である。例えば、人口統計や外国為替相場の推移などがこれに当たる。これらの統計データの値は、政府や国際連合の専門機関などが実施している厳密な環境において観測されていたり、センサを用いて取得される。そのため、これらのデータは正確である。また、折れ線グラフとして描画することでデータの変化や概要を容易に把握することができる。

地図は地球や地表、架空の世界の全部もしくは一部を平面上に縮尺表現したものである。例えば、地球全体もしくは大部分を表現している世界地図や統計データを地図上に表した統計地図などがこれに当たる。地図は地理的な位置関係や方向、距離、面積、形、高さを知る上で有用である。また、ある時点の統計量を地図にマッピングすることで、出来事の規模や地理的な広がりや把握できるといった特徴を持っている。

これらの情報はそれぞれ単体でも用いることができるが、ユーザの興味となる要素をトリガとしてインタラクティブに情報を提示することで、円滑な情報アクセスが可能なインタフェースを実現できると考えている。

### 4.2 データの相補的利用

図 4 に、本研究で対象とするデータの関係性を示す。地図はグラフの興味を持った時点の統計量をマッピングすることで、その事象の地理的な広がりや把握したり (図 4-①)、地理的な広がりの変化を見て興味を持った地域の統計グラフを参照することで、その地域の統計量の時間的な変化を把握したりできる (図 4-②)。また、そのような地理的な広がりの変化を見て興味を持った地域の新聞記事を参照することで、その変化の要因となる理由や背景を把握したり (図 4-③)、記事に言及されている統計量に対応するグラフを参照することで、その統計量の概況を把握したりできる (図 4-④)。さらに、記事で言及されている事象 (e.g., エボラ熱の患者数) に興味を持って、その事象に関する地理的影響や規模を把握したり (図 4-⑤)、グラフの特徴的な箇所の新

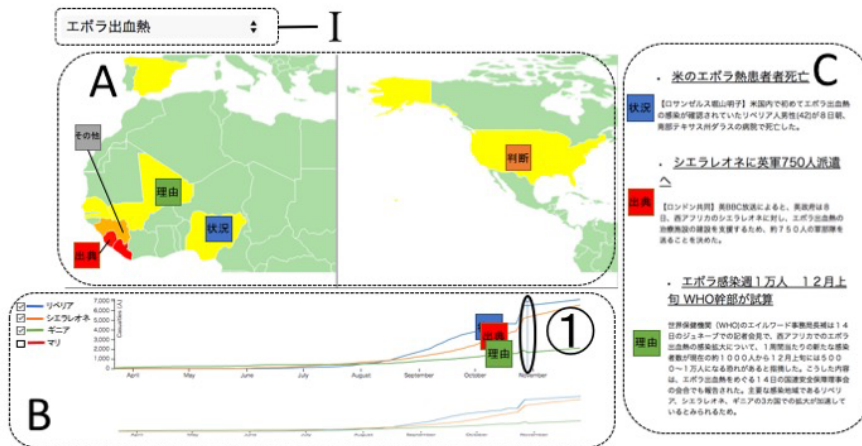


図 3: 理想のシステムの全体像

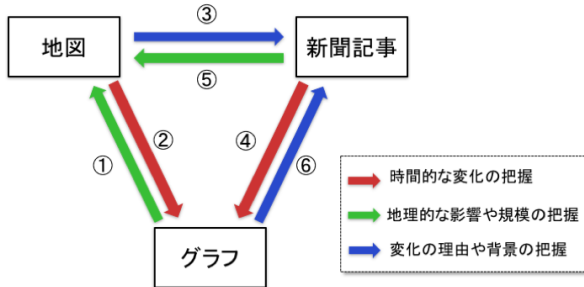


図 4: 対象とするデータの関係性 (文献 [6] より図引用)

聞記事を見ることでその変化の理由や背景を知ることができる (図 4-⑥)。本研究ではこのような情報アクセスを行えるようにシステムの機能をデザインした。これによって、ユーザの探索行為を円滑にする。

## 5 時系列データの探索的分析を支援する可視化システム

### 5.1 システムの構成

提案システムの構成を図 5 に示す。システムは、新聞記事と統計データ、地図の 3 つのデータベースとモジュール、入力判断部、コンテンツ生成部、画面生成部から構成されている。

システムはユーザの入力に応じて、これらのデータベースに格納されているデータを可視化表現 (e.g., グラフ, 地図, 記事) として生成する。ユーザの入力は、(1) 記事をクリックする, (2) グラフの日付を選択する, (3) 地図上の都道府県をクリックする, といった

直接操作を対象とする。次に、システムが実行する処理の流れを説明する。

まず、システムはユーザからの入力を入力判断部で判断する。入力判断部は、ユーザがどのような操作を行ったのかを解釈し、その結果をコンテンツ生成部へ伝える。コンテンツ生成部は、入力判断部から送られる結果をもとに、各モジュールに処理を伝達する。伝達を受けたモジュールは、データベースから必要なデータを受け取り、可視化表現を生成する。

地図モジュールでは、(1) 地理情報をもとに地図データベースから位置情報を取得し、地図を生成する、(2) 地図上の都道府県をハイライトする、という 2 つの処理を行う。都道府県のハイライトでは、あらかじめ統計モジュールから引き渡された統計量の最小値と最大値に基づいて、統計量を 0 から 255 に変換し、それを色の RGB 値に用いる。地図データベースは、地理情報とそれに関する位置情報を格納している。

記事モジュールでは、(1) トピックをもとに記事データベースから本文とその見出しを取り出し、記事のスニペットを生成する、(2) 記事のスニペットをハイライトする、という 2 つの処理を行う。記事データベースは、トピックごとに分類された新聞記事データを格納している。新聞記事データは本文とその見出し、記事の発行日を抽出したデータである。

統計モジュールでは、(1) 統計情報をもとに、統計データベースから統計量を取り出し、グラフとして描画する、(2) グラフ上の日付をハイライトする、という 2 つの処理を行う。また、グラフを描画する際に、記事へのアクセスが可能なトリガとしてアノテーションを付与する。統計データベースは、地域ごとの統計量を時系列順に格納している。

このように、各モジュールでは、(1) 可視化表現を



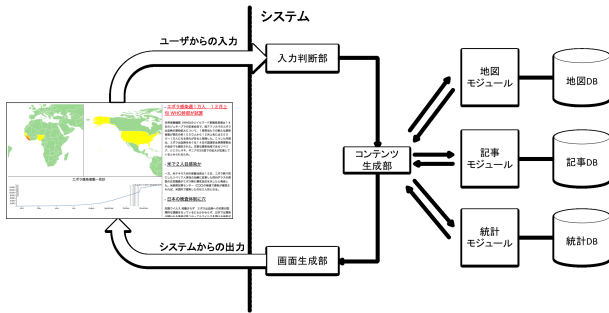


図 5: 提案システムの構成図 (文献 [1] より図引用)

生成する, (2) その可視化表現の一部をハイライトする, という 2 つの処理を行っている。

また, 各モジュールはコンテンツ生成部を介して, 他のモジュールと連携している。

記事モジュールは, データがタグ付けられた記事のスニペットから, コンテンツ生成部を介して, 日付を統計モジュールへ引き渡す。これは, グラフの描画やアノテーションの付与, 日付のハイライト, 都道府県のハイライトに用いられる。

統計モジュールでは, グラフ上のアノテーションの付与を行うために, 統計データの値の補間を行っている。本システムでは, 記事の有無を示すアノテーションとして円を折れ線グラフ上に配置している。この円を折れ線グラフ上に配置するためには, 記事が持つ日付に当たる統計データの値が必要になる。しかし, 統計データは, 1 週間毎に観測されたデータやところどころ抜けているデータがほとんどである。そのため, これらのデータの値を線形補間することで, 折れ線グラフ上に円を配置した。

統計モジュールは, コンテンツ生成部を介して, 統計量とその最大値と最小値, その地域名を地図モジュールへ引き渡す。これは, 地図のマッピングに用いられる。また, 日付情報をコンテンツ生成部を介して記事モジュールへ引き渡す。これは, 記事のハイライトに用いられる。

地図モジュールは, コンテンツ生成部を介して, 地理情報を記事モジュールや統計モジュールに引き渡す。これは, 記事のハイライトやグラフの描画に用いられる。

コンテンツ生成部は, 記事モジュールから引き渡された統計量とそれに関する日付を統計モジュールへ引き渡すといった処理を行う。このように, コンテンツ生成部と各モジュールがデータの受け渡しを繰り返すことにより, 可視化表現を生成する。そして, 生成された可視化表現は, コンテンツ生成部によって集められ, 画面生成部へ処理を移行する。画面生成部では, コンテンツ生成部で生成された可視化表現をユーザに提示する。

表 1: 統計データの例

| 日付        | 合計    | 岩手県  | 宮城県  | 福島県  | その他 |
|-----------|-------|------|------|------|-----|
| 2011/3/18 | 6495  | 3370 | 1926 | 1139 | 60  |
| 2011/3/22 | 9415  | 3370 | 4846 | 1139 | 60  |
| 2011/3/31 | 11310 | 3370 | 6741 | 1139 | 60  |
| ...       | ...   | ...  | ...  | ...  | ... |
| 2011/7/29 | 15645 | 4615 | 9364 | 1600 | 66  |
| 2011/3/18 | 15735 | 4649 | 9417 | 1603 | 66  |

## 5.2 対象データ

本システムでは, 統計データとして農林水産省の果樹生産出荷統計のみかんの出荷量<sup>1</sup>と土地価格<sup>2</sup>, 東日本大震災における死者の推移に関するデータ<sup>3</sup>を用いた。また, 統計データを都道府県別かつ時系列順に整理し, csv 形式にした。

統計データの例を表 1 に示す。東日本大震災の統計データには, 東日本大震災における死者・行方不明者の推移の wikipedia から死者の推移の統計データを人手で抽出したものを用いた。土地価格には, 土地価格が分かる土地代データという Web サイトから各都道府県の公示地価, 基準地価の総平均値を用いた。みかんの出荷量には, 農林水産省に掲載されている平成 15 年から 27 年までの果樹生産出荷統計の都道府県別の結果樹面積・10a 当たり収量・収穫量・出荷量のみかんの csv データの出荷量を用いた。

新聞記事データは, 野菜・果物と地価, 東日本大震災のニュース記事である。地価と東日本大震災のニュース記事には, Yahoo!ニュースの地価<sup>4</sup>と東日本大震災に関連するアーカイブ一覧<sup>5</sup>を用いた。アーカイブ一覧から各記事の見出しと本文, 記事の発行日を抽出し, csv 形式で保存するプログラムを作成した。記事の発行日は, 「年/月/日」のとなるようにフォーマットを統一した。

野菜・果物のニュース記事には, 農業協同組合新聞の野菜・果物のニュース<sup>6</sup>を用いた。ニュース一覧から 2013 年 1 月 7 日から 2016 年 12 月 1 日までの記事を前述したプログラムを用いて抽出した。

## 5.3 提案システムの概要

提案システムを図 6 に示す。選択ボックス (図 6-I) には, システムが提示しているトピック名が表示されており, ユーザは選択ボックスを操作することで分析

<sup>1</sup><http://www.maff.go.jp/j/tokei/kouhyou/sakumotu/>

<sup>2</sup><http://www.tochidai.info/>

<sup>3</sup><https://ja.wikipedia.org/wiki/東日本大震災における死者・行方不明者の推移>

<sup>4</sup>[http://news.yahoo.co.jp/list/?t=2011sanrikuoki\\_eq](http://news.yahoo.co.jp/list/?t=2011sanrikuoki_eq)

<sup>5</sup>[http://news.yahoo.co.jp/list?t=price\\_of\\_land](http://news.yahoo.co.jp/list?t=price_of_land)

<sup>6</sup><http://www.jacom.or.jp/yasai/news/>

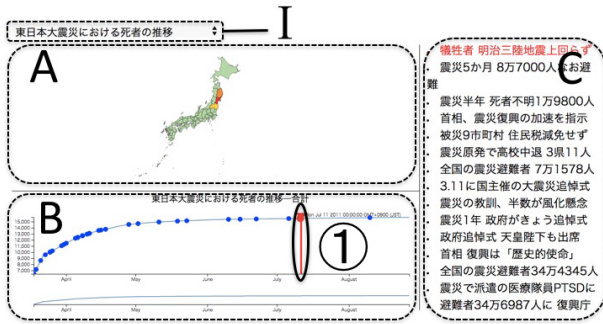


図 6: 提案システム

したいトピックを選択することができる。グラフペイン (図 6-B) には、2つのグラフが表示されている。上部のグラフ上には、赤色のスライダーと青色の円 (図 6-①) が表示されている。スライダーはマウスによる直接操作が可能であり、左右にドラッグすることで日付を選択し、それに併せて地図上にマッピングする統計データを切り替えることができる。青色の円は、その時点で記事があることを示している。スライダーがこの円に重なるとその時点のニュース記事がスクロールされ、スクロールされた記事と重なった円が赤くハイライトされる。また、下部の小さいグラフを操作することで上部のグラフの表示する期間を変更したり、拡大・縮小を行うことができる。記事ペイン (図 6-C) に表示されている記事はクリックすることが可能であり、クリックすると、その記事の本文を表示することができる。同時に、選択された記事に該当するグラフ上の円が赤くハイライトされる。

## 5.4 実装概要

5.2節で述べたプログラムの実装には、Python の version 3.5.2 と Python のライブラリとして BeautifulSoup の version 4.5.1 を用いた。5.3節で述べたシステムは、ウェブアプリケーションになっており、作成には HTML と CSS, JavaScript を用いた。また、JavaScript のライブラリとして、D3.js<sup>7</sup> の version 3.4.13 と TopoJSON の version 1.6.18 を用いた。

## 6 実験

実験では、提案システムの問題点と改善点を明らかにするために、先行システムと提案システムとのユーザビリティの比較を行った。実験参加者には、シナリオの課題を達成するようにシステムを操作してもらい、アンケートに回答してもらった。これらの手順を1つ

表 2: 実験参加者と課題の対応表

| 実験参加者 | 先行システム   | 提案システム   |
|-------|----------|----------|
| A     | 東日本大震災   | 土地価格     |
| B     | みかんの出荷統計 | 東日本大震災   |
| C     | 土地価格     | みかんの出荷統計 |
| D     | 東日本大震災   | 土地価格     |
| E     | みかんの出荷統計 | 東日本大震災   |
| F     | 土地価格     | みかんの出荷統計 |
| G     | 土地価格     | 東日本大震災   |
| H     | 土地価格     | 東日本大震災   |
| I     | みかんの出荷統計 | 土地価格     |
| J     | みかんの出荷統計 | 土地価格     |
| K     | 東日本大震災   | みかんの出荷統計 |
| L     | 東日本大震災   | みかんの出荷統計 |

のシステムで2回行った。実験参加者には、提案システムで2回、先行システムで2回の計4回アンケートに回答してもらった。シナリオ1はグラフから記事へアクセスする機能を、シナリオ2は記事からグラフへアクセスする機能を用いることが想定されている。アンケートの項目は、(1)シナリオの目的を達成できたか、(2)機能は十分だったか、(3)使う時に迷ったことなど不満な点はありませんでしたか、(4)こうした方が良いというような意見があれば教えてください、の4つである。項目1と2には、5段階のリッカート尺度を用いた。

実験参加者は筆者らが所属している研究室の大学生12人 (男性10人, 女性2人) である。実験のシステムに用いたトピックは、「東日本大震災」と「土地価格」、「みかんの出荷統計」の3種類である。実験では、実験参加者が2つのシステムにおいて必ず異なるトピックを用いるようにした。また、2つのシステムの順序効果を統制した。実験参加者と実験に用いたシステムのトピックとの対応表を表2に示す。

アンケート結果のシナリオ1の目的の達成度を表3に、シナリオ2の目的の達成度を表4に示す。シナリオ1におけるシステムの機能の満足度を表5に、シナリオ2におけるシステムの機能の満足度を表6に示す。各システムのシナリオの目的の達成度と機能の満足度の平均スコアから、シナリオ1の目的の達成度を除いて提案システムのスコアが先行システムのスコアを上回っていることがわかる。シナリオ1において先行システムのスコアが提案システムのスコアを上回った要因としては、提案システムと比べて記事が見づらいこととグラフと記事との対応付けがわかりづらかったことが挙げられる。記事が見づらいことについては、先

<sup>7</sup><https://d3js.org/>

表 3: シナリオ 1 の目的の達成度

| 実験参加者 | 先行システム | 提案システム |
|-------|--------|--------|
| A     | 4      | 5      |
| B     | 2      | 4      |
| C     | 4      | 2      |
| D     | 4      | 4      |
| E     | 4      | 4      |
| F     | 5      | 4      |
| G     | 5      | 3      |
| H     | 4      | 2      |
| I     | 4      | 4      |
| J     | 4      | 5      |
| K     | 4      | 4      |
| L     | 4      | 4      |
| 平均    | 4.0    | 3.8    |

表 4: シナリオ 2 の目的の達成度

| 実験参加者 | 先行システム | 提案システム |
|-------|--------|--------|
| A     | 5      | 3      |
| B     | 2      | 5      |
| C     | 4      | 5      |
| D     | 2      | 2      |
| E     | 1      | 4      |
| F     | 4      | 5      |
| G     | 5      | 4      |
| H     | 3      | 3      |
| I     | 4      | 2      |
| J     | 5      | 5      |
| K     | 4      | 5      |
| L     | 5      | 5      |
| 平均    | 3.7    | 4.0    |

行システムでは記事と記事の間隔が空いていたものの、提案システムではその間隔が空いていなかったため、記事が見つらかったと 1 人の実験参加者が述べていた。グラフと記事との対応付けのわかりづらさについては、1 人の実験参加者がグラフから記事にアクセスするために円をクリックしたり、記事の発行日を見るために、グラフの表示期間外の記事をクリックしたり、していた。

## 7 議論

グラフから記事へアクセスする機能では、7 人の実験参加者が記事を見ようとしてグラフ上の円をクリックしていた。このことから、操作がわかりづらいことが明らかになった。この問題は、青い丸をクリックするとその記事にアクセスできる機能を設けることで解決できると考えている。

記事からグラフへアクセスする機能について、5 人の実験参加者から本機能を用いて記事がグラフのどの部分にあたるのか把握しようとする発話が確認された。

日本地図について、4 人の実験参加者から地図が小さいというコメントがあった。これを解決するためには、地図を拡大したり、縮小したりする機能が望ましい。

記事の選定について、提案システムでは、スクレイピングした記事をすべて用いている。そのため、提示している時系列データに関係がない記事や提示している時系列データの期間とは異なる記事も含まれている。また、時系列データの分析においてユーザは特徴的な変化に興味を持ち、その変化の原因を調べるために変化した付近の記事を参照する。このようなユーザにとって、変化した付近以外の記事や原因に関係がない記事

を参照することは時間や手間を要する。そのため、このような時系列データを分析するユーザにとってはグラフの特徴的な変化している時点付近に、グラフの変化について述べられている記事が対応付けられていることが望ましい。提案システムでは、このような対応付けを実現する方式について検討する必要がある。

新聞記事と時系列データのアラインメントについて、提案システムにおける記事と時系列データの対応付けは、記事の発行日を用いて行っている。しかし、記事には過去のある時点のデータの変化や過去の出来事について言及されているものが多くある。そのため、このような記事では記事の発行日を用いて対応付けるのではなく、記事の文章を読み取り、適切な箇所に対応付けることが望ましい。このような取り組みとして、著者らは記事と時系列データをアラインメントする手法について検討している [6]。この手法では、記事中の日付と値の変動に関する曖昧表現の 2 つに着目し、時系列データの該当箇所を特定している。この手法を提案システムに組み込むことで、より適切な記事と時系列データのアラインメントが可能となる。

## 8 おわりに

本研究の目的は、ユーザの興味や関心に応じて様々なモダリティの情報へのアクセスを繰り返しつつ時系列データを分析するための支援システムの実現である。著者らはこれまでに、ユーザが時系列データの経時的変化とその変化の要因の把握を可能にする可視化インタフェースを提案してきた [1]。先行インタフェースでは、(1) 容易に興味を持った時点の記事を参照するこ

表 5: シナリオ 1 におけるシステムの機能の満足度

| 実験参加者 | 先行システム | 提案システム |
|-------|--------|--------|
| A     | 4      | 5      |
| B     | 2      | 4      |
| C     | 4      | 3      |
| D     | 3      | 4      |
| E     | 3      | 4      |
| F     | 3      | 5      |
| G     | 4      | 4      |
| H     | 4      | 2      |
| I     | 4      | 4      |
| J     | 2      | 5      |
| K     | 4      | 4      |
| L     | 4      | 4      |
| 平均    | 3.4    | 4.0    |

表 6: シナリオ 2 におけるシステムの機能の満足度

| 実験参加者 | 先行システム | 提案システム |
|-------|--------|--------|
| A     | 4      | 3      |
| B     | 3      | 4      |
| C     | 2      | 5      |
| D     | 3      | 2      |
| E     | 3      | 5      |
| F     | 3      | 4      |
| G     | 4      | 3      |
| H     | 3      | 3      |
| I     | 3      | 4      |
| J     | 4      | 5      |
| K     | 4      | 4      |
| L     | 5      | 4      |
| 平均    | 3.4    | 3.8    |

とができない, (2) データの推移を把握しづらい, という問題があったため, 本稿では, (1) グラフと記事との連携の強化, (2) グラフのインタラクションの追加, によりユーザの探索行為の円滑化を図った. 加えて, ユーザ実験を行い, システムの到達点と今後の拡張の方向性について検討した. 今後は, (1) 記事と地図との連携の強化, (2) 分析に必要なデータの選定, を行うことでより効率的に時系列データを分析できるようにする.

## 謝辞

本研究の遂行にあたり, 文部科学省科学研究費(課題番号:15H02780)の助成を受けた, 記して謝意を表す.

## 参考文献

- [1] Naito, S., Matsushita, M.: Supporting Consecutive Data Exploration by Visualizing Spatiotemporal Trend Information, in *Proceedings of the 2015 Conference on Technologies and Applications of Artificial Intelligence*, pp. 227-231 (2015)
- [2] 松下光範, 加藤恒昭: 数値情報の補填とグラフ概形の示唆による複数文書からの統計グラフ生成, 知能と情報, Vol. 18, No. 5, pp. 721-734 (2006)
- [3] 山本健一, 殿井加代子, 谷岡広樹: タグ付きコーパスを用いた動向情報とその要因の可視化, 言語処

理学会第 12 回年次大会ワークショップ「言語処理と情報可視化の接点」(2006)

- [4] Rosling, H.: Gapminder, (2011) <http://www.gapminder.org> (2017/2/19 確認)
- [5] Inoue, M., Hasegawa, S., Suyama, A., Kakehashi.: Development of a Web-based Data Visualization System for Comprehensible Ascertainment of the Spatiotemporal Extent of Infectious Diseases, 医療情報学, Vol. 33, No. 1, pp. 27-32 (2013)
- [6] 内藤峻, 古田遼樹, 松下光範: 時系列データの探索的分析を支援する可視化システム:記事と時系列データのアラインメント方式の提案, 人工知能学会第 13 回インタラクティブ情報アクセスと可視化マイニング研究会, pp. 1-6 (2016)
- [7] 松下光範, 加藤恒昭: 言語情報と数値情報の相補的利用を目指した可視化手法, 第 21 回人工知能学会全国大会, 3H8-3 (2007)