技-1 345

汎用データ解析ソフトウェア S-PLUS と
アドオンモジュール S+WAVELETS
—データ解析、可視化とオブジェクト指向プログラミング環境—

田澤 司 (株) 数理システム)

Data Analysis Software S-PLUS,
and its add-on module S+WAVELETS
—Object Oriented Programming Environment for analysis and visualization—
Tsukasa TAZAWA

ABSTRACT

An introduction of objects oriented data analysis environment designed by Bell Lab., S language, and its enhanced product, S-PLUS. S-PLUS is now used in processing of data, exploratory data analysis, and data visualization in over 3000 domestic site. S+WAVELETS, developed for wavelet analysis using S language, is add-on module of S-PLUS.

Keywords: Software, Visualization, Data analysis, Object oriented, Wavelet analysis

1. はじめに

S-PLUS は汎用のデータ解析ソフトウェアであり、業種、業務を問わず研究開発の様々なフェーズで科学者、技術者の手軽な解析ツール、可視化ツールとしてすでに国内 3000 以上のサイトで利用されている。本稿では、その S-PLUS の機能概要と、さらにウェブレット解析用のアドオンモジュールである S+WAVELETS について紹介する。

2. S 言語と S-PLUS

S-PLUS の核となるのは、米国ベル研究所で開発されたオブジェクト指向データ解析言語「S」である。S-PLUS の 2000 以上の解析関数は、主にこの S 言語を用いて開発されている。ベル研究所では C や C++ などの言語、UNIX などの OS が開発されているが、S 言語の設計思想にもこれらの影響が色濃く現れている。すなわち、「プログラムの再利用がやりやすいように、プログラム単位を汎用かつ小さい機能単位で分割し、その組み合わせでより大きな機能を実現する」という点である。

上記はプログラミング環境としての設計思想であるが、解析ソフトウェアとしては探索的データ解析 (Exploratory Data Analysis) を容易に行える解析環境を目指したものになっている。すなわち、定型の解析というよりもむしろ、解析結果を見ながらモデルや手法の改良を行い、さらにそれを繰り返す、という作業を能率的に行える環境となっている。具体的にいえば、プログラムの再利用だけでなく、データの再利用も快適に行えるためのデータ管理方式が採用されている。

S 言語は 1984 年に原型ができ、その後何度か言語仕様の改定が行われている。最も最近の言語仕様は 1998 年の改定にともなう Version 4 である。このバージョンにより、S 言語は ACM の 1998 年 Software Award を一受賞している。

S-PLUS は、この S 言語をベースに (1) 様々な分野に向けて、各種の解析手法を多数追加実装し (2) データの可視化のためのグラフィックスを豊富に備え (3) コンピュータの専門家でなくとも利用できるようにユーザインターフェースを備えた (Fig.1) 商用ソフトウェアである。

Fig.1 一般的な S-PLUS 利用画面

3. 機能概要

以下に S-PLUS (S 言語) の主な機能と特徴を記す。

3.1 データハンドリング

S-PLUS は通常のテキストファイルなら可変長固定長、

デリミタやヘッダ&フッタの有無を問わずおおむねどのようなものも読み込むことが可能である。データに関する情報がファイル内に記されている場合も、S 言語で入カルーチンを作成すれば自動処理が可能である。また、Excel や 123、あるいは Matlab のようなアプリケーションのデータフォーマットを直接読み込める。ODBC インターフェースも実装しているので、データベースからのデータの直読みも問題ない。

読み込んだデータは、対話型の S 言語を一行ずつコマンド的に利用しながら、抽出操作や併合などのデータ操作を加えて、解析に適したように自由に加工することができる。

3.2 プログラミング

デバッグ環境を含めた、S 言語の完全なプログラミング環境を備えている。オブジェクト指向設計により、データクラスの設計が可能であり、様々なデータにあわせた処理（関数）を自動実行できるように定義することにより操作を簡単なものにすることができる。例えば S-PLUS での“plot”という関数は、データに適した作図処理を選択するように構築されている。

Windows 版ではオートメーションに対応しているので、外部のプログラムとの連携動作も容易である。

3.3 解析機能

予め用意されている解析機能としては、基本的な数学関数、線形演算、基本統計、多変量解析、時系列解析、信号処理、品質管理、シミュレーション、その他 2000 以上の関数が用意されている。

また、複素数型データ形式を備え、線形代数演算が得意なものも S-PLUS の特徴である。

3.4 グラフィックス

多種類のグラフィックスを備えている。S-PLUS はプレゼンテーション用のソフトウェアというより、むしろデータの観察のためのグラフィックスを解析の初期段階で効率的に多数出力できるという部分に力が注がれているので、この点いわゆる CG ツールとは一線を画する点に注意されたい。

また、こうしたグラフィックの軸やレイアウトなどのディテールを細かくカスタマイズすることができる。もちろん、こうしたカスタマイズは S 言語を用いればすべて一括定義可能である。

3.5 対応機種

Windows(95,98,2000,NT)、主要 UNIX、Linux に対応する。元来 UNIX をベースにした S-PLUS であったが、現在は Windows の方が利用本数としては増加している。また、最近では Linux 版が着実に増加している。

4. S+WAVELETS

S+WAVELETS は S-PLUS の優れた線形代数演算能力と、オブジェクト指向のプログラミング環境を利用してウェーブレット解析のために開発された S-PLUS アドオンモ

ジュールである。Wavelet 関数としては、table 1 に記してあるものが利用できる。

Table 1 wavelet 関数

型	Wavelet 関数名
Haar	“haar”
Daubelets	“d4” “d6” “d8” “d10” “d12” “d14” “d16” “d18” “d20”
Symlets	“s4” “s6” “s8” “s10” “s12” “s14” “s16” “s18” “s20”
Coiflets	“c6” “c12” “c24” “c30”

主な機能を下記に述べる。

- 離散 Wavelet 変換（2 次元を含む）
- 多重解像度解析（2 次元を含む）
- 探索的ウェーブレット解析
- Wavelet 変換を用いたノンパラメトリック推定
- Wavelet Shrinkage
- Wavelet パケット解析
- 最良基底選択 (Fig. 2)
- コサイン・パケット解析
- Matching pursuit 解析

Fig.2 2次元最良基底選択による画像処理

5. その他

最近ではこうした S-PLUS の機能（出力）を、WWW を通して広域に提供するサーバソフトウェア StatServer もリリースされている。ブラウザをインターフェースとした、手軽に利用できるデータの可視化システムへの活用が見込まれている。その他、アドオンモジュールとして GIS との連携モジュール、空間統計モジュール、数値的最適化のための数値計画法モジュールなどが人気を集めているようである。これらの詳細な情報は <http://www.msi.co.jp/splus/> から簡単に得られるので、ぜひ一度アクセスされることをお勧めする。