

[連載]フリーソフトによるデータ解析・マイニング 第8回

GGobi とデータの視覚化

1. GGobi とは

GGobi は、高次元データを対話型かつダイナミックな視覚化システム XGobi を Windows 上で使用可能にし、さらに色を柔軟に管理することや XML ファイルを扱うことができるように改良したものである。

主なグラフの種類は、2次元の散布図(scatterplot)、散布図行列(scatterplot matrices)、平行座標グラフ(parallel coordinate plot)、時系列グラフ(time series plot)、棒グラフ(bar charts)であり、各変数の組み合わせは自由に繰り返し置き換えることが可能である。また3次元、多次元空間で回転するダイナミックなグラフ環境が備わっている。

本稿では、まず Windows 環境上の GGobi について説明を行い、最後に R 上の GGobi パッケージ Rggobi について説明する。以下 GGobi を ggobi と記する。

2. ggobi のインストール

ggobi に関しては次に示す URL から情報を得ることができ、自由にダウンロードすることができる。

<http://www.ggobi.org>

<http://www.ggobi.org/Download.html>

ページを開き、図1に示すような画面の「Self-installing Windows binary」から ggobi をダウンロードし、ダウンロードされたソフトのアイコンをクリックすると ggobi が自動的にインストールされる。



図1 ggobi のダウンロードページの画面

インストールが終了するとデスクトップの画面に ggobi アイコンが生成される。

3. ggobi の操作とデータの視覚化

3.1 データの読み込み

ggobi のアイコンをクリックすると ggobi が起動され、図 2 のような ggobi.exe ダイアログボックスが開かれる。

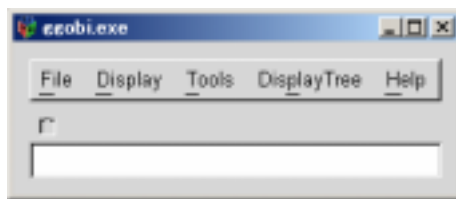


図 2 ggobi.exe ダイアログボックス 1

ggobi.exe のメニューの File Open をクリックすると図 3 のようなデータファイルを読み込むダイアログボックスが開かれる。ggobi をインストールするといくつかのデータファイルが ggobi のフォルダ中の data というフォルダに自動的にセットされている。

以下では、その中の flea というデータを用いて説明を行う。データ flea は複数の拡張子になっているが、ここでは拡張子が xml になっている XML ファイルを用いることにする。図 3 のようにデータ flea.xml を選択し、[OK] ボタンを押すと図 4 のように ggobi.exe のダイアログボックスが変わり、また図 5 のような散布図(XYPlot, Scatterplot)画面が開かれる。

これでデータが ggobi に読み込まれ、データを探索的かつ視覚的に考察を行うことができる。

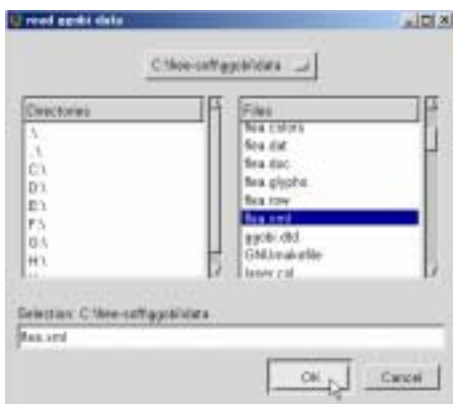


図 3 データセットを読み込む画面



図 4 ggobi.exe のダイアログボックス 2

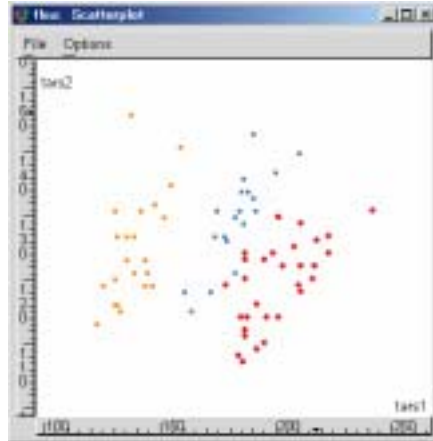


図5 ggobi の散布図画面 1

3.2 散布図

図4に示すような ggobi.exe ダイアログボックスの右段に示されているのは変数である。変数の X, Y ボタンを押すことで、その変数を横軸(X)、縦軸(Y)にした散布図が作成される。

メニューの ViewMode Identify をクリックし、マウスのポインタを散布図の画面の点に合わせるとその個体に関する情報が返される。

メニューの ViewMode Brush をクリックし、マウスポインタを散布図の画面上でクリックすると四角マーカがマウスのポインタの位置に移動される。その四角の範囲内にいくつの個体が集中しているかなどを確認することができる。ViewMode Brush を実行すると ggobi.exe のダイアログボックスが次のようになる。



図6 ggobi.exe 画面コピー 1

ggobi.exe のダイアログボックスの左段の[Choose color & glyph]ボタンを押すと散布図に用いる点の色と形を選択する図7のようなダイアログボックスが開かれる。

点の色や形を選択し、[Close]ボタンを押し、ggobi.exe のダイアログボックスの左下部にある[Color & glyph groups...]ボタンを押すと図8のようなダイアログボックスが開かれる。ダイアログボックス Symbol を右クリックすると散布図の点の色と形が変わる。このように散布図の点の色や形、サイズを好みに合わせて設定することができる。



図7 Choose color & glyph ダイアログボックス

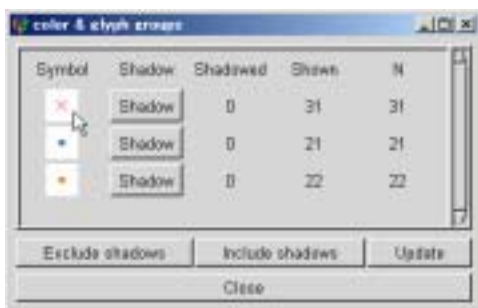


図8 Color & glyph groups...ダイアログボックス

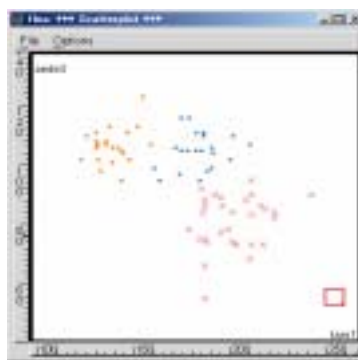


図9 ggobi の散布図画面2

3.3 グラフの種類

図10に ggobi.exe メニューの Disply を開いた画面コピーを示す。ggobi では図10に示す5種類のグラフを作成することができる。

Scatterplot は上記で示した散布図で、Scatterplot Matrix は図10に示すような散布図行列である。

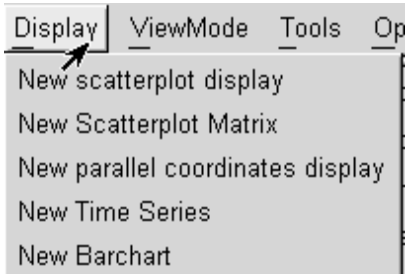


図 10 Display メニュー画面

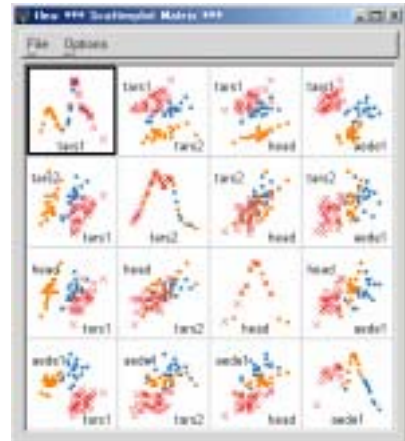


図 11 散布図行列

parallel coordinates display は図 12 のような平面座標図である。平面座標図は、1 つひとつの変数を横軸に等間隔で配置し、その値の大きさを縦軸で示した多次元データのグラフである。平方座標グラフでは、Option メニューで線を描くかか、描かないかを選択することも可能である。図 13 に線を描いていない平行座標グラフを示す。

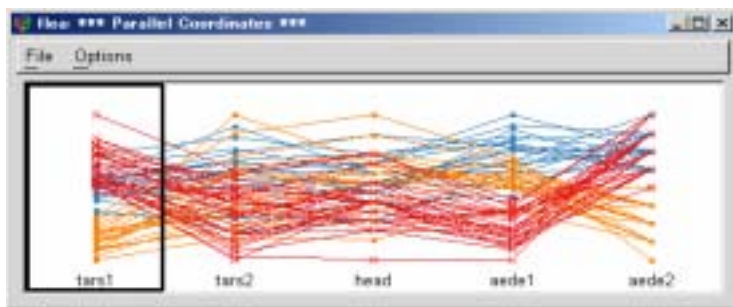


図 12 平面座標グラフ 1



図 13 平面座標グラフ 2

Barchart は図 13 のような棒グラフである。このような棒グラフでは、グループの重なりの度合いが確認できる。

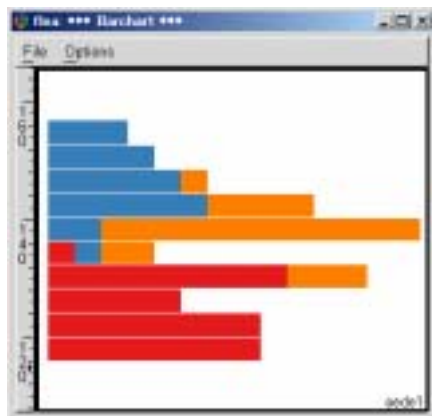


図 14 棒グラフ

Time Series は時系列のデータのグラフを作成する。時系列グラフを作成するためには、時系列データを読み込まなければならない。data フォルダの中には pigs.xml という時系列データがある。図 15 に pigs.xml の時系列グラフを示す。

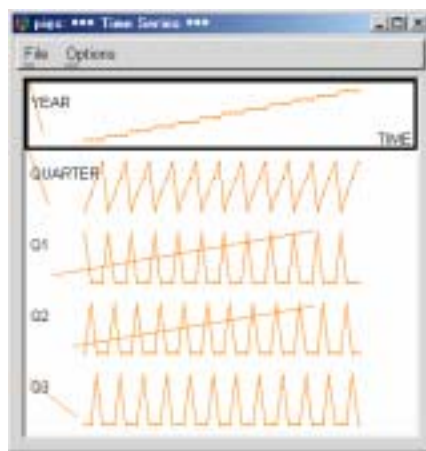


図 15 時系列グラフ

3.4 ダイナミックなグラフ

メニューViewMode のサブメニュー画面コピーを図 16 に示す。メニューの中の 1D Tour、2D Tour、Rotation、2 × 1D Tour はいずれも動的なグラフである。Rotation は 3 次元空間に 3 変数のデータに対応付け、角度を自動的に変えて見せる動的なグラフである。このようなグラ

フは特にクラスター分析の考察に有効である。図 17 に Rotation 画面コピーを示す。3 変数の組み合わせはダイアログボックスの変数ボタンを押すことで換えることができる。またダイアログボックスの右にグラフの動きの速度を調整するバーがある。バーを右に移動すると動く速度が早く、左に移動すると遅くなる。

2D Tour は 3 次元以上の変数を組み合わせた動的グラフである。

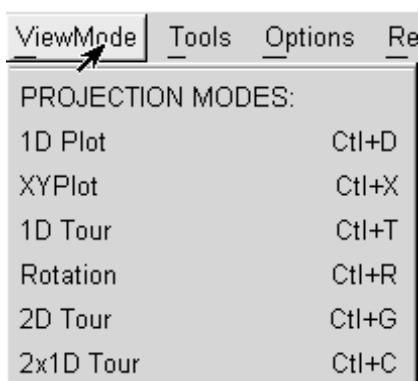


図 16 ViewMode メニュー画面



図 17 Rotation の画面

これ以外にも ggobi には多く機能があるが紙面の都合によりこれ以上紹介する余裕がないのが残念である。ggobi のホームページには 40 ページに上るマニュアルがある。

4. Rggobi

ggobi の R バージョンを Rggobi と呼んでいる。Rggobi のインストールは ggobi より若干煩雑であり、機能も多少限定されている。

Rggobi は ggobi と同じのサイトからダウンロードすることができる。

インストールの手順

- (1) 各自のマシンにダウンロードする。
- (2) 解凍する。
- (3) 解凍されたフォルダの中から、Rggobi フォルダをコピーし、R がインストールされている library フォルダの中に貼り付ける。R をインストールする際、インストールのフォルダについて特別な指定を行わなかった場合には、通常 library までのパスは C:\Program Files\¥rw****¥library である。
- (4) パス(Path) を通す(Windows XP を例としてその手順を説明する)。
 - A) 「スタート」 「コントロールパネル」 「システム」を右クリックし、「システムプロパティ」ダイアログボックスを開く。

- B) 「システムプロパティ」ダイアログボックスの「詳細設定」タブを右クリックする。
 - C) 「詳細設定」のダイアログボックスの[環境変数]ボタンを押し、「環境変数」ダイアログボックスを開く。
 - D) 「環境変数」ダイアログボックスの「システム環境変数 (S)」の PATH の行を右ダブルクリックし、「システム変数の編集」ダイアログボックスを開く。「システム変数の編集」ダイアログボックスの「変数値 (V)」のテキストボックスに Rggobi の dll ファイル(Rggobi の中拡張子が dll になっているファイル)が入っているフォルダまでパスを入力する。通常 dll ファイルは Rggobi¥libs の中にある。
 - E) [OK]ボタンを押し、すべてのダイアログボックスを閉じる。
- (5) コンピュータを再起動する。

Rggobi を使用するためには、R を起動し、まずパッケージ Rggobi をロードしなければならない。

```
>library(Rggobi)
```

次に iris データを用いて ggobi の使用方法を簡潔に説明する。

```
>data(iris)
>ggobi(iris[,1:4])
```

を実行すると、次の図のようなダイアログボックスとデータの散布図画面が開かれる。

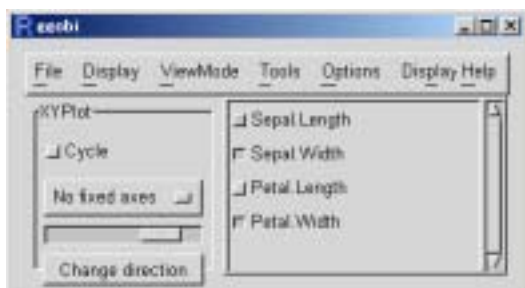


図 18 Rggobi のダイアログボックス

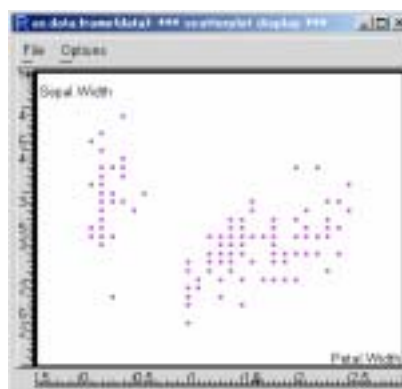


図 19 Rggobi の散布図 1

ggobi の主な機能は Rggobi で実行することが可能であり、操作の方法も基本的には同じであ

る。また R のコンソール上でコマンドにより操作するのも可能である。

XML 形式で作成したデータファイルを用いるためには、XML 形式のデータファイルを指定したフォルダの中に入れておかなければならない。

Rggobi のフォルダの中には data というフォルダがある。XML データファイルをフォルダ data の中に入れておくと、R に読み込むことができる。例えば、iris データの XML 形式のデータファイルを作成し library¥Rggobi¥data フォルダに入れると、次のコマンドで Rggobi に読み込むことができる。

```
>ggobi(system.file("data", "iris.xml", package="Rggobi"), args= "-noinit")
```

図 20 に iris の平面座標グラフ、図 21 に iris の動的グラフの画面コピーを示す。

iris の XML データセットを作成していない場合は、iris.xml の代わりに flea.xml など data フォルダの中にあるデータを用いて Rggobi を体験することができる。

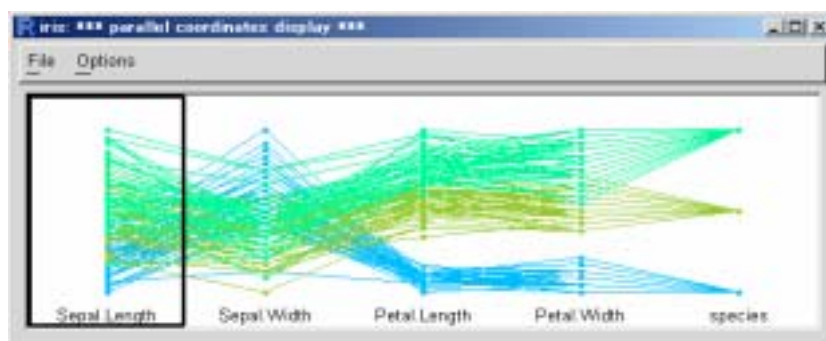


図 20 Rggobi の平行座標グラフ

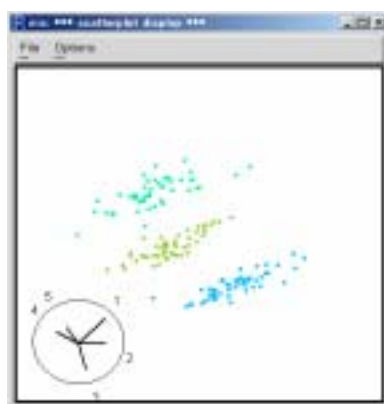


図 21 Rggobi の 2D Tour グラフ