視覚的要素の配置・色を用いた多次元カテゴリデータの分析ツール

白石　宏亮†　　三末　和男†　　田中　二郎†
†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

1　はじめに

カテゴリデータはビジネスや科学分野など様々な領域で現れ，分析目的に利用することが多い．例えば，代表的なカテゴリデータの一つであるアンケートデータはマーケティングリサーチなどにおいて製品などの市場調査をするために分析する．

本論文では，カテゴリデータの視覚的な分析を目的とした分析ツールの開発について述べる．本ツールで用いる「つぶつぶ表現」はデータにおける個々のエンティティを視覚的要素として表現し，インタラクティブな操作によって多次元データの視覚的な分析が可能である．

2　カテゴリデータ分析

従来の一般的なカテゴリデータ分析のプロセスを図1に示す．まず，カテゴリデータの生データはリスト形式のデータである．行にエンティティ，列に属性が対応する．エンティティとはデータにおいて，一つの単位としてまとめられる対象のことを呼ぶ．例えば，アンケートデータにおける「人」や，商品データにおける「商品」がエンティティと考えられる．生データにおいていくつかの属性を選び，カテゴリのクロス集計 (Cross tabulation) をとることで集計表 (Contingency table) として表現することができる．最終的にこの集計表を棒グラフや円グラフなどのグラフ表現により可視化を行い，人間が視覚的に分析しやすい形に表現する．

多次元データはデータ中の属性をいくつか選び，次元を切り替える操作が必要である．しかし，このような表をベースとした方法では専門的な知識が必要となることがあり，時間と労力がかかってしまう．

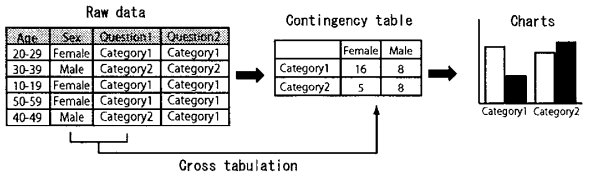

図1: カテゴリデータ分析のプロセス

A Tool for categorical data analysis using color and position of visual elements
†Kousuke Shiraishi, Kazuo Misue and Jiro Tanaka
†Department of Computer Science, University of Tsukuba

3　カテゴリデータの視覚的表現

3.1　つぶつぶ表現

ここでは，本ツールで用いるカテゴリデータの視覚的表現について述べる．本研究で開発した視覚的表現は，データにおける個々のエンティティを視覚的要素として表現する．我々はこの視覚的要素を大きさを持った円として表現する．円として表現する理由は，図形における最も基本的な形状の一つであるということと，散布図などにおけるデータのプロットは点として表現されるため，個々のエンティティとしての理解が容易であると考えた．この視覚的表現を円として粒のように表現する特徴から「つぶつぶ表現」と名付けた．つぶつぶ表現における個々の円を要素と呼ぶ．図2に本研究で開発した視覚的表現と，比較として棒グラフ表現を並べた図を示す．

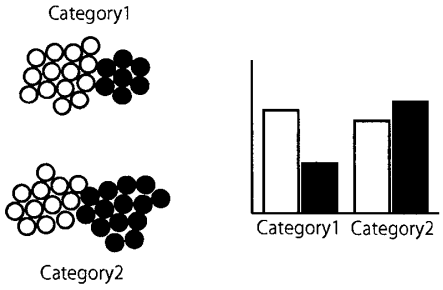

図2: つぶつぶ表現 (左) と棒グラフ表現 (右)

棒グラフ表現は棒の長さによって何らかの値の相対量を表現するのに対して，つぶつぶ表現は個々のエンティティを視覚的に表示することで絶対量を表現する．そして，要素の配置・色によって属性におけるカテゴリを表現する．

4　ツールの開発

我々は [1] において分析ツールの第一プロトタイプの開発を行った．本稿では，[1] で開発したツールの機能などを吟味し，再実装を行ったツールについて述べる．新しい機能として，「要素の非アクティブ化」「任意のラベル作成」を開発した．

Copyright © 2010 Information Processing Society of Japan.
All Rights Reserved.

4.1 ラベルによる要素の操作

多次元データ分析には複数ある次元を切り替える操作が必要である．つぶつぶ表現ではこの操作をクエリを模した操作で行う．ユーザは着目したいカテゴリのラベルを操作することで、要素の配置を移動させる．ラベルをマウスによって動かすと，要素がラベルの方向にアニメーションによって移動する．図3はまず要素を男性と女性の要素に分け，10代のラベルを用いて男性と女性の要素から10代の要素を移動させた例である．

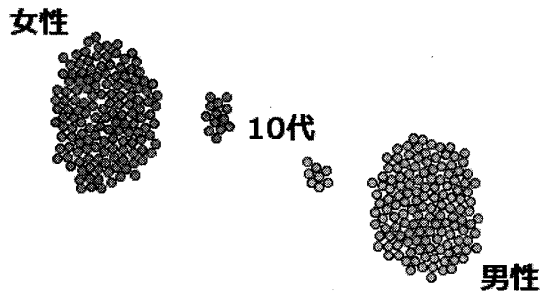

図3: ラベルによる要素の操作

4.2 要素の非アクティブ化

分析する過程において，データの一部に着目したい場合がある．例えば，「10代の男性」のように一部のカテゴリのみの要素に着目して詳細に分析したい場合などである．

要素の非アクティブ化は，ラベルによる操作の影響を受けなくする機能である．図4では要素を男性と女性に分け，男性を非アクティブ化している．非アクティブ化をすると，要素の色が白くなり，枠が点線となる．この状態で10代のラベルで引き寄せると，女性の要素からのみ要素が引き寄せられる．

非アクティブ化によって，複数のカテゴリのAND関係で要素を分けることができ，より詳細な分析が可能となる．

4.3 任意のラベル作成

分析する過程において，データ上では定義されていないカテゴリをユーザ自身で作成したい場合がある．例えば，年代の属性を含むアンケートデータでは10代・20代・30代…のように10歳ごとにカテゴリ分けされたデータとなっているのが一般的であるが，若い年代とそれ以上の年代との2つに分けたい場合などがある．

任意のラベル作成はユーザ自身で定義したカテゴリを作成することができる．全ての要素を選択してラベルを作成することで，全ての要素を引きつけるラベルを作成することや，ユーザが着目していない要素を選択して，「その他」のようなラベルを作成する使い方が可能である．

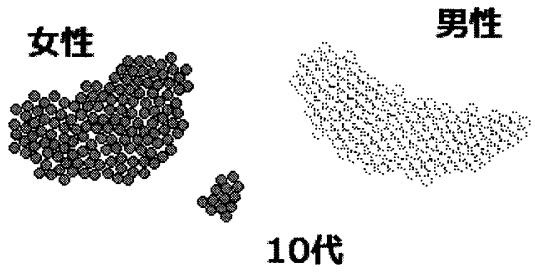

図4: 要素の非アクティブ化

5 関連研究

Dust&Magnet[2]では多次元の数値データを磁石のメタファを用いて分析する手法を提案している．データにおけるエンティティをダストのように表現し，属性を磁石とみたてて操作をすると，ダストが持つ各属性の値によって引き寄せられる早さが変化する．これにより，軸のない多次元の散布図のように，ダストを2次元上に布置することができる．本研究はカテゴリデータを対象としてる点や，カテゴリデータ分析に適した機能を備えている点で異なっている．

6 まとめ

本稿ではカテゴリデータ分析を目的とした視覚的表現及び，ツールの開発について述べた．視覚的表現はデータにおけるエンティティを視覚的要素として表現し，それらの配置・色を用いて属性におけるカテゴリを表現する．開発した分析ツールはクエリを模した操作によって，多次元データにおける次元の切り替えをインタラクティブに行うことが可能である．

参考文献

[1] 白石宏亮，三末和男，田中二郎．つぶつぶ表現を用いたカテゴリデータの視覚的分析ツール．インタラクション2009, pp. 105–112. 情報処理学会, 2009.

[2] Ji Soo Yi, Rachel Melton Ponder, John Stasko, and Julie Jacko. Dust magnet: multivariate information visualization using a magnet metaphor. In *Information Visualization*, Vol. 4, pp. 239–256, 2005.

Copyright © 2010 Information Processing Society of Japan.
All Rights Reserved.