

平成 30 年 6 月 21 日現在

機関番号：32644

研究種目：挑戦的萌芽研究

研究期間：2015～2017

課題番号：15K11997

研究課題名(和文) オープンデータ活用のためのデータ統合・可視化サイトに関する研究

研究課題名(英文) Research on data integration and visualization site for utilizing Open data

研究代表者

山本 義郎 (Yamamoto, Yoshiro)

東海大学・理学部・教授

研究者番号：80301943

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：オープンデータやSNS情報などインターネットを利用することにより入手できる情報について整理し、それらを有効活用する方法について調査した。当初は、その種のデータを統合するサイト構築を想定していたが、それらを有効に活用したシステム構築例を多く示す方が有益考えた。第一の対象領域は減災であり、Twitter, GoogleAPI, GoogleMapなどを利用しGIS可視化システムを構築した。その他、Rで簡単にWebアプリケーションが構築できるShinyを利用し、心肺停止患者の救急搬送に関するウツタインデータの可視化システムを構築した。更に、よりインタラクティブにユーザの利便性を上げる機能も実装した。

研究成果の概要(英文)：We organized the information available by using the Internet such as open data and SNS information and investigated ways to make effective use of them. Originally, we assumed the site construction to integrate such data, but it was beneficial to show many examples of system construction that effectively utilized them. The first target area was disaster reduction, and we construct a GIS visualization system using Twitter, Google API, Google Map, etc. In addition, Shiny which can easily construct Web application with R is used to construct a visualization system of Utsutinta data on emergency transportation of cardiopulmonary arrest patients. In addition, we also implemented a function that makes user's convenience more interactive.

研究分野：計算機統計学

キーワード：オープンデータ SNSデータ 可視化 データ分析 インタラクティブ Webアプリケーション

1. 研究開始当初の背景

インターネットの普及と公的情報の公開からの、公共データのビジネス活用への期待や様々な研究分野での外部データとの連携による研究の推進への期待などがあり、オープンデータへの取り組みが進んでいる。しかしながら、PDF ファイルに記録されたデータの公開など、実際にデータ解析に活用しようとする問題のあるオープンデータも多く存在する。実際にオープンデータを有効活用するためには、まだまだ様残な障壁があるのも事実である。

研究代表者はこれまで、教育面でオープンデータを活用する他、地理情報などのオープンデータを活用し、他の情報と結合し、統計解析やデータの可視化に取り組んでいる。特に Web 上のデータの可視化については科学研究費を獲得し新たな方法も提案している。さらに近年では、Twitter に代表される SNS (ソーシャル・ネットワーク・サービス) から様々な有益な情報を得ることができている事例も報告されるようになった。本研究計画では、インターネット上からオープンデータや SNS データなどを入手し、統計解析システムを利用して分析・可視化するための環境を検討するとともに、有益な可視化手法が提案できるのではないかと考えた。

2. 研究の目的

本研究室では、インターネット上に公開されている統計情報、地理情報などのオープンデータを活用し、他の情報と結合し、統計解析やデータの可視化に取り組んでいる。特に Web 上のデータの可視化については科学研究費を獲得し新たな方法も提案している。近年では、Twitter やブログなどインターネット上に拡散されている情報を要約し、Web 上の地図と連携することで、地域に関連する有用な情報の可視化も可能となっている。

本研究課題では、インターネットと統計解析システムを利用して、オープンデータや地理情報を有効に活用した、新たな統計データの可視化の実現と、そのような可視化に必要なデータを統計解析可能な形に統合し、解析や可視化を実現するシステムを構築すること、また新たな可視化手法を提案することが目的である。

3. 研究の方法

今年度の計画はとしては、(1)「インターネット上のオープンデータを WebAPI 等で自動収集する方法の調査・開発」および、(2) SNS からの情報収集方法とオープンデータと統合する方法の開発であり、それぞれについて進捗を以下に示す。

(1)「インターネット上のオープンデータを WebAPI 等で自動収集する方法の調査・開発」として、国内のオープンデータについて調査を行うとともに Excel、R、PHP などでデータを取得する方法について整理するために Web

からの情報収集にとどまらず、国内外の学会、研究会においても情報収集を行った。データ解析コンペ、スポーツデータ解析コンペで利用可能なオープンデータの取得方法について優先して調査し、コンペ参加者への情報提供を行った。

(2) SNS からの情報収集方法とオープンデータと統合する方法の開発に関しては、Twitter からの情報収集および興味対象のデータについてデータの蓄積に関して調査および研究をすすめた。データベースに保有する方法と、R などでリアルタイムに分析する方法、Shiny などでサーバアプリケーションとしてユーザがインタラクティブにデータ取得、解析できる方法についても研究し、単純なシステムの構築を開始した。

平成 28 年度までに、様々なオープンデータについて API などを利用して、データの取得と可視化する方法について検討してきたが、現状としては API は仕様変更が多く、データの提供方法が変更されることが多々あり、情報をまとめて公開しても情報が古くなるケールが多いことが確認された。そのため、API などの仕様の特化せず、どの種の情報をオープンデータとして入手でき、それらの組み合わせや、独自のデータに追加して用いることで得られる知見についてまとめる方向とした。

平成 29 年は、(1)「インターネット上のオープンデータを WebAPI 等で自動収集する方法の調査・開発」として平成 28 年度までの研究成果をまとめ、解析および可視化が実施可能なシステムのプロトタイプを構築した。(2)「SNS からの情報収集方法とオープンデータと統合する方法の開発」については、平成 28 年度までの研究成果をまとめ、対象を Twitter からの災害情報の抽出に絞り、オンラインでの情報抽出とその可視化についての研究を行った。また、より即時的な分析のために、情報収集結果について Shiny によりインタラクティブな処理が可能なサーバアプリケーションとして実装について開発を行った。更に(3)スポーツの分野についてオープンデータと新規取得データを用いたデータ解析について検討し、動画からのデータ抽出やインターネットサイトからのスクレイピングなどを利用したデータの取得と取得された情報をサーバ上で即時的に解析する方法について検討した。(4)救命救急処置に関するウツタインデータのように完全にはオープンではないものの、研究目的で入手でき、その研究成果を発表できるタイプのデータについても分析と有効な可視化について検討し国内学会においていくつかの報告を行った。

4. 研究成果

(1)「インターネット上のオープンデータを WebAPI 等で自動収集する方法の調査・開発」については、政府統計の総合窓口

(e-Stat) が API でデータ取得可能となったことで、当初目的としていたものについては、独自に開発する必要がなくなり、R では estatapi パッケージにより簡単にデータを取得することができるようになった。参考 <https://notchained.hatenablog.com/entry/2015/11/01/084909>

(2)「SNS からの情報収集方法とオープンデータと統合する方法の開発」については、対象を Twitter からの災害情報の抽出に絞り、オンラインでの情報抽出とその可視化についての研究を行った。また、より即時的な分析のために、情報収集結果について Shiny によりインタラクティブな処理が可能なサーバアプリケーションとして実装について開発を行った。オープンデータとしては、気象情報や災害情報などを API で取得し、Googlemap 上に GUI 可視化するプロトタイプを構築した。(発表論文、学会発表など)

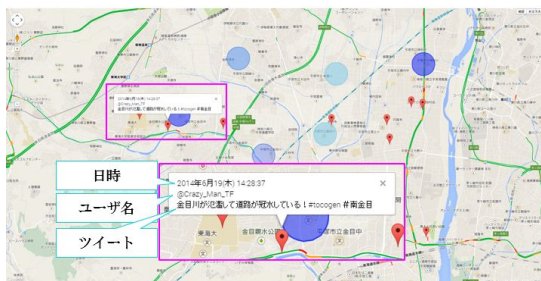


図 1：ツイート、気象情報、人口の可視化

また、減災のための情報活用として、平塚市の避難所について、人口を考慮した場合の避難所の不足についての検討や、熊本地震における土砂崩れデータを活用した、土砂崩れを起こしたと思われる場所の推定と可視化などについても学会・シンポジウムで分析結果を公表している。(学会発表 など)

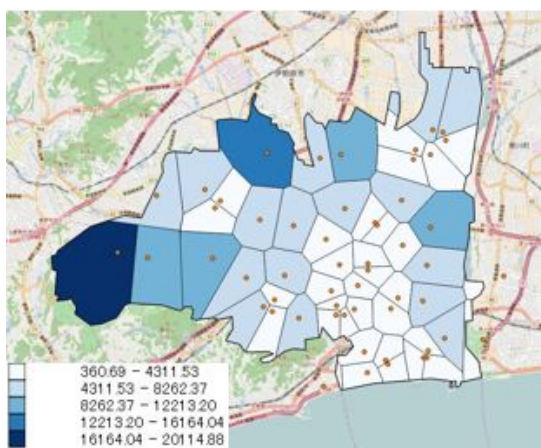


図 2 . 平塚市の避難所のボロノイ図

(3)スポーツの分野についてオープンデータと新規取得データを用いたデータ解析について検討し、動画からのデータ抽出やインターネットサイトからのスクレイピングなどを利用したデータの取得と取得された情

報をサーバ上で即時的に解析する方法について検討した。具体的な方法としてはクラウドの利用が適切と考えられたため、Microsoft Azure を利用して、仮想環境の Linux 上にデータを取得するシステムを構築し、大量の映像データのデータの処理を行える環境を作ることができた(学会発表)



図 3 . ビデオ映像からのボールの推定

(4)救命救急処置に関するウツタインデータのように完全にはオープンではないものの、研究目的で入手でき、その研究成果を発表できるタイプのデータについても分析と有効な可視化について検討した。ウツタインデータはオープンデータではないため、クラウド環境で利用者を限定して共有し分析を行った。

この分析においては、データの取得方法として通常の解析方法と変わらないが、これまでの本研究で取り組んだ、Web 上での解析結果の表示についてのノウハウを活かすことができ、解析は本研究室で行うが、Shiny により医学部においても結果の可視化と、独自視点の更なる考察ができるシステムを提供することができ、現在もより深い研究が進められている(学会発表 など)。

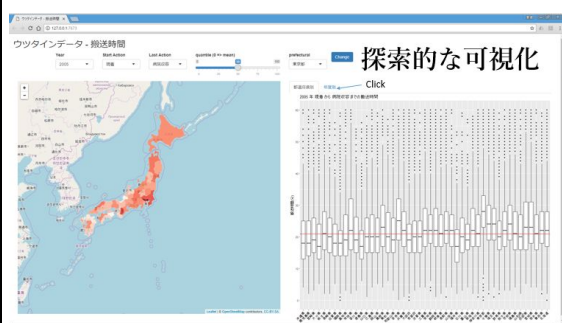


図 4 . ウツタインデータの探索的可視化

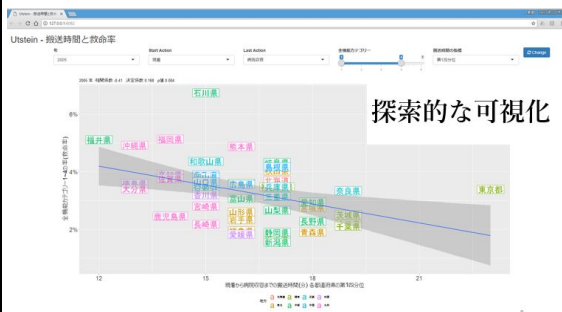


図 5 . 図 4 で設定を変えた可視化

(5) ウツインデータの解析と可視化
 ウツインデータの分析はこれまでも行われているが、その分析結果には、状況を考慮せずに全体のデータをまとめて分析したため、特定器具を使わない方が救命率が高いので、救急搬送時には何もしない方がよい、と結論付けた論文が受理された。しかしながら、心肺停止の要因ごとに分けて考えることにより、適切な状況において、特定器具を用いることが救命率が高くなることを確認できた。このように、適切な可視化によりデータから正確に情報を取得することができること、そのために適した可視化の方法があることが示された(学会発表 など)。

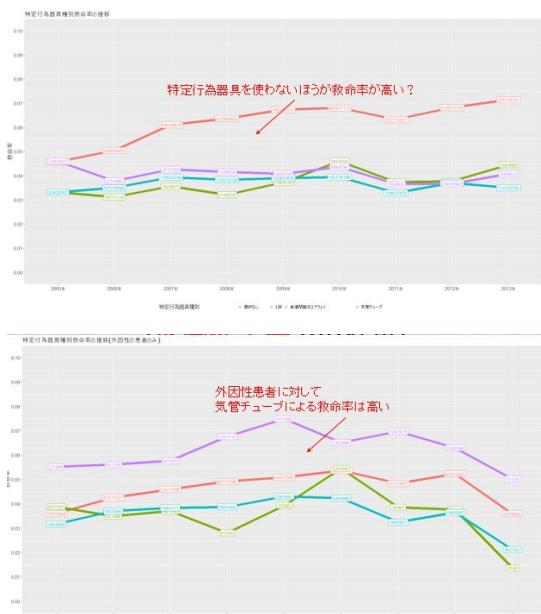


図6．年度ごとの救命率(下)外因性のみ

(6) 新たな可視化手法の提案

本研究を進める過程で、様々な統計解析手法を組み合わせ、その可視化により、さらにデータ解析結果の解釈を助ける可視化方法について研究してきた。

その1つとして、データマイニングの代表的な方法である、アソシエーションルール分析と多変量データ解析手法として古典的な対応分析を組み合わせた分析・可視化手法を開発した(雑誌論文)。このアプリケーションでは、パラメータ調整の必要があるため、Web アプリケーションとしての利点がある。

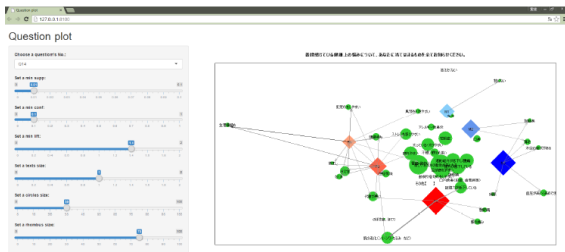


図7：属性特化型特徴抽出アソシエーションプロットとその Shiny アプリケーション

また、スポーツデータの可視化として、サッカーのボール支配率と利用しているフィールドエリアの可視化を、Shiny アプリケーションとして実装することができ、この種の Web アプリケーションの開発、公開が簡単にできるようになったことについて論文として公表することができた(雑誌論文 など)。



図8．サッカーデータの可視化

(7) インタラクティブな可視化

Shiny アプリケーションとして、リストボックスやスライダーで変更をおこなうことで、利用者の見たい状況を確認できることとなったが、さらに、visNetwork などグラフを線や点をマウス操作(スマホではタップ)で移動することができなど、より利用者のニーズに沿ったインタラクティブな可視化が提供できることについてもある程度まとめることができた(学会発表 など)。この成果について、平成 30 年以降継続して研究を行っていく。

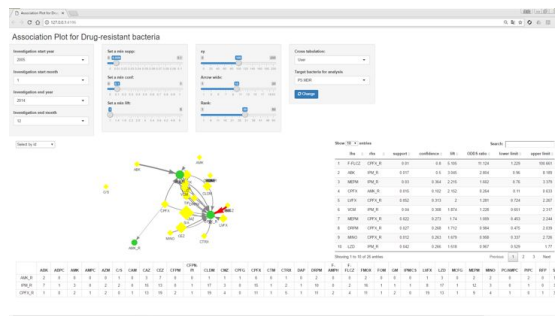


図9．アソシエーションプロットがマウスで移動可能、表のドラックでハイライト可能

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 25 件)

M. Kosugi, K. Utsu, M. Tomita, Y. Kajita, Y. Yamamoto, O. Uchida, Improvement of Twitter-based Disaster-related Information Sharing System, Proc. 4th International Conference on Information and Communication Technologies for Disaster Management, 2018/2,

<https://doi.org/10.1109/ICT-DM.2017.8275693>

T. Funayama, Y. Yamamoto and O. Uchida, Development of Visualization Application of Tweet Data for Extracting Information in case of Disaster, 2017 15th International Conference on ICT and Knowledge Engineering, 45-49, 2018, <https://doi.org/10.1109/ICTKE.2017.8259620>

山田実俊・山本義郎, RStudio を利用したサッカー試合のゾーン使用率・ゾーン支配率の可視化, 計算機統計学, 30(1), pp.59-71, 2017, https://doi.org/10.20551/jscswabun.30.1_59

S. Yamada, K. Umezawa, Y. Yamamoto, T. Ozeki, Visualization System for Analyzing Regional Characteristics of Ustein Data, 2017 Fifteenth International Conference on ICT and Knowledge Engineering, 40-44, 2018, <https://doi.org/10.1109/ICTKE.2017.8259619>

山田実俊・山本義郎, 対応分析を用いたアソシエーションルールによるアンケート結果の可視化, データ分析の理論と応用, 5(1), 3-15, 2016.

T. Funayama, Y. Yamamoto, M. Tomita, Y. Kajita, S. Tajima, K. Utsu, O. Uchida, Disaster Mitigation Support System using Web Services and SNS Information, 2015 13th International Conference on ICT and Knowledge Engineering, 42-48, 2015.

<https://doi.org/10.1109/ICTKE.2015.7368469>

〔学会発表〕(計 8 2 件)

船山・山本・内田, 災害時の Tweet 情報の変化に関する分析, 第 17 回西東京統計研究会, 2018 年 3 月

S. Tajima, O. Uchida, K. Utsu, M. Tomita, Y. Kajita, Y. Yamamoto, Information Behavior of University Students Living Outside the Disaster Area After the Kumamoto Earthquake, The 9th International Conference on Society and Information Technologies, 2018 年 03 月

今田・山本, 野球投球ビデオ動画からの投球の動体検出のための画像の分類について, 日本分類学会第 36 大会, 2017 年 12 月

山田・梅澤・尾関・猪口・山本, ウツタインデータから見える地域差 1 搬送時間 第 12 回日本病院前救急診療医学会, 2017 年 12 月

山田・山本・他 5 人, 薬剤耐性誘導の影響を分析する探索的可視化システムの開発,

日本計算機統計学会第 31 回大会, 2017 年 5 月.

船山・山本・古山, Shiny によるソフトウェア開発プロセスの要因分析, 日本計算機統計学会, 2017 年 5 月

日比・河野・船山・梶田・山本, オープンデータと地域活動データの可視化, 第 15 回西東京統計研究会, 2016 年 3 月

〔図書〕(計 4 件)

山本義郎, 統計学序論 改訂版, 東海大学出版会 (2017/4/20)

山本義郎・藤野友和・久保田貴文, R によるデータマイニング入門, オーム社 (2015/11/21)

〔その他〕

ホームページ等

<http://stat.sm.u-tokai.ac.jp/~yama/>

6. 研究組織

(1) 研究代表者

山本義郎 (YAMAMOTO, Yoshiro)

東海大学・理学部・教授

研究者番号: 80301943