氏　　名	本多啓介
学位（専攻分野）	博士（学術）
学位記番号	総研大甲第1150号
学位授与の日付	平成２０年３月１９日
学位授与の要件	複合科学研究科　統計科学専攻 学位規則第６条第１項該当
学位論文題目	A 3-Dimensional Extension of Parallel Coordinate Plot
論文審査委員	主　査　准教授　佐藤　整尚 教授　中野　純司 教授　田村　義保 准教授　金藤　浩司 教授　田中　豊

論文内容の要旨

The three dimensional parallel coordinate plot (3-D PCP) is a visualization method to detect hidden information in data by using human spatial perception. The 3-D PCP proposed in this dissertation can display several characteristics of multiple variables simultaneously. First, it can show all the information of each observation at a glance. Second, it can make some non-linear structures in data clear. Finally, it is useful to find piecewise linear relationships of variables and their conditions.

The basic idea of the 3-D PCP has already appeared in the parallel coordinate plot. The parallel coordinate plot can visualize multi dimensional data on a two dimensional plane. Coordinates of all variables are set in parallel. In the standard form of parallel coordinate plot the bottom position of each axis corresponds to the minimum value of each variable, and the top to the maximum value. One observation corresponds to one set of connected lines. Parallel coordinate plot can shows the characteristic between two adjoining axes of variables directly. If two variables have a correlation coefficient of 1, lines expressing observations are located horizontally. If two variables have correlation coefficient -1, lines of them cross in one point in the middle between two axes.

However, relations between two variables whose positions are apart more than two axes are not clearly shown immediately. Another serious problem of static parallel coordinate plot is that it is not easy to distinguish one observation from another when the number of observations is large. To solve these problems, several interactive techniques have been developed including highlighting by brushing operations. The 3-D PCP can show the same effect as the highlighting by brushing operation in a parallel coordinate plot by extending it into 3-dimensional space. We choose one variable as a reference variable, usually a response variable. The 3-D PCP places connected lines expressing observations in 3-dimensional spaces by sorting them according to the values of the reference variable. This observation-wise 3-D PCP representation is useful for illustrating the characteristics of observations such as outliers.

The 3-D PCP has another representation in which values of observations on each variable are connected by lines. It is called variable-wise connection representation and is useful to see relations between the reference variable and other variables. For example, the connected lines expressing the variable which has strong linear relationships with the reference variable are located around a straight line expressing the reference variable. It is well known that the scatterplot matrix can show relationships between variables clearly. However, if the number of variables is large, the single scatterplot elements become too small to be seen properly. The 3-D PCP can show more variables than a scatterplot matrix. It is sometimes more suitable to show characteristics of data simultaneously than using a static representation of scatterplot

matrix.

We note that the 3-D PCP can detect particular non-linear relations, i.e. interaction by two variables, through observation-wise connection representation. This relationship is detected by the special pattern of the angles of connected lines expressing observations. As explained earlier, a correlation coefficient near 1 or -1 between two variables whose axes are adjacent produces parallel or crossing patterns in parallel coordinate plots. Similar patterns can be detected in 3-D PCP. If we find the change of such patterns at specific values of the reference variable, we conclude that the structures of data are different for each region of the reference variables. In such cases, it is natural to divide the data into several groups in which the structures have no more changes. We propose to draw many 3-D PCPs corresponding to the groups simultaneously and use a lattice layout, which places many graphs on a grid. We show that they are useful to identify the interaction between two variables by analyzing simulation and real data.

We realize our 3-D PCP by using the Java language. Java has several advantages for implementing modern data visualization methods. It is a pure object oriented programming language and has well-designed standard graphics libraries which are useful to realize 2-D and 3-D graphics and interactive graphical user interfaces. These libraries can work as useful components of statistical graphics and be in incorporated by using so-called design patterns. Design patterns are suggested solutions to common problems often appearing in object-oriented software development. Our implementation is based on several design patterns for generality and reusability. Our software enables us to analyze data by utilizing advanced interactive operations given by Java. We show that the 3-D PCP and the software are expected to lead to new achievements in the field of data visualization.

This dissertation is set out as following. Chapter 1 surveys issues of information visualization and basic statistical graphics used in multivariate data visualization such as scatterplot, scatterplot matrix, 3-D scatterplot and parallel coordinate plot. It also discusses dynamic techniques of data visualization and existing software products for data visualization such as Mondrian, ParallAX, and GGobi. In Chapter 2, 2-D parallel coordinate plots are discussed. Important issues at visual data analysis with parallel coordinate plot are considered. We introduce several existing works for extending parallel coordinate plot into 3-dimensional spaces. In Chapter 3, we discuss our extension of parallel coordinate plot into 3-dimensional space, and several of its characteristics. We show usefulness of lattice layout to display several 3-D PCP at a time in Chapter 4. Chapter 5 analyzes three data sets by using our 3-D PCP. In Chapter 6, we explain details of our software design. Finally, concluding remarks are given in Chapter 7.

論文の審査結果の要旨

提出論文は全 7 章 73 頁からなり、英語で執筆されている。本論文では、多変量データを表示するための（2 次元）平行座標プロットを 3 次元空間に拡張することが提案され、それがデータ解析で有効であることが示されている。

第 1 章は序論で、情報可視化の総論から始まる。そのひとつであるデータ可視化においては、現在、統計グラフィックスとして、散布図、散布図行列、平行座標プロットなどがよく用いられていることが述べられ、またそれらの対話的操作が有用であることが説明される。同時に有力な統計グラフィックスソフトウエアも簡単に紹介される。第 2 章では、平行座標プロットがより詳しく説明される。例えば特有の操作として、一つの変数の大きさによって個体を表す折れ線の色を変えて表示する Zebra が紹介される。さらに平行座標プロットを 3 次元に拡張するためにこれまで提案されているいくつかの試みを紹介する。ここまでは序論である。第 3 章で出願者が提案する 3 次元平行座標プロット(3-dimentional parallel coordinate plot, 3-D PCP)が導入される。そのアイデアは 2 次元平行座標プロットを用いた統計解析でよく利用されている、ある変数に関する動的なブラッシング操作、または Zebra を、静的な 3 次元のグラフで表現するというものである。具体的には一つの変数を基準変数として選択して、それを 3 番目の直交座標にとる。そしてその座標の各観測値の位置にもとの平行座標プロットでその観測値をあらわす折れ線を配置する。すると外れ値などの各観測値の特徴が見やすくなるとともに、観測値ごとに折れ線を描く他に、変数ごとに折れ線を描くことによって、基準変数と他の変数の線形関係や区分的な線形関係のような非線形関係が直接表示できるようになる。第 4 章では、複数の 3-D PCP を 2 種類の格子表示することが提案される。ひとつの格子表示は、変数の平行座標の順序を変えることによる格子表示である。これにより、すべての変数を隣接させることができ、その結果、変数間の交互作用を可視化できる場合がある。そして観測値を表す折れ線をその角度によって色分けすることで、より視認性が高められる。もう一つはある変数（ここでは 1 または 2 個の変数）の大きさによってデータを分割して、各分割に含まれるデータに関する 3-D PCP を格子表示することである。これにより変数間の区分的な線形関係などを見やすく表示することができる。これらの格子表示が有効なことが人工的なデータを用いることによって示される。第 5 章では 3-D PCP を利用して 3 種類の実データが解析され、データのいくつかの性質が明らかにされる。第 6 章では、出願者が開発した 3-D PCP ソフトウエアの実装とそこでのデザインパターン技術の利用が説明される。第 7 章はまとめの章である。

本論文は、多変量データを表示・解析することのできる 3 次元平行座標プロットを提案したものである。（2 次元）平行座標プロットを 3 次元へ拡張することは自然なアイデアであるので、本論文にも述べられているようにすでにいくつか提案されている。ただしそれらが、観測値の挙動を見やすくすることを目標としているのに対して、出願者の提案では、変数間の関係（区分的な線形関係、2 変数間の交互作用）を可視化することを目標としている。このように、可視化ツールを探索的な非線形統計モデリングに利用できるようにしたという点は高く評価できる。そして、提案した可視化手法をソフトウエアとして実現している。3 次元グラフィックスとその対話的な操作を可能とするようなソフトウエアの実現は簡単なものではなく、提案した手法が有効に利用できることを示すアプリケーションを作成し、現実のデータに適用したことは統計科学の発展に寄与し、十分な評価に値する。また、これらの成果に関連する学術論文として、出願者が第一著者である和文論文が 1 本と査読付き国際会議録（英文）が 1 本、出願者が第三著者の和文論文が 1 本ある。

以上から、博士論文審査委員会は、出願者の学位請求論文が学位に十分値する水準にあると全員一致で判定した。