

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 6 月 14 日現在

機関番号：36102  
 研究種目：基盤研究(C)  
 研究期間：2009～2012  
 課題番号：21500285  
 研究課題名（和文） データ解析システムにおけるコラボレーション支援機能に関する研究  
 研究課題名（英文） A research of the collaboration work supporting system for the data analysis  
 研究代表者  
 山本 由和 (YAMAMOTO YOSHIKAZU)  
 徳島文理大学・理工学部・准教授  
 研究者番号：80240133

研究成果の概要（和文）：本研究では、データ解析システムにおけるコラボレーション支援機能に必要な機能とそれを実現するためのソフトウェアについて研究を行い、その成果を確認するためのソフトウェアを作成した。この成果を利用して、一般的な使用を目的とした通信フレームワークを Java 言語によって作成した。このフレームワークの設計と実装、評価結果については、学会発表および雑誌論文により報告した。

研究成果の概要（英文）： In this research, we had tried to develop the collaboration work supporting system for the data analysis. Therefore, we discussed required functions and software technologies and implemented software using them. We created a framework for general use by the result. We had gave a talk and papers about the design and implementation of the framework and the evaluation result of some software.

## 交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009 年度	900,000	270,000	1,170,000
2010 年度	1,200,000	360,000	1,560,000
2011 年度	600,000	180,000	780,000
2012 年度	700,000	210,000	910,000
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：情報学

科研費の分科・細目：情報学・統計科学

キーワード：Web サービス・Java・Jasp・サーバ/クライアント

## 1. 研究開始当初の背景

(1) 本研究に関連する国内・国外の研究動向及び位置づけ

統計解析ソフトウェアに関連した国内・国外の研究は R (<http://www.r-project.org/>) を利用したものを中心として多くの研究が行われている。しかし、本研究のようなコラボレーション機能に関するものは見当たらない。ソフトウェア開発においては、

NetBeans (<http://www.netbeans.org/>) や Eclipse (<http://www.eclipse.org/>) などの開発環境では、開発者間のコラボレーション機能が搭載されているものもある。

(2) これまでの研究成果を踏まえ着想に至った経緯

現在、われわれは、汎用的な統計解析システム Jasp (Java based statistical

processor)の開発を行っている。この作業は、四国と関東という地理的に離れた場所のメンバーがバージョン管理ツールなどを利用して行っている。そこで、データ解析にもネットワークを利用できれば、さらに便利になるのではないかと考えるようになった。一方、近年のコンピュータネットワーク技術が進歩したおかげでインターネット環境においてもデータ解析システムを利用できるようになったと判断したため、本研究のような機能の着想に至った。

## 2. 研究の目的

本研究の目的は、データ解析のためのコラボレーションシステムを開発することである。これは、複数のユーザが共同でデータ解析を行うための支援機能を持ったデータ解析ソフトウェアの開発である。

データ解析は、“解析対象のデータの専門家”や“統計学の専門家”などによって行われることが多い。あるデータを解析する場合には、そのデータに関する固有知識と統計学の知識の両方が必要不可欠である。固有知識は、解析対象のデータの専門家が知っているものであり、統計学の知識は、統計学の専門家が持っている。しかし、これら両方の知識を持った専門家がデータ解析を行うことは非常に少ない。そのため、データ解析は、複数の専門家が協力して行うことが多い。

複数の専門家によってデータ解析を行う場合を考える。統計学の専門家が探索的に行った解析結果を解析対象のデータの専門家と一緒に検討を行う。ここで、固有知識の観点から解析結果に問題があった場合には、統計学の専門家は、解析をもう一度行わなければならない。これは、効率的な作業とは言えない。一方、最近のデータ解析では、計算やグラフを描くためにソフトウェアを必ず利用する。もし、このソフトウェアが複数のユーザによるデータ解析支援機能を持っていれば、効率的に解析を行うことが期待できる。さらに、データ解析に関する教育に利用することもできる。

データ解析ソフトウェアは、商用のものから無償で利用できるものまで多くの種類がある。しかし、コラボレーション機能を持ったデータ解析ソフトウェアは、ほとんど見ることができない。そこで、本研究では、複数のユーザが共同でデータ解析を行うための支援機能の開発を考える。特に、

- (1) 遠く離れた場所の複数のユーザが共同でデータ解析を行う
  - (2) データ解析の教育に有効利用する
  - (3) 計算資源の共有と有効利用を行う
- などを可能にするために必要な機能と実装するためのソフトウェア技術についての検討を行う。

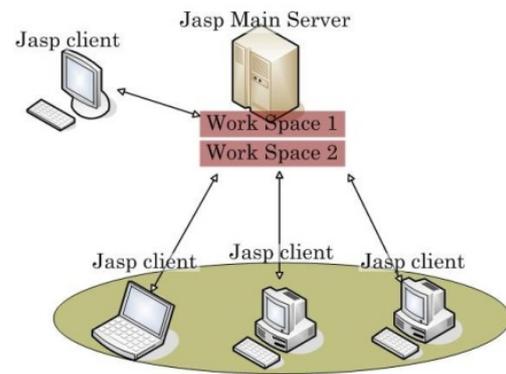


図 1: コラボレーション機能

## 3. 研究の方法

- (1) 図 1 は、本研究で考えているコラボレーション機能である。これを実現するために、まず、データ解析のコラボレーション支援のために必要な情報の整理を行う。ある程度の検討を行った後で、インタプリタに必要な機能の検討とコラボレーション支援機能のために必要なネットワーク機能の検討も開始する。この時に、ソフトウェア開発のために Eclipse などで行われているコラボレーション支援機能や Ajax (Asynchronous JavaScript + XML) 技術などを参考にする。

- ① 複数のユーザによって 1 つのデータ解析を同時に行うためには、排他制御を考えなければならない。あるユーザが参照している変数の値を他のユーザが変更することを許可しないことや、計算結果をある変数に代入することによってすべてのユーザが参照できるようにすることなどである。このような機能をインタプリタが持つことによって、対話的に命令を実行してデータ解析を行う場合や事前に作成したプログラムを実行する場合でも、複数のユーザによる共同作業が可能となる。
- ② 複数のクライアントプログラムが 1 つのサーバプログラムとの通信によって 1 つ仕事を行うためには、同期制御と通信方法の検討が必要である。同期制御は、サーバプログラムとクライアントプログラムが持つ情報に矛盾が発生しないようにするために必要である。あるクライアント上で変化した値をできるだけ早くサーバに送り、サーバプログラムはその変化をできるだけ早くクライアントに送る。そして、クライアントは、表示の更新を行わなければならない。ここで、変化するのは、数値や表示しているグラフィックスなどが含まれる。通信方法は、通信量と実装方

法の両方のことを考慮して検討しなければならない。

- (2) 排他制御と同期制御を考慮しながら、サーバプログラムとクライアントプログラムとの通信に Web サービスを利用することについて検討する。
  - ① データ解析のコラボレーション支援機能のために必要なこととして、計算エンジンとしての利用を考えている Jasp の通信機能の拡張である。Jasp は、サーバ/クライアント型のシステム構成となっているが、複数のクライアントプログラム間で情報の共有を行うことは考えられていない。しかし、この機能は、データ解析のコラボレーション支援機能の実現のためには必要不可欠なものである。
  - ② データ解析を考えた場合には、排他制御と同期制御の対象となる情報には、数値やグラフィックスなどがある。通信方法は、インターネット環境において、安全に利用できることと、高速な通信が可能でなければならない。
  - ③ Web サービスによって、以上のような条件を満足するような機能の実装を行いながら設計を検討する。データ解析において共有しなければならない情報としては、数値や文字によって表されたデータの集合とグラフィックが考えられる。これらを Web サービスによって送受信しなければならない。一方、Web サービスは、送受信できる情報に関してある程度の制約がある。これらの条件を考えた場合、利用する Web サービスを実装するためのソフトウェアの選定とそのソフトウェアを利用した設計についての検討を行う。
- (3) Jasp のサーバプログラムとクライアントプログラムとの通信にこれらの技術を利用することを具体的に考えて、実装作業を行う。

これには、RESTful Web サービスと HTTP によるプッシュ型の通信の実装に利用するソフトウェア技術を決定する。次に、この実装に適したサーバプログラムとクライアントプログラム間の通信方法とデータ形式を決定する。これには Java 言語を利用する。これは、Jasp のサーバプログラムとクライアントプログラムは、Java 言語によって実装されているためである。

この作業と同時に、運用面を考慮したサーバプログラムとクライアントプログラムの設計も検討する。これには、オンラ

インゲームプログラミングの技術を参考にする予定である。オンラインゲームプログラミングは、特にネットワークの高度な知識が必要とされているものである。また、運用面では非常に多くのユーザの接続を高速に処理する超大規模サービスの代表例である。これらの技術から本研究で考えている通信機能に応用できるものを、データ解析のコラボレーション支援機能のために必要なものという観点から検討する。

#### 4. 研究成果

Java 言語によって実装されたクライアントプログラムとサーバプログラムを Web サービスによって通信を行う一般的な利用も可能なフレームワークを作成した。このフレームワークは、プッシュ型の通信を行えることと RESTful Web サービスであることが特徴である。既存の Java アプリケーションをクライアントプログラムとサーバプログラム分けて、Web アプリケーションにすることも考えている。

このフレームワークの実装には、次のようなソフトウェアを利用した。プッシュ型の通信を行うために、サーバ側は Atmosphere、クライアント側は AsyncHttpClient を利用して、WebSocket による通信を行えるようにした。また、RESTful Web サービスに対応させるために Jersey を利用した。サーバプログラムとクライアントプログラム間でやり取りされるデータ形式には JSON (JavaScript Object Notation) を採用した。

このフレームワークを Jasp に適用することによって、インターネット環境におけるデータ解析のためのコラボレーション支援機能を実現できる。さらに、いくつかのアプリケーションの作成して、有効性を検討した。これらの成果については、学会や雑誌論文において発表した。

#### 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 3 件)

- ① 澤村・河井・磯本・山本、Web サービスを利用した分散処理アプリケーションの開発、徳島文理大学研究紀要、第 82 号、査読無、pp. 65-71, 2011
- ② 樋笠・磯本・山本、サーバプッシュ技術を利用した Web アプリケーションの開発、徳島文理大学研究紀要、第 82 号、査読無、pp. 73-80, 2011
- ③ Y. Yamamoto・J. Nakano・T. Fujiwara, Parallel Computing in a Statistical

System Jasp, Computational Statistics, Vol. 25, No. 2, 査読有, pp. 291 - 298, 2009.

[学会発表] (計 12 件)

- ① 山本・中野, 主成分分析のための平行座標プロットとその集約的シンボリックデータへの拡張, 2012 年度統計関連学会連合大会, 2012 年 9 月 11 日, 北海道大学
- ② 山本・中野・藤原・尾崎, 可視化とネットワーク分析による閲覧・購買行動の解析, 日本計算機統計学会第 26 回大会講演論文集, 2012 年 5 月 12 日, 高松
- ③ 磯本・山本, Web サービスを利用した通信ライブラリの作成, 平成 24 年度電気関係学会四国支部連合大会, 2012 年 9 月 29 日, 高松
- ④ Y. Yamamoto・J. Nakano, Statistical Graphics for Aggregated Symbolic Data, In Bulletin of the International Statistical Institute 58th Session, Dublin, 2011 年 8 月 24 日, Dublin Ireland
- ⑤ Y. Yamamoto・J. Nakano・T. Fujiwara, A visualization of aggregated symbolic data, In The 2011 Taipei International Statistical Symposium and 7th Conference of the Asian Regional Section of the IASC, 2011 年 12 月 17 日, Taipei Taiwan
- ⑥ 山本・中野・藤原・尾崎, 大量 pos データの可視化と解析, 日本計算機統計学会第 25 回大会, 2011 年 5 月 8 日, 函館
- ⑦ 山本・中野, 集約的シンボリックデータの可視化, 2011 年度統計関連学会連合大会, 2011 年 9 月 6 日, 九州大学
- ⑧ J. Nakano ・ Y. Yamamoto, Data Visualization and Aggregation, Yves Lechevallier and Gilbert Saporta, editors, COMPSTAT 2010 Proceedings in Computational Statistics, 2010 年 8 月 31 日, Paris France
- ⑨ 藤原・山本・中野, 統計ライブラリのシステム間での利用とその実装, 日本計算機統計学会第 24 回シンポジウム, 2010 年 11 月 12 日, 箕面
- ⑩ 山本・中野, シンボリックデータの視覚化, 2010 年度統計関連学会連合大会, 2010 年 9 月 8 日, 早稲田大学
- ⑪ 山本・中野, 統計グラフィックスと集計, 2009 年度統計関連学会連合大会, 2009 年 9 月 9 日, 同志社大学
- ⑫ 橋本・山本, 統計グラフィックスにおける平行処理技術の利用, 平成 21 年度電気関係学会四国支部連合大会, 2009 年 9 月 26 日, 愛媛大学

[その他]  
ホームページ等  
<http://se.bunri-u.ac.jp/~yamamoto/>

## 6. 研究組織

### (1) 研究代表者

山本 由和 (YAMAMOTO YOSHIKAZU)  
徳島文理大学・理工学部・准教授  
研究者番号: 80240133

### (2) 研究分担者

なし

### (3) 連携研究者

なし