

多変量データの視覚化技法

— プラントグラフの提案 —

7K-7

菅原 護 池亀 一

富士通㈱ 通信事業推進本部

1. はじめに

ある事象の現状を把握し、そこから問題を見つけ出し何らかの改善行動をとるために、一般にその事象を色々な観点から多段階に評価することを行う。例えば、ある製品の出来映えを評価するために、その製品に対する評価項目をいろいろと用意し、ユーザに5段階で評価してもらうことを行う。このように、多くの項目を多段階に評価することによって得られたデータを「多変量多段階評価データ」と呼ぶことにする。

多変量多段階評価データは、非常に主観的であいまいであり、異常値や外れ値に相当するデータが多く混入していて信頼性が低く、多くの変量からなり複雑であるという特徴を持っている。このようなデータに対して、統計理論上の前提条件(例えば無作為性、正規性、独立性、等分散性など)を仮定した論理的な解析(例えば多変量解析など)を行うことによって、問題を見つけ出そうとすることは、危険であり意味がない場合が多い。

我々はこの多変量多段階評価データを、生データのまま視覚化することによって、人間の感性に訴え、人間の右脳に分析させて問題を発見するためのグラフ「プラントグラフ」を考案した。

本論は、このプラントグラフの開発思想、定義、および適用事例について述べたものである。

2. プラントグラフの開発思想

多変量データをグラフ化する手法は、顔形グラフ、体形グラフ、星座グラフ、木形グラフ、連結ベクトルグラフなど数多く提案されているが¹⁾、多変量多段階評価データを我々が要求する観点からグラフ化する手法は見当たらなかった。

一般にデータを分析する目的は、データを使って重要な問題を客観的に発見し、それを基に何らかの改善行動をとることにある。しかし、人間はかなり強い外的な刺激を受けないと、改善行動のような革新的でパワーを必要とする行動は起こしにくいという特徴を持っている。そこで我々は、複雑で、あいまいで、信頼性が低いという特徴をもっている多変量多段階評価データから、問題

を発見し、改善行動を起こさせるための手法として以下の要件を満たした手法「プラントグラフ」を開発した。

- ①データは、なるべく論理的な解析をせずに、生データのまま扱う。
- ②データを視覚的、刺激的に表現することによって、人間の感性(右脳)に訴え、問題を発見させる。
- ③複雑で、あいまいなデータを、日頃我々が見慣れている具体的なイメージに置き換えることによって、データの語らんとするところを、人間が直観的に理解できるようにする。
- ④ミクロよりマクロを見せることによって、全体像を一目でわからせるようにする。
- ⑤なるべく単純に表現することによって、誰もが簡単に同じように読めるようにする(客観的に読めるようにする)。

3. プラントグラフの定義と特徴

下記の原則に基づいて描かれたグラフをプラントグラフという。

【プラントグラフの5原則】

- ①多変量多段階評価データを樹形イメージで表す。
- ②各評価項目の評点の割合を作図データに使う。
- ③1つの評価項目のデータを1本の枝に対応させる。
- ④評点が低いと枝が下を向き、評点が高いと枝が上を向くように、各評点に対応させて枝の角度を設定する。
- ⑤枝の全長は一定とし(枝の全長×各評点の割合)によって、各評点对応の枝の長さを定める。

図1に、5段階評価データをプラントグラフで表す場合の一例を示す。

図1からもわかるように、プラントグラフは「評点が悪いと枝が下を向き枯れ木の状態になり、評点が良いと枝が上を向き元気な木の状態になる」という特徴を持っている。したがって、このプラントグラフからは、誰もが次のことを容易に読み取れる。

- ①木全体の感じをみることにより、評価対象の総合的な状態がわかる。
- ②各枝ぶりをみることにより、どこに重要な問題があるかがわかる。

③2枚のプラントグラフを重ね合わせるにより、他の集団との相対比較が容易に行える。

④木の右側の枝群と左側の枝群の分類に意味を持たせることにより、1枚のグラフ上で2つのカテゴリ間の比較を行える。

⑤時系列データの場合、データの古い順に幹の下の方から枝を割り付けることにより、経時変化をみることができる。

評点5が100% 評点3が100% 評点1が100%

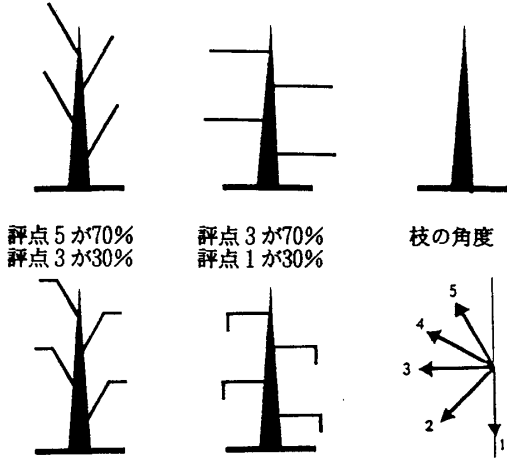


図1 5段階評価データの評点と枝ぶりの関係

4. プラントグラフの適用事例

プラントグラフをソフトウェアの開発データに適用した事例を述べる。現在我々は、各プロジェクトごとにソフトウェアの品質や開発工程の状態を1~5の5段階で評価するシステム「あゆみ」を運用している^{2) 3) 4)}。この評価結果は、プロジェクトごとに表1の評価表にまとめられる。しかし、このままの情報では、プロジェクトの総合的な状態や問題の有無・問題の箇所などが非常にわかりにくい。

図2は、表1のデータを使って評価要素のレベルでプラントグラフを描かせたものである。プロジェクトAは比較的良好なプロジェクト、Bはあまり良くないプロジェクトの例である。プロジェクトの総合的な状態や問題の多い評価要素が一目で読み取れると思う。2つのプロジェクトを直接比較したい場合は、2つのグラフを重ね合わせるとより効果的である。

図3は、プロジェクトAのデータについて評価基準のレベルでプラントグラフを描かせたものである。評価要素のレベルでは、あまり問題がなさそうにみえたこのプロジェクトも、評価基準のレベルでみると、ところどころに問題があるのがわかる。このように評価項目が多い場合は(20以上の場合は)、評価項目を階層化し、大分類でプラントグラフを1つ描かせ全体の傾向をみるようにし、中分類ごとに小さなプラントグラフをいくつか描かせ、具体的な問題を見つけるようにすると良い。

表1 評価表

評価要素	評価基準	評点の割合 (%)				
		1	2	3	4	5
外観検査	納入品	0.0	0.1	4.7	20.1	75.0
	ドキュメントの体裁	0.0	0.1	3.1	17.8	79.1
	ファイルの体裁	0.0	0.0	9.2	32.8	58.0
	納入要件	0.0	0.0	1.4	41.4	57.2
	検印	8.5	0.0	0.0	0.8	90.9
レビュー実施度	設計レビューの量	2.0	0.0	24.1	37.8	36.0
	製造レビューの量	3.4	0.0	25.7	41.3	29.6
	設計レビューの質	0.5	0.9	15.5	66.2	16.9
	製造レビューの質	1.1	1.4	16.3	64.2	17.0

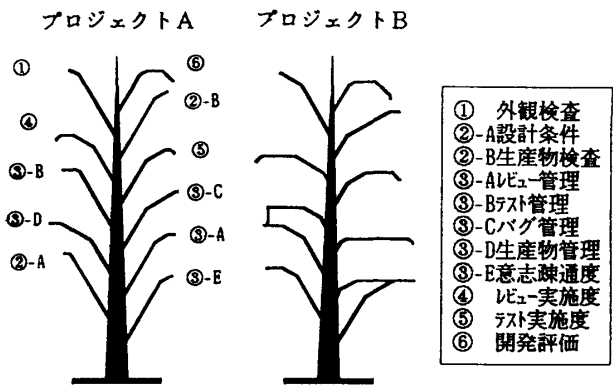


図2 評価要素のプラントグラフ

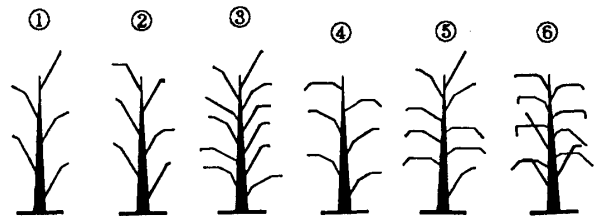


図3 評価基準のプラントグラフ

5. まとめ

多変量多段階評価データを視覚的に表現し、人間の感性(右脳)に訴え、問題を発見させる手法「プラントグラフ」を提案した。

【参考文献】

- 1) 脇本和昌 他, 「多変量グラフ解析法」朝倉書店
- 2) 菅原, 小栗, 中村, 「外注ソフトの品質向上活動「あゆみ」」, 第9回ソフトウェア生産における品質管理シンポジウム発表報文集, pp71-78, (1989)
- 3) 池亀, 阿部, 菅原 「ソフトウェア品質の5段階評価法」, 情報処理学会36回全国大会論文集, pp.943-944, (1988)
- 4) 大森, 池亀, 菅原 「ソフトウェア開発へのQuality Feedback」, 品質, Vol.19, No.2, pp16-24, (1989)