

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 21 日現在

機関番号：32644

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500358

研究課題名(和文)データの高度視覚化のためのグラフ表現法とグラフ表示環境に関する研究

研究課題名(英文)Studies on the methods and the environment of graphical representation for advanced data visualization

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) 3,900,000円、(間接経費) 1,170,000円

研究成果の概要(和文)：本研究は、インターネットを通じたデータの公開やオンラインデータなどについて、Webアプリケーションや新たな環境を利用することで可能となる統計データの可視化についての実装方法も含め行われたものである。本研究費による助成により、統計データの可視化法として新たな方法がいくつか提案でき、また、Webブラウザなどを利用したデータの可視化方法のプロトタイプを提案できた。この成果は、今後のデータの可視化の多様化に対して、その可能性を提供するものとなった。

研究成果の概要(英文)：This research is a theoretical study of the visualization methods for statistical data. It is also a proposal on how to implement these visualization to the current environment on PC. Results in the open data through the Internet is available, the need for visualization of statistical data is increasing. In this research, we propose a visualization method made possible to utilize the new features of the statistical analysis software. In addition, we realized the visualization of statistical data as a Web application. Awaiting to this grant of research funding, we could suggest new ways of the visualization method for statistical data. In addition, we were able to propose a prototype of a method for visualizing data using, such as a Web browser.

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：統計グラフ表現

1. 研究開始当初の背景

PC の処理速度とディスプレイの大型化に伴い、統計情報の可視化に関して大きな変化が起こった。その第一は、インタラクティブグラフィックスである。モザイクプロット (Friendly, 1994) や並行座標プロット (Inselberg, 1985) は一部のデータを選択し、ハイライトすること、また他のグラフとのリンクビュー (Tufte, 1983; Diaconis and Friedman, 1983) によりその能力を発揮する。これらのグラフ表示法以外にもインタラクティブ性、リンクビュー機能が有効に用いられるグラフ表現が提案されており、地図を利用した micromap (Carr, 2001) など様々なものが有効に活用され始めている。それらの成果は、Handbook of Data Visualization (Chen, Heardle and Unwin eds., 2008) により多面的にまとめられている。本研究グループとしてもその 1 章を担当し XML ベースのグラフ表示に関して研究成果をまとめている。また、そのようなグラフの実装に関しては XML を用いた SVG (2 次元ベクターグラフ用フォーマット)、X3D (3 次元モデリング言語) に代表されるオープンな規格の有用性を示していたが、HTML5 により次世代の Web における標準となった。統計ソフトウェア R においても、iplots ライブラリを用いることで、インタラクティブグラフが利用可能となり、統計グラフをとりまく環境が大きく変化しており、そのようなグラフ表示を活用した新たなグラフ表示が提案可能である。

統計情報の公開に対する Web の果たす役割も大きくなっており、統計解析システムとしてもインターネットにおいて、データベースを利用した Web ベースでの解析システムの必要性が高まっている。そのため Web における統計解析システムの研究も広く行われ、実用的なものも見られるようになった。Web におけるグラフ表現に関しては、静的な表示以外にも、Flash を用いた動的な表示などが増え、ダイナミックなグラフ表現も数多く見られるようになってきている。また、統計パッケージにおいても、多機能なインタラクティブなグラフが利用可能となっている。しかしながら、統計パッケージにおけるグラフ表現は、そのパッケージに限定されるものであったり、Web における動的なグラフの多くは、特別なシステムの上に構築されているもので、必ずしも一般的とはいえないのが現状である。

これまでの開発環境では、Web 上のグラフ表現は、Web 専用で作る必要があったが、現在で OS とブラウザを含んだアプリケーションの垣根は低くなり、HTML5 によるアプリケーション開発においては、よりその傾向が強くなることが予想される。ブラウザを利用した解析システムはオンライン利用でもオフライン利用でも可能であり、Web ブラウザを利用するシステムは、マルチプラット

フォームに対応する非常に有効な方法であることが認識されている。このように、表示するための環境に関する研究も必要である。特に、新しい表示環境として Web ブラウザの果たす役割が大きくなっており、Google Map, Google Earth, Google chart など Google の先進的な表示法を活かしたグラフ表現が可能となり、久保田 他 (2007) などでも研究されている。

2. 研究の目的

4 つの目的の個々に対して研究機関内に研究を行い、実現する内容とその詳細を以下に示す。

(1) インタラクティブグラフ・ダイナミックグラフに関する研究

散布図やヒストグラムなど古典的統計グラフに関しても、インタラクティブな処理が実装されたグラフは探索的データ解析 (EDA) のための強力なツールとなり得る。更に複数のグラフに対しリンク機能があると、データの様相を多角的に眺めることができることは、インタラクティブグラフの特徴としてすでによく知られていることである (Symanzik, "Interactive and Dynamic Graphics" in Handbook of Computational Statistics, Springer, 2004)。また、空間統計学や地理情報システムの分野でも、地図グラフと統計グラフの活用は非常に有効な視覚表現である。このような機能を持った統計グラフの Web 上での実現を行う。

(2) 探索的データ解析のための統計的グラフ表現法に関する研究

SOM (自己組織化マップ) や相関ルール (Association rule) などのデータマイニング手法による新たなデータの視覚化が提案されている。本研究では、大量データ、多次元データに対する視覚化手法について新たな方法を提案するため、まず最近の研究を調査し、実データが解析可能なよう実装し、新たな手法についての提案を行う。

(3) 統計解析ソフトにおけるグラフ表現とその実装に関する研究

スタンドアローンの環境では、クラス作成者は何らかのソフトウェアで作成することになる。そのため、新たなグラフの提案は、それらのソフトウェアのライブラリ等で提供する必要が生じる。フリーの統計解析環境 R のライブラリとしての提供を計画している。このライブラリには、サンプルデータも提供し、新たな表現手法の理解を助けるものとする。

(4) Web におけるグラフ表現とその実装に関する研究

Web におけるグラフ表現について、インタラクティブな希望を有するようになるための環境も Java, PHP, Flash, SVG などサーバー型から Web アプリケーション型まで多様に考えられる。本研究では、それらの長所、短所を評価し、方法論の提供を主眼に置いて、

成果を公表していく。まずは、現在分担者がすでに開発しているシステムについて、汎用的に利用できるものを公開し、順次新たな実装方法について公開を進める。

Web 上でのインタラクティブグラフも、いくつかのサイトで利用できるようになり、PHP また Google Map や Google Chart などの Web2.0 のテクノロジーも効率よく活用する方法についても研究し、ホームページ等においてその成果を公表する。

3. 研究の方法

研究遂行のための研究体制（役割分担）計画している 4 つの研究分野についてそれぞれ担当をおき、まず研究交流のための研究会を実施し、国内外の研究者との連携を確立する。その度、各担当者を中心に、新たな研究グループを構成し、それぞれのテーマに沿った形で研究を進めた。各グループで研究を推進し、年に 1 回の合同研究集会を実施することにより情報の共有と、新たな研究へのアイデアについて議論し、次年度の体制を整えた。研究成果は、ホームページを通じて公開するとともに、国内の研究集会、国際会議などで報告するとともに、論文投稿をおこなってきた。

研究遂行のための研究体制（役割分担）は以下の通りである。

- (1) インタラクティブグラフ・ダイナミックグラフに関する研究（担当：藤野、山本）
- (2) 探索的データ解析のための統計的グラフ表現法に関する研究（担当：山本）
- (3) 統計解析ソフトにおけるグラフ表現とその実装に関する研究（担当：飯塚、藤野）
- (4) Web におけるグラフ表現とその実装に関する研究（担当：山本、藤野）

研究目的にあげた個別の研究内容の計画に関して、以下の計画を推進した。

- (1) インタラクティブグラフ・ダイナミックグラフに関する研究

山本と藤野は SVG および X3D を用いた統計グラフに関して研究を行っており、山本・藤野 (2007) など で 紹介 し、<http://www.fwu.ac.jp/fujino/Xg4stat/> (XML-based Graphics for Statistics) には、グラフの例や利用例を与えている。しかしながら、インタラクティブグラフについては種類・機能ともにまだ十分ではない。平成 23 年度は基本的な統計グラフであるヒストグラム、ボックスプロット、散布図についてのプロトタイプを作成する。作成されたグラフについては、現在用意しているフリーの統計解析ソフトである R の XML グラフ作成パッケージの更新という形で随時公開していく。

- (2) 探索的データ解析のための統計的グラフ表現法に関する研究

H23 年度は、山本はクラスター分析の視覚的表示法について、散布図を基本としたアプローチにより研究を行う。さらに、これらの方

法を実現するシステムを、Java および SVG、X3D をベースに開発し、XML に基づいてリレーショナル・データベースとの接合を図る。

- (3) 統計解析ソフトにおけるグラフ表現とその実装に関する研究

飯塚と山本はこれまで統計グラフについてインタラクティブ性を考慮した活用法について、R における開発を行っている。それらの成果を、ライブラリとして提供する方法について整理し、汎用化して実装しやすいものから公開を始める。また、R におけるインタラクティブグラフとして、iplots ライブラリがあるがその活用例に関して情報交換を行う。

- (4) Web におけるグラフ表現とその実装に関する研究

インタラクティブグラフを実現するための Web テクノロジーについて長所・短所について考察を行う。特に、Web2.0 と呼ばれるさまざまな Web テクノロジーの有効活用について整理することから始める。

4. 研究成果

研究目的にあげた個別の研究内容の計画に関して、以下の計画を推進した。

- (1) 「インタラクティブグラフ・ダイナミックグラフに関する研究」として、山本と藤野は 2 次元グラフとして SVG および 3 次元グラフとして X3D を用いた統計グラフに関して研究を行ない、基本的な統計グラフについてプロトタイプを作成した。

- (2) 「探索的データ解析のための統計的グラフ表現法に関する研究」については、山本がデータマイニング結果の表示方法について研究を行った。

- (3) 「統計解析ソフトにおけるグラフ表現とその実装に関する研究」については、全員が統計グラフについてインタラクティブ性を考慮した活用法について、R における開発を行っているものについて新たな方法を試行した。

- (4) 「Web におけるグラフ表現とその実装に関する研究」については、HTML5 を用いたグラフ表現について各自調査を行い、その情報交換を行った。

研究代表者および、共同研究者の研究成果は、国内外の学会で報告し、雑誌にも投稿した。また、これまでの結果の一部のまとめとして、共立出版の R で学ぶデータサイエンスから「統計データの視覚化」として書籍をまとめ、5 月に出版した。

本研究で実現した可視化の例として、Rstudio の機能を利用して対話的に機能を実現したグラフが図 1 に、google マップでの地理情報の可視化の例を図 2 に与えた。

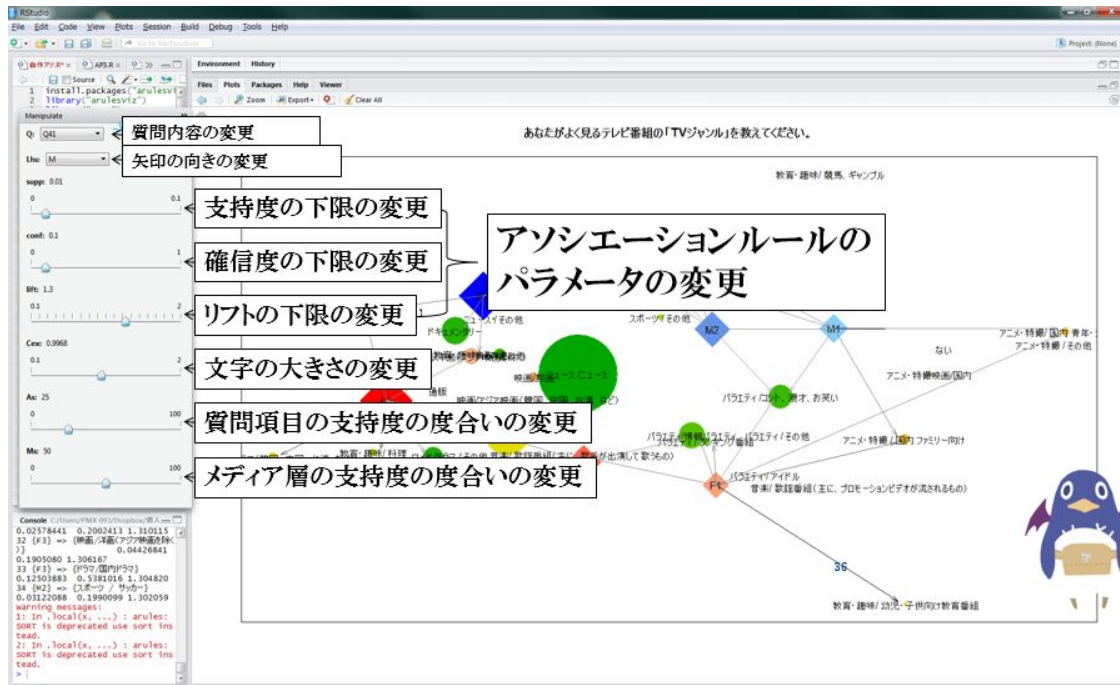


図 1 . Rstudio によるインタラクティブな操作が可能なグラフ

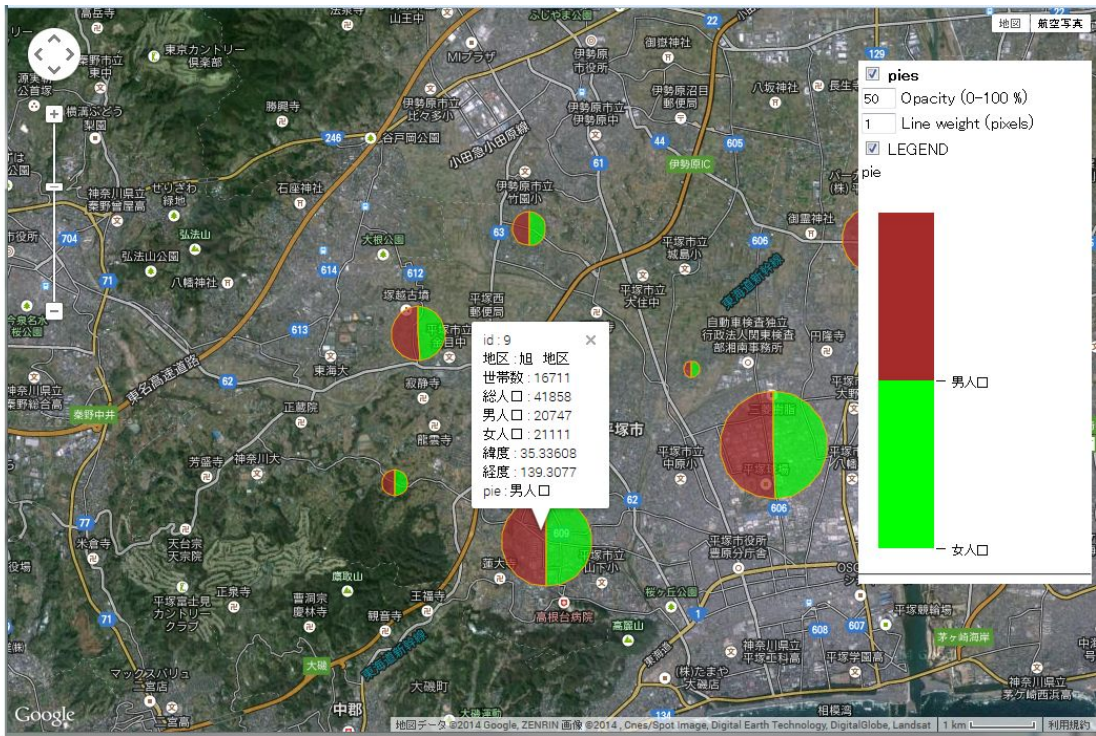


図 2 . Google マップによる地理情報の可視化例

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10件)

久保田 貴文, 藤野 友和, 富田 誠, 石岡 文生, 藤田 利治. 空間データベースを用いた隣接情報の作成と自殺データの集積性への応用, 統計数理, Vol.61 No.1, 2013, pp.167-176, ISSN:0912-6112, 査読有

S. Yamada, S. Munakata, K. Yagi and Y. Yamamoto, Visualization of zone activity and possession in soccer games, ICT and Knowledge Engineering (ICT&KE), 2013, 1-4, DOI:10.1109/ICTKE.2013.6756274, 査読有

K. Yagi, T. Funayama, Y. Yamamoto, and Y. Hoshi, Use of stand-alone and web-based systems for DNA analysis, ICT and Knowledge Engineering (ICT&KE), 2013, 1-3, DOI:10.1109/ICTKE.2013.6756273, 査読有

Kuroda, M., Mori, Y., Iizuka, M. and Sakakihara, M., Alternating least squares in nonlinear principal components, Computational Statistics, Vol.5 No.6, 2013, pp.456-464, DOI:10.1002/wics.1279, 査読有

T. Kubota and T. Tarumi. A Simulation study of Geometric Anisotropy Detection Methods. Journal of the Japanese Society of Computational Statistics, Vol.25 No.1, pp. 13-24, 2012, 査読有

T. Kubota, T. Hayashi and T. Tarumi. Cluster Analysis of Car Parking Data, and Development of their Web Applications. Communications of the Korean Statistical Society, Vol.18 No.4, pp.549-557, 2011. 査読有

T. Fujino, Y. Yamamoto, Statistical Graphics for Spatial Clustering Based on Inter-temporal Variation, Analysis and Modeling of Complex Data in Behavioural and Social Sciences, pp.1-4, 2012, 査読有

[学会発表](計24件)

Yoshiro Yamamoto, Clustering for Large Data Sets: Computation and Visualization, ISBIS-2012 International Symposium on Business and Bangkok, Thai Industrial Statistics, 2012年06月18日~2012年06月21日, Bangkok, Thailand

T. Fujino and Y. Yamamoto, Statistical Graphics for Spatial Clustering Based on Inter-temporal Variation, JCS-CLADAG, 2012年09月04日~2012年09月05日, Capri island, Italy

久保田 貴文, 富田 誠, 石岡 文生, 椿 広計, 自殺死亡の地域統計の可視化について, 日本計算機統計学会第26回シンポジウム, 2012年11月01日~2012年11月02日, 東京(東京大学)

松居 俊宏, 飯塚 誠也, アソシエーションルールの視覚化とアイテムの分類, 北海道大学情報基盤センター共同研究「ビッグデータの効率的情報縮約と発見的情報表現の研究」「大規模データに対するシンボリックデータ解析法に関する研究」(北海道大学情報基盤センター共同研究) 2013年2月

T. Kubota, H. Tsubaki, Text Mining for Tweet Data of Suicide in Japan, Joint International Session of the 2013 Korean Statistical Society Meeting, 2013

T. Fujino, Y. Yamamoto, Application Of The Parallel Coordinate Plot For Ranking Comparison Between Two Groups, 58th World Statistics Congress of the International Statistical Institute, 2011, Dublin.

宗像 昌平, 山本 義郎, アンケート集計結果の可視化と設問のクラスタリング, 応用統計学会2014年度年会, 統計数理研究所, 2014年5月

船山 貴光, 山本 義郎, 移動 OPS による打者の調子の波の可視化, 応用統計学会2014年度年会, 統計数理研究所, 2014年5月

八木 圭太, 山本 義郎, プロ野球打者の特徴の数値化と分類, 日本計算機統計学会第28回大会, 中央大学, 2014年5月

山田 実俊, 山本 義郎, 多肢選択アンケートのメディア層の反応の可視化, 日本計算機統計学会第28回大会, 中央大学, 2014年5月

亀岡 瑠, 船山 貴光, 宗像 昌平, 山田 実俊, 八木 圭太, 山本 義郎, 条件付きアソシエーションによる購入の特徴抽出, 日本計算機統計学会第28回大会, 中央大学, 2014年5月

宗像 昌平, 山田 実俊, 船山 貴光, 八木 圭太, 趙 斌, 亀岡 瑠, 山本 義郎, アンケートの可視化によるメディア層の嗜好の分析, 日本計算機統計学会第28回大会, 中央大学, 2014年5月

八木圭太, 山本義郎, プロ野球打者の特徴による分類, 日本分類学会第 32 回大会, 2014 年 03 月

宗像昌平, 山本義郎, カーネル密度推定を用いたクーポン共同購入サイトの解析, 日本計算機統計学会第 27 回シンポジウム, 熊本市, 2013 年 11 月

郡川貴子, 山本義郎, クラスタリングの計算の効率化と視覚的表示法について, 日本分類学会第 29 回研究報告会, 2011 年 8 月, 同志社大学

郡川貴子, 山本義郎, 大量データに対するクラスタリングの視覚的表示方法に対する提案, 日本計算機統計学会第 25 回大会, 2011 年 5 月, 函館

〔図書〕(計 3 件)

山本義郎, 飯塚 誠也, 藤野 友和, 共立出版, 統計データの視覚化, 2013, 219

山本義郎, 鳥越規央, 東海大学出版, 統計学序論, 2013, 197

山本義郎(松原望, 美添 泰人 編), 丸善出版, 統計応用の百科事典, 2011, 698

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

〔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

(2) 研究分担者

飯塚 誠也 (IIZUKA, Masaya)
岡山大学・環境学研究科・准教授
研究者番号: 60322236

藤野 友和 (FUJINO, Tomokazu)
福岡女子大学・文理学部・講師
研究者番号: 40364161

久保田 貴文 (KUBOTA, Takafumi)
統計数理研究所・リスク解析戦略研究所・
特任助教
研究者番号: 30379705