

つぶつぶビュー：カテゴリデータのグラフィカルな可視化インタフェース

Tutubuvview: a graphical visualization interface for categorical data

白石 宏亮 三末 和男 田中 二郎*

Summary. カテゴリデータを可視化する際、一般的には広い視点からデータの傾向を俯瞰することが多く、個々の細かい情報まで柔軟に参照することは難しい。我々はカテゴリデータをマクロな観点からだけでなく、個々のレベルへの詳細情報まで直感的に参照することが可能なインタフェースを実装した。本インタフェースではデータの要素をひとつひとつグラフィカルに表示する。ひとつひとつを表示することによってデータの個々のレベルにアクセスが可能である。また、グラフィカルに表示することで、アンケートデータでは人の集まり、商品データでは商品の集まりのようにオブジェクトの集まりとしてデータの本質を直感的に理解することができる。これによって、属性数が多いデータにおいてもユーザーは直感的にデータの探索を行うことができる。

1 はじめに

カテゴリデータとはデータ中の変数が量的な値をもたない、すなわちカテゴリで区別されるデータである。身近な例では、アンケートデータなどが挙げられる。これらの一般的な可視化方法はチャート表現を用いることが多い。チャート表現はデータの傾向をつかむのには有用であるが、狭い視点、例えばアンケートデータにおいてある特定の人、グループへの局所的な参照をすることは従来のチャート表現では難しい。

我々はこれらの問題点を解決するために、データの要素ひとつひとつをグラフィカルに表示し、可視化するインタフェースを実装した。アンケートデータでは一人ひとりをグラフィカルに表示する。これによって、データを個々のレベルからひとつひとつ柔軟に参照が可能である。また、グラフィカルに表示することによってデータをオブジェクトの集まりとして直感的に操作することが可能である。

2 つぶつぶ表示

2.1 要素のつぶつぶ表示

カテゴリデータはその性質上、アンケートデータでは人、商品データでは商品のようにデータの要素をオブジェクトとして捉えることができるデータが多い。この特徴を直感的に理解しやすいように、本手法ではデータの要素をひとつひとつグラフィカルに表示する。要素ひとつひとつを表示することによって、データひとつひとつを視覚的に捉えることができる。また、ある要素のみに着目するといった個々

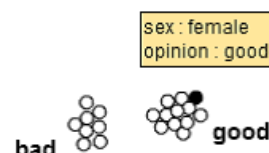


図 1. 要素のつぶつぶ表示

への参照が柔軟に行える。このように要素をひとつひとつ表示することをつぶつぶ表示と名付ける。

ある飲料水に対する評価を「good」と「bad」のどちらかで、被験者男女各 10 人に行ってもらったようなアンケートを考える。このデータを本インタフェースで可視化した例を図 1 に示す。ここでつぶつぶに表示されている丸がデータの要素である。すなわちこのデータではこのつぶつぶひとつひとつを「人」として捉えることができる。

2.2 カテゴリの区別

つぶつぶに表示された各要素は属性を持っている。例えば図 1 では黒色の要素が持つ属性がポップアップして表示されている。この黒色の要素は「female」で、かつ意見として「good」を持っていることを示している。ここで、図 1 では「good」と「bad」の 2 つのカテゴリのラベルが表示されているが、つぶつぶの要素がどちらかのラベルの近くに寄っているのがわかる。表示されたラベルをドラッグすると、そのラベルのカテゴリを持つ要素が周りに近寄ってくる。つまり、ここでは「good」と「bad」の意見を持っている人が配置によって分けられている。このように、表示されたラベルを操作することで、つぶつぶの各要素が移動し、ユーザーは自分の着目したいカテゴリに関してつぶつぶの要素を区別していくことによって可視化を行っていく。

Copyright is held by the author(s).

* Kousuke Shiraiishi, Kazuo Misue and Jiro Tanaka, 筑波大学大学院 システム情報工学研究科 コンピュータサイエンス専攻

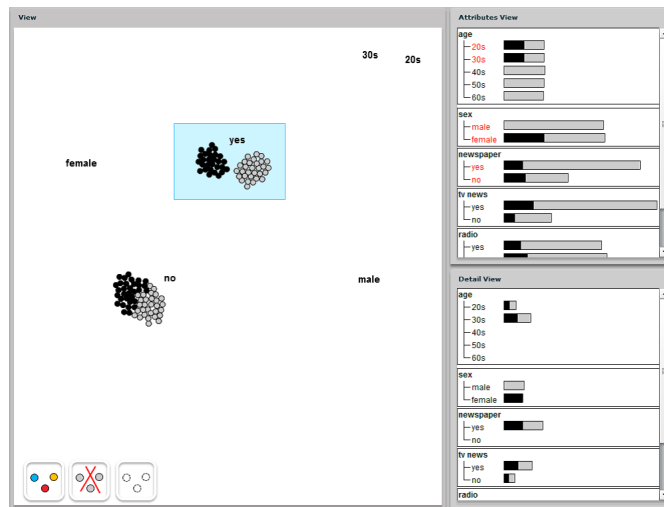


図 2. インタフェース概観

3 つぶつビュー

3.1 インタフェース構成

実装したインタフェース概観を図 2 に示す．本インタフェースはメインパネル，属性パネル，詳細パネルの 3 つのパネルから構成されている．

メインパネルはつぶつぶの要素やカテゴリのラベルが表示される画面で，ユーザーは主にこの画面で操作をし，可視化を行っていく．図 2 では，あるアンケートデータの可視化を示している．ここでは黒色のつぶつぶが「female」，白色が「male」を表している．そして，「yes」，「no」の 2 つのカテゴリは新聞を読むか否かを示している．それぞれつぶつぶはいずれかのラベルの近くに寄っていて，それぞれの要素が持つ属性が配置によってわかる．

属性パネルは全体の属性ごとのカテゴリの集計値が棒グラフ表示され，詳細パネルはメインパネルで選択したある特定つぶつぶのみにに関する情報が属性パネルと同じ形式で表示される．つまり，図 2 では新聞を読む人だけに関する情報が表示されている．

3.2 可視化の流れ

本インタフェースによる可視化の流れを述べる．まずデータの傾向をマクロな視点から概観する．これは属性パネルに表示された属性ごとの棒グラフにより行う．そして，より詳細に着目したい属性，カテゴリを選択すると，そのカテゴリに属する要素がつぶつぶ，カテゴリのラベルがメインパネルに表示される．すなわち，メインパネルでは全体のデータの中の一部をひとつひとつの要素の視点から参照していく．カテゴリのラベルを操作することによってつぶつぶは移動し，ユーザーは着目したい部分のカテゴリ分けを行う．そして，カテゴリ分けを行った要素を範囲指定することによって，指定された要

素のみの詳細情報が詳細パネルに表示される．このように，属性パネルで全体を概観しながら，メインパネルでは全体の中の一部に焦点を当て操作することによって，要素ひとつひとつへのミクロな視点での参照を行っていく．

4 関連研究

Mosaic display[1] ではクロス集計表中のセルの値を長方形の大きさとして表現する手法を提案している．また，[2] では Mosaic display と parallel coordinate を発展させた表現を提案している．上記の研究は本研究のように個々のレベルに柔軟にアクセスできない問題点がある．

Dust & Magnet[3] では多変量解析のアプローチとして要素をひとつひとつドットで表示して，マグネットのメタファーを利用して数ある商品の中から自分の欲しいものを探す支援をしている．要素ひとつひとつをドットで表すという点は似ているが，本研究の対象は量的データではなく，複数の要素から特定の要素を探し出すという目的も異なっている．

参考文献

- [1] M. Friendly, Visualizing categorical data: Data, stories and pictures, In *Proceedings of SAS User Group International Conference*, pp.190-200, 1992.
- [2] F. Bendix, R. Kosara, and H. Hauser, Parallel Sets: Visual Analysis of Categorical Data, In *Proceedings of InfoVis2005*, pp.133-140, 2005.
- [3] J. S. Yi, R. Melton, J. T. Stasko, and J. A. Jacko, Dust & Magnet: multivariate information visualization using a magnet metaphor, *Information Visualization*, vol. 4, pp.239-256, 2005.