

特集「統計データの可視化」にあたって

中野 純司[†] (オーガナイザー)

まったく未知の統計データを解析しなければならない場合、最初にグラフなどを描いてそのデータ全体の特徴を可視化することが多い。もちろん、平均などの基本統計量を計算することも行われるが、そのような少数の数値だけで複雑なデータの特徴を明らかにすることは難しいからである。適切に描かれたグラフは、データの特徴を一度に明らかにすることが可能である。

経験的に、人間の情報入力手段の中で視覚が他の感覚(聴覚、触覚など)に較べてきわめて多量の情報を入力できることは明らかである。一説によると 1 秒間あたりの情報伝達量は

視覚	10^7 bit
聴覚	10^4 bit
触覚	10^2 bit

と言われている。さらに、人間の視覚が非常に微妙なパターンを識別できることもよく知られている。例えば、顔の表情を見ればその人の喜怒哀楽の感情を一瞬で区別することができるのである。

このような優れた特徴をもつ視覚を、データ解析で利用することは自然であり、そのための統計グラフは古くから統計学における重要な手段であった。現代人の常識として、基本的な棒グラフ、円グラフ、線グラフなどは初等教育で、ヒストグラム、散布図などは基本統計学で教育されており、それらは実際に新聞などでもよく利用されている。また、これらより新しい統計グラフィックスとして、EDA(探索的データ解析)において幹葉表示や箱ひげ図などが提案され、積極的に利用されるようになった。

最近の計算機の進歩、特に安価で高性能なパーソナルコンピュータとインターネットの爆発的な普及は人間社会を劇的に変えつつあるが、統計グラフィックスにおいても事情は同じである。まず、グラフを非常に美しく容易に描くことが可能になった。それだけではなく、これらのグラフを紙の上に描いていた時代には不可能であったような対話的・動的な機能や、複雑な計算の結果を利用する機能も実現されてきた。さらに計算機の進歩は、解析すべきデータの巨大化、構造の複雑化を引き起こしており、それらを解析するための統計手法をも変化させている。例えばデータマイニングの勃興は、統計学における新しいパラダイムと考えてもよい。したがってこのような変化に対応した新しい統計グラフィックスや可視化手法が必要となってきている。本特集は、このような問題意識による日本における現時点での研究を集めたものである。

本特集の論文は大きく 3 種類に分類することができる。可視化実現のためのツールに関するもの、平行座標プロットに関するもの、GIS (Geographic Information System)に関するもの、である。

現代ではほとんどの統計グラフィックスは計算機によって描かれるが、そのためには計算機上の描画に関するツールを利用しなければならない。そのようなツールは発展を続けているが、その中でも XML と Java は特に注目を集めている。山本義郎 他の論文は XML およびそれに

[†] 統計数理研究所：〒106-8569 東京都港区南麻布 4-6-7

関連する技術の統計グラフィックスへの応用について述べたものである。山本由和 他の論文は Java による統計グラフィックスライブラリの設計と実装を説明している。

平行座標プロットは 1980 年代に考案された比較的新しい統計グラフィックスであるが、多量のデータに対する「ビジュアルデータマイニング」のツールとしてもよく用いられるようになっていく。熊坂夏彦 他の論文では平行座標プロットを拡張したグラフィックスが論じられるが、特に、計算機実行環境の部分が詳しく述べられている。本多啓介 他の論文は平行座標プロットを 3 次元空間に描画し、さらにいくつかの変数の値の大小で条件を付けたグラフィックスを描くことを提案している。

人間社会に関する情報の多くは特定の場所に付随する地理情報であり、それらを表示するためには、地図を利用する GIS (Geographic Information System) を用いることが普通になっている。さらに GIS は地理情報の空間的な統計データ解析のためにも利用される。亀川佳美 他の論文は Web 上の簡易 GIS を作成する試みである。久保田貴文 他の論文は表示に GoogleMaps を利用し、空間データ分析を行うことを提案している。小林郁典 他の論文は条件付きコロプレスマップと平行座標プロットを組み合わせることによって多変量地理データを表示している。

統計データも、それを表すための統計モデルも、ますます精緻・複雑になっており、その全体の特徴を式や数字などで捉えることは容易ではない。可視化は人間にとってもっとも直観的なインタフェースであるために、適切に利用すればそれらをわかりやすく一度に表示するために有用な手段となる。本特集が可視化という統計学の重要な一分野の研究をより盛んにする一助となれば幸いである。