

氏 名 藤 原 丈 史

学位（専攻分野） 博士(学術)

学 位 記 番 号 総研大甲第570号

学位授与の日付 平成14年3月22日

学位授与の要件 数物科学研究科 統計科学専攻

学位規則第4条第1項該当

学 位 論 文 題 目 統計解析システムJaspの言語と拡張機能に関する研究

論 文 審 査 委 員 主 査 教授 馬場 康維  
教授 大隅 昇  
教授 田村 義保  
教授 中野 純司  
助教授 泰地 真弘人  
教授 許 文烈（成均館大学）

## 論文内容の要旨

本論文は、新たに作成した統計解析システム Jasp (Java based Statistical Processor) について述べたものであり、特に、新しい統計解析言語と拡張機能、およびその実装法を提案したものである。Jasp は、現在のコンピュータ環境を統計解析に活かすとともに、急速に発展し続けるコンピュータ技術にも迅速に対応できる汎用的な統計解析システムを目的として開発された。幅広い層のユーザが、それぞれの目的や必要性に応じて、快適に統計解析が行なえるような、操作性、機能、拡張性を目標にし、進んだユーザインタフェース、統計解析言語、分散処理、多様な拡張機能を設計および実装した。また、様々なコンピュータ上で稼働することも特長である。

ユーザインタフェースとしては、視覚を利用して容易に操作できる GUI(Graphical User Interface) と、統計解析言語を用いて複雑な処理も柔軟に実行できる CUI (Character User Interface) とが互いに関係を取り、等価的な操作を可能にした。したがってユーザは目的に応じて自由に 2 つのユーザインタフェースを選択できる。また、GUI では、解析履歴が階層構造で参照できるオブジェクト表示や、ドラッグアンドドロップによる解析の実行といった機能を持たせ、より直観的な操作を可能とした。

統計解析言語は、関数を中心としたプログラミングが可能で、さらに、それらをまとめるためのオブジェクト指向プログラミングも行なえるように設計した。これは、システムの初心者が使う場合や簡単な解析については、記述しやすい関数を中心とするプログラミングを行ない、熟練者が再利用や統計手法の整理などを考慮にいたったプログラミングを行なうときには、クラスを利用するようにしたものである。プログラムした関数を変更することなく、ひとまとまりの統計手法としてひとつのクラスにして整理を行なうことは、これまでの統計解析言語では考えられていない。ここでは、この機能が統計手法を系統的に整理するために重要であることを示した。

分散処理では、特にインターネットの利用を考慮した。従来のサーバ/クライアント方式を拡張し、ひとつのクライアントが複数のサーバを利用できるように設計した。これにより、ユーザのクライアントマシンの性能にかかわらず、複雑な計算や処理が可能である。また、Java アプレットを使用することで、Web ブラウザをクライアントとして使用することも特長のひとつである。

Jasp は Java 言語で作成されているので、Java が利用できる環境なら実行できる。したがって Windows マシンから 高速の Unix ワークステーションまで、さまざまなプラットフォーム上で利用可能である。さらに、Java ライブラリによる拡張、C 言語や Fortran などのネイティブ言語とのインタフェース、Microsoft Excel などの COM (Component Object Model) とのインタフェースなどの拡張性も備えている。

以上のように、Jasp はこれまでにない、またはより洗練された多くの特長および利点をもつが、これを実現するための方法も重要である。本論文では特に、統計解析言語、および拡張機能の実装のための方法を提案した。

Jasp の統計解析言語 (Jasp 言語) は、プリプロセッサアプローチという方法を用いて実装した。プリプロセッサアプローチは、実際の処理を行う既存の言語(ベース言語)を選択し、新たな言語の処理は、そのプログラムをプリプロセッサによりベース言語に変換した後、ベース言語で実行を行なうものである。この方法によると、既存言語を利用し、開発、保守にかかわるコストを押えると同

時に、短い期間での設計、開発が可能になる。Jasp 言語の実装は、ベース言語として Pnuts を採用し、基本的な言語仕様はできるだけ Pnuts を利用し、統計解析言語として必要な変更や拡張をプリプロセッサアプローチによって実装した。Pnuts は、修得が容易で単純な文法を持つ言語で、Java 言語で作成されており、当然のことながら Java 言語と非常に親和性が高く、また拡張性も高い。Jasp の実装過程を通して、プリプロセッサアプローチは、統計解析言語を設計、実装するために有利な方法であることが確認できた。

統計解析システムにおいては、システムの組み込み機能として、多様なユーザの望む機能や操作性などを、すべて提供することは不可能である。さらに、コンピュータ技術の急速な進歩は現在も絶えず続いており、これに追随し続けることは容易ではない。また、新しい技術への対応だけでなく、既存の資源、すなわち過去のシステムのプログラムやライブラリの再利用に対応することも重要である。そのため、Jasp においては、新たな機能は Java ライブラリを利用して追加できるようになっている。現在、数多くの Java ライブラリが作成され、インターネット上で提供されている。このような Java ライブラリを、Jasp ではクラスファイルのまま直接利用することができる。また、その機能を用い Java でプログラムを開発すれば、システム自体をユーザ自身で拡張することも可能である。開発者だけでなく、ユーザが自分の要求を実現するようにシステムを拡張、変更できるというのは大きな利点である。

既存の資源の再利用については、Jasp では 3 つの方法を提案した。第一に、C 言語や Fortran などのネイティブ言語へのインタフェースである。統計学の分野では Fortran によるライブラリが数多く存在しており、これを無視することはできない。例えば、統計数理研究所で開発された TIMSAC は、Fortran で書かれた時系列解析のライブラリであり、現在でも幅広い分野で使用されている。Jasp ではそのような既存ライブラリを、ほとんど手を加えることなく利用することが可能である。なお、TIMSAC は Jasp の基本ライブラリとして取り込まれている。第二に、Jasp では、COM コンポーネントへのインタフェースを実現した。COM は、Microsoft が定めたオブジェクト呼出しに関する規約であり、Microsoft Office を始め、Windows のアプリケーションのほとんどはこの規約に基づいて作られた COM コンポーネントである。Jasp では、COM コンポーネントへの一般的なインタフェースを持っている。例えば、Jasp から Excel を起動し、ワークシートのデータを Jasp へ読み込み、処理結果を再びワークシートに出力する、などの処理が可能である。第三に、Jasp は、統計解析システムである XploRe のプログラムを実行できる。XploRe は、豊富な統計ライブラリ、分かりやすい言語などのさまざまな特長を持つ統計解析システムである。Jasp では、この XploRe 言語で書かれたプログラムを、Jasp 言語のプログラムへと変換するトランスレータを利用することにより、実行することができる。本論分で提案されているこの機能は、異なる統計解析システム間の相互利用のための一つの方法である。

以上のように、Java 言語を用い、プリプロセッサアプローチを利用することで、急速に発達している最新のコンピュータ技術に対応できる使いやすい統計解析システムのプロトタイプとして、統計解析システム Jasp を構築した。

## 論文の審査結果の要旨

審査委員会は、数物科学研究科統計科学専攻の藤原文史氏の論文について、数物科学研究科における課程博士の授与に係わる論文審査等の手続き等に関する規定に基づき、公開の論文発表会を開催し審査を行った結果、下記の理由により博士論文として十分な内容を備えていると判断した。

### 論文概要：

本論文は、統計解析のための新しい言語と拡張機能を提案し、さらに Java 言語の特長を利用してそれを実装することを提案したものである。

論文は全5章と付録からなっている。第1章は新しい統計解析システムの必要性を述べたものである。特に、パーソナルコンピュータやインターネットの普及により変化した計算機環境を統計解析に活かすこと、さらに急速に発展し続ける計算機技術に迅速に対応できる拡張性が重要であることが強調されている。それらを実現するために新たに構築した統計解析システムである Jasp (JAVa based Statistical Processor) が第2章で説明される。Jasp の簡単な操作例、提案された統計解析言語、ユーザインタフェース、分散処理機能、拡張機能等が順に述べられる。第3章と第4章が本論文の中心をなす章である。第3章で新しい統計解析言語である Jasp 言語が提案される。関数を中心としたプログラム言語に、オブジェクト指向の機能を加えることが統計解析に有用であることを主張している。また、その実装のために、プリプロセッサアプローチと、Java 言語および Pnuts 言語を用いることを提案する。これにより、既存のライブラリを取り入れながら、比較的容易に新言語を構成できることが、その手順を例示しながら説明される。第4章は Jasp の拡張機能を述べる。ネイティブ言語(Fortran や C)へのインタフェースを用意することにより、TIMSAC(TIME Series Analysis and Control package)などの既存のライブラリを統合・再利用すること、Microsoft Windows 上のプログラムとのインタフェースである COM(Component Object Model)を利用して Excel などとの連携をはかること、トランスレータを用いて他の汎用統計解析システム(ここでは XploRe)のプログラムを利用すること、が提案される。第5章では、現在の Jasp に不足している機能(エラー処理やヘルプ機能等)と、現在、開発を進めている付加機能(データの可視化、データマイニング手法等)についての今後の展開に触れるとともに、本論文のまとめを述べている。付録では Jasp 言語の詳しい仕様などが述べられている。

### 論文の評価：

インターネットの普及など、計算機環境が急速に変化している現在において、それらの利点を十分に活用できる新しい統計解析システムは、データ解析ツールを必要とする人にとっては不可欠なものである。また、統計学者にとっても、自分で新しい統計手法を容易に実現できる統計解析システムは、研究のための道具として欠かすことができない。本論文はそのようなシステムのプロトタイプを構築を試みた好例であり、このような研究の方向は現代の統計学が扱うべき重要な課題の一つである。世界的に見ても、本論文と同様な問題意識の下で、いくつかの新しい統計解析システム作成のプロジェクト(例えば XploRe, Omegahat)が進行中である。それらの多くは、Java 言語を用いているが、その理由の一つは、この言語が最近の計算機技術の発展を的確に反映していることにある。本論文で報告された統計解析システム Jasp は、プロトタイプではあるが、この Java 言語を

効果的に用いた新しい統計解析システムの好例であり、その発展可能性を示しているという点で、高く評価される。

さらに、統計解析言語 (Jasp 言語) が、それを実装するためのシステム言語である Java と非常に密接に、しかしお互いの利点を阻害しない形で結合されており、そのため機能の拡張が容易に行えるように設計されている点は、これまでの統計解析システムに見られない新しい特長である。さらに、オブジェクト指向の考え方を、統計手法の整理という観点から導入したことも注目すべき点である。実行処理速度に若干の問題を残すとはいえ、プリプロセッサアプローチの導入により、多くの言語機能を比較的短期間で実現したことは評価される。また、Java の諸機能を用いて、多様な機能を統一的、かつ平易に実現できるよう設計されていることは、今後のこの種のシステム開発にとっても参考とすべきところが多い。このような理由で、本論文で提案された機能や実現方法は、新しい統計解析システムのプロトタイプとして、注目に値するものである。

本論文の内容に関して申請者は英文 1 編、和文 1 編の論文を学術誌に投稿しており、既に掲載が確定している。また、単位取得状況、研究指導を受けた状況等すべて順調であり本研究科の規定を満たしている。

以上により、総合的に判断して、本研究は計算機統計学の分野において有意義な貢献を果たしており、博士の学位を与えるのに十分な内容を備えているものと判断した。