

# 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

白石 宏亮<sup>†</sup> 三末 和男<sup>†</sup> 田中 二郎<sup>†</sup>

本論文では、カテゴリデータを分析する手法としてデータのレコードを一つ一つ視覚的に表示し、それらの配置、色、大きさによってカテゴリの集合的側面を表現する手法の開発について述べる。この手法をデータのレコード、すなわち要素一つ一つを表示することからつぶつぶ表現と名付けた。つぶつぶ表現によって、データをオブジェクトの集合として直感的にイメージしやすくなり、インタラクティブにカテゴリデータの多次元分析、局所的な特徴の分析が可能である。また、要素一つ一つを参照することができるため、従来方法では困難であった個々の観点からの分析が柔軟に行える。我々はこのつぶつぶ表現とグラフ表現を統合することによって、質的と量的の両側面からカテゴリデータの分析が可能なツールの実装を行った。開発したツールを用いて具体的なデータの可視化を行い、本手法の有用性を示した。

## A Tool for Analyzing Categorical Data Visually with Granular Representation

KOUSUKE SHIRAIISHI,<sup>†</sup> KAZUO MISUE<sup>†</sup> and JIRO TANAKA<sup>†</sup>

In this paper, we propose "Granular Representation", a new visualization method for analyzing categorical data visually. Given a raw categorical data, Granular Representation displays each data record as a visual element one by one and represents differences of categories by arrangement, color, and size of elements. By displaying as visual elements, categorical data can be imagined as a set of objects intuitively and analyzed from an individual point of view. We developed a tool for multi dimensional analysis of categorical data by integrating Granular Representation and bar charts. The effectiveness of the tool is demonstrated using a real data.

### 1. はじめに

カテゴリデータ（質的データ）とはデータ中の変数がカテゴリによって区別されるデータであり、温度や売上など、データ中の変数が数値をとる量的データの対として区別される。カテゴリによって区別される変数の例として、アンケートデータにおける性別や血液型、出身地などが挙げられる。このカテゴリデータは、新聞やテレビなどのマスメディアやマーケティングリサーチといった日常生活の多くの場面で使用されており、その使用目的によって大きく2つに分けることができる<sup>5)</sup>。一つはマスメディアで使用されているようなデータの傾向を伝えることを目的しているものである。世論調査では反対、賛成など、評価では良い、悪いといったような意見の比率がどのような傾向にある

かを相手に掲示することが重要である。しかし、マーケティングリサーチなど、データを詳細に分析する場合、反対や賛成といった比率に加えて、反対と言っている人達の性別、年齢、職業はどんな傾向にあるか？といったような、複数属性間の内訳、すなわち多次元分析を行う必要がある。しかし、このようにカテゴリデータを細部に渡って分析することは、専門的な知識を必要とする場合が多く、直感的に行うことは困難である。また、一般的なカテゴリデータの分析の手順は、大規模なテーブルから探索を行い、複数のグラフを見比べるといった作業が必要であるため、非常に時間と労力がかかる。

我々はカテゴリデータを分析目的に使用するために、視覚的に直感的な多次元分析、局所的な特徴の分析が可能ツールを開発している。本ツールに用いる手法、「つぶつぶ表現」はデータのレコード、すなわちアンケートデータでは一人一人の人といった要素を視覚的に表示する。そして、配置、色、大きさを用いてカテゴリの集合的側面を表現する。それらをインタラクティ

<sup>†</sup> 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  
University of Tsukuba, Department of Computer Science

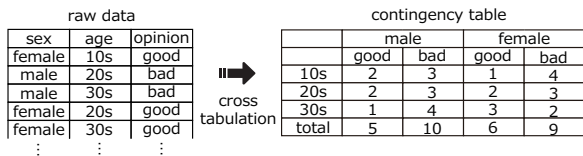


図 1 クロス集計  
Fig. 1 Cross tabulation

に操作することで、カテゴリデータの着目したい部分へと直感的にドリルダウンすることが可能である。要素を一つ一つ視覚的に表示するという特徴から、要素の数で量的側面の分析を、個々の要素を参照することで質的側面からの分析が可能である。

## 2. カテゴリデータの可視化

### 2.1 カテゴリデータ

カテゴリデータ (categorical data) とはデータ中の変数が名義的 (nominal)、または順序的 (ordinal) な尺度で測られる変数である。名義変数とは性別、血液型など名詞的な値であり、カテゴリによって区別される<sup>2)</sup>。例えば、性別では男性と女性の 2 つのカテゴリに分けられる。順序変数とはカテゴリ間の順序付けが可能変数である。例えば、階級や映画の年齢指定などが挙げられる。

データ中の変数が数値をとるものは量的データであるが、カテゴリ化を行うことでカテゴリデータとして扱うことができる。例えば、年齢を 10 歳ごとに一つのカテゴリとして分けていくといった方法である。

### 2.2 クロス集計

データが収集され未加工の段階、すなわち生データはレコードがそれぞれの属性のカテゴリを持つデータである。例えば、ある商品を「良い」または「悪い」のどちらかで評価してもらった簡単なアンケートを考える。この時、属性として性別と年齢を考え、性別のカテゴリとして「男性」、「女性」、年齢のカテゴリとして「10代」、「20代」、「30代」の 3 つをとる。商品に対する評価も一つの属性として考えられ「良い」と「悪い」の 2 つのカテゴリとなる。このようなカテゴリデータの生データは各レコードが 3 つのデータフィールド、すなわち属性を持つデータで図 1 左のようなデータとなる。

この生データを各属性ごとに集計することによって図 1 右のようなクロス集計表によって表すことができる。これは 3 つの属性間の度数を全て表した 3 次元のクロス集計表である。すなわちこの表から 3 つの属性間の関係が得られる。このクロス集計表の性質として、属性数を  $n$ 、属性のカテゴリ集合を  $C_i (i = 1, \dots, n)$  と

すると、 $n$  次元クロス集計表中の総セル数は  $\prod_{i=1}^n |C_i|$  となる。上式からわかるように、データ中の属性数が増えると、とりうる集計の数は多くなる。

### 2.3 クロス集計表の可視化

クロス集計表中の各セルの値は数値をとるから、量的データの可視化手法を適用できる。クロス集計表を可視化する方法としては、棒グラフ、円グラフなどのグラフ表現が幅広く用いられている (なお、これ以降「グラフ表現」という言葉を何度か用いるが、本稿で用いる「グラフ表現」とはグラフ理論やグラフ描画ではなく、棒グラフや円グラフなどの視覚的な図表現のことを指す)。属性数の多いカテゴリデータを分析する際は、とりうる集計の数が多くなるため、表全体を一度に可視化するという事は視覚的に煩雑になってしまう。そのため、複数の属性の中から着目したい項目だけのクロス集計表を作成し、グラフ化を行う。そして、得られた知見をもとに別の属性のクロス集計表を作成し、グラフ化を行うといったような作業を繰り返す。こうした作業の中で、複数のグラフを見比べることや、クロス集計表とグラフとの対応を把握することは難しい。

### 2.4 個々への参照問題

カテゴリデータにはコメントなどのようにカテゴリによって区別することが困難な値が含まれる場合がある。例えば 1,000 人のアンケート中のコメントを全て順に閲覧していくことを考える。順に閲覧していくことで、データ全体としてどんなことがコメントとして言われているかという傾向はつかめる。しかし、分析していく上で重要なのは、どのような人達がどのようなコメントを言っているか?といったように、コメントと属性間との関連である。このようにデータ全体の傾向を概観しつつ、局所的な特徴を同時に分析することは従来方法では困難である。

## 3. つぶつぶ表現

Shneiderman によって提唱されたマントラ<sup>12)</sup>によると、可視化におけるプロセスとはまず全体を概観し、ズームやフィルターを行い、さらに必要に応じて詳細に分析するとされている。カテゴリデータの分析においてもまず全体を概観し、得られた知見を元に次元から多次元へ関係へとドリルダウンしていくプロセスが重要である。

属性数  $n$  のカテゴリデータでは最大で  $n$  次元の関係が得られるが、その全てに意味があるわけではない。分析を行っていく上で、関係のありそうな属性同士を選び、特徴や傾向など、得られた知見を元にドリルダ



図 2 つぶつぶ表現  
Fig. 2 Granular representation



図 3 配置による関係付け  
Fig. 3 Categories are represented by arrangement

ウンしていくようなプロセスを柔軟に行える必要がある。

### 3.1 個々の視覚的表示

カテゴリデータはアンケートデータでは人のように、一つ一つのオブジェクトがそれぞれ属性のカテゴリを持った集合と考えられる。この性質を直感的に想像しやすいように、データのレコードを一つ一つ視覚的に表示する。図 2 に図 1 で例に挙げたデータを本手法で可視化した例を示す。図中のつぶ一つ一つがデータのレコードに対応している。また、ここでは黒色のつぶの持つ属性が右側の枠内に表示されている。つまり、この人は 30 代の男性であり、「悪い」という評価をしていることがわかる。このようにデータのレコード一つ一つを視覚的に表示することを「つぶつぶ表現」と呼ぶ。さらに、このつぶつぶ表現におけるつぶを要素と呼ぶ。

図 2 は要素に何の関係付けもされていない状態である。すなわち、一つ一つ参照しなければ、要素の持つ属性のカテゴリがわからない。つぶつぶ表現ではカテゴリの集合的側面を表現するために、配置、色、大きさを用いる。

### 3.2 配置による関係付け

要素の集合が配置によっていくつかの塊に分かれている場合、人間はそれぞれの塊を別の関係として知覚する。この知覚はゲシュタルト要因の近接の要因によるものである<sup>14)</sup>。この性質を利用し、配置によって要素が持つ属性のカテゴリを表現する。実際に図 1 のクロス集計表のデータをつぶつぶ表現を用いて表し、配置によって属性のカテゴリを表現した例を図 3 に示す。ここで、図中のラベルは属性のカテゴリを表している。

近接の要因が働き、要素群は 2 つの塊に分かれていると知覚する。また、ラベルと要素群の配置によ



図 4 色による関係付け  
Fig. 4 Categories are represented by color



図 5 大きさによる関係付け  
Fig. 5 Grouping elements by size

て、それぞれの要素が持つ属性を認識することができる。この場合、男性と女性の要素群に分かれているとわかる。

### 3.3 色による関係付け

配置と同様に、色を用いて要素の持つ属性のカテゴリを表現する。ある要素群の中で、色が異なる要素を別々の関係、同色である要素同士を同一の関係と知覚する。これはゲシュタルト要因の類同の要因によるものである。例えば、性別の属性を持つ要素群の中で、男性と女性を図 4 のように別々の色で表現すると、要素の関係を一目で知覚することが可能である。要素群の両端に同色のラベルを配置することによって、要素の持つ属性が認識できる。色は配置とは独立して利用できるため、色と配置によって 2 次元の関係を表現することが可能である。

### 3.4 大きさによる関係付け

要素の大きさは一つの要素中に含まれる要素の数に比例する。すなわち、複数のレコードを一つの要素として表現する。これを要素のグループ化と呼ぶ。図 1 では一つの要素が一つのレコードに対応していたが、グループ化では図 5 のように一つの要素に複数のレコードが含まれる。要素の大きさはその要素に含まれるレコード数に対応し、図 5 右では一つの要素に 10 レコードが含まれている。すなわち、図 2 では 30 個の要素として表現されていたが、グループ化では 3 つの大きい要素として表現される。レコード数が多いデータでも、グループ化を行うことで視覚的に煩雑にならずにつぶつぶ表現を用いることができる。

## 4. インタラクション手法

前章では要素の関係付けについて述べた。ここでは配置分けの方法やグループ化における要素の分離・結合といったインタラクション手法について述べる。直感的なインタラクション手法を用いて要素の操作を行い、

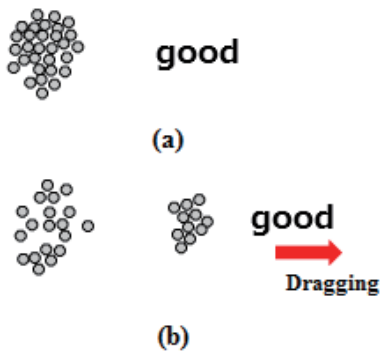


図 6 要素の配置制御

Fig. 6 Controlling elements with the label

要素のカテゴリ分けを行っていく。

#### 4.1 要素の配置制御

要素の配置制御は属性のラベルをドラッグすることによって行う。例えば、図 6 (a) は何の関係付けもされていない状態である。この要素群から「良い」のカテゴリを持つ要素を取り出すには、カテゴリのラベルをドラッグする。すると、ドラッグしているラベルのカテゴリを持つ要素に、ラベルの方向に引き寄せられる力が働き、ラベルの方向に移動する (図 6 (b))。具体的にこの操作はラベルにカーソルを合わせ、プレスしている時にだけ引き寄せられる力が働くので、配置を柔軟に制御することが可能である。

#### 4.2 ラベルの力制御

前項で述べたラベルによる操作は、表示された要素群の全てを引き寄せるため、要素群を 2 次元以上の属性の関係に分けたいといった場合には適さない。例えば、図 1 のアンケートデータを「良い」と「悪い」の要素群に分けた後、さらに「男性」と「女性」の要素群に分ける場合などは、一度「良い」と「悪い」に分けた要素群も、「男性」と「女性」のラベルで引き寄せることでまとまってしまう、2 次元の属性の関係に分けることができない。このような時に、要素群を 2 次元以上の属性の関係に分けていく方法として線による力制御を用いる。まず、「良い」と「悪い」の要素群に分けた後、その間に線を描く。すると、線を跨いではラベルで要素を引き寄せる力が働かなくなる。この時、描いた線上を行き来するようにラベルを動かすことによって、4 つのカテゴリへと要素群に分けることができる (図 7)。この力制御により、要素群を 2 次元以上の属性の関係へと柔軟に分けていくことが可能である。

#### 4.3 要素の分離・結合

大きさの関係付けによるグループ化は一つの要素中

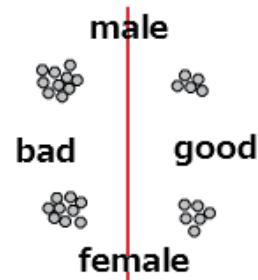


図 7 ラベルの力制御

Fig. 7 Controlling attractive force of labels by drawing the line

に複数のレコード数が含まれるため、つぶつぶ表現の利点である個々への参照が行えなくなってしまう。グループ化された要素は分離をすることで、これによって個々への参照が可能である。

また、グループ化におけるラベルによる操作では、例えば一つの要素の中に「男性」と「女性」のカテゴリが含まれている場合に「男性」のラベルを用いて引き寄せた場合は、一つの要素から「男性」の要素、「女性」の要素の 2 つの要素に分離し、「男性」の要素のみが引き寄せられる。この際に、別々に分離した要素を結合することが可能である。このように、要素を柔軟に分離・結合することで、グループ化においても個々への参照が可能である。

### 5. ツールの開発

つぶつぶ表現は要素のカテゴリ分けによるドリルダウンが直観的に行え、個々への参照が行えるというのが利点である。しかし、一つ一つ視覚的に表示するという性質上、量的比較にはグラフ表現の方が優れていると思われる。一方、グラフ表現は量的比較には優れているが、柔軟なドリルダウンや個々への参照は行えない。本ツールはこれらの二つの表現を統合することによってそれぞれの表現の利点を生かし、相互に補完するツールを開発する。

#### 5.1 Linking & Brushing

つぶつぶ表現とグラフ表現の 2 つの異なる表現間の統合を行うために、Linking & Brushing の手法をツールに適用する。これによってそれぞれの表現を互いにリンクさせることが可能である。

Linking とは異なる可視化表現間で共通する部分を統一する手法である<sup>3)</sup>。本ツールではつぶつぶ表現における要素の色と、グラフ表現におけるグラフの色を対応させることで Linking を実現する。Brushing とは全体から着目したい部分を指定することによって、



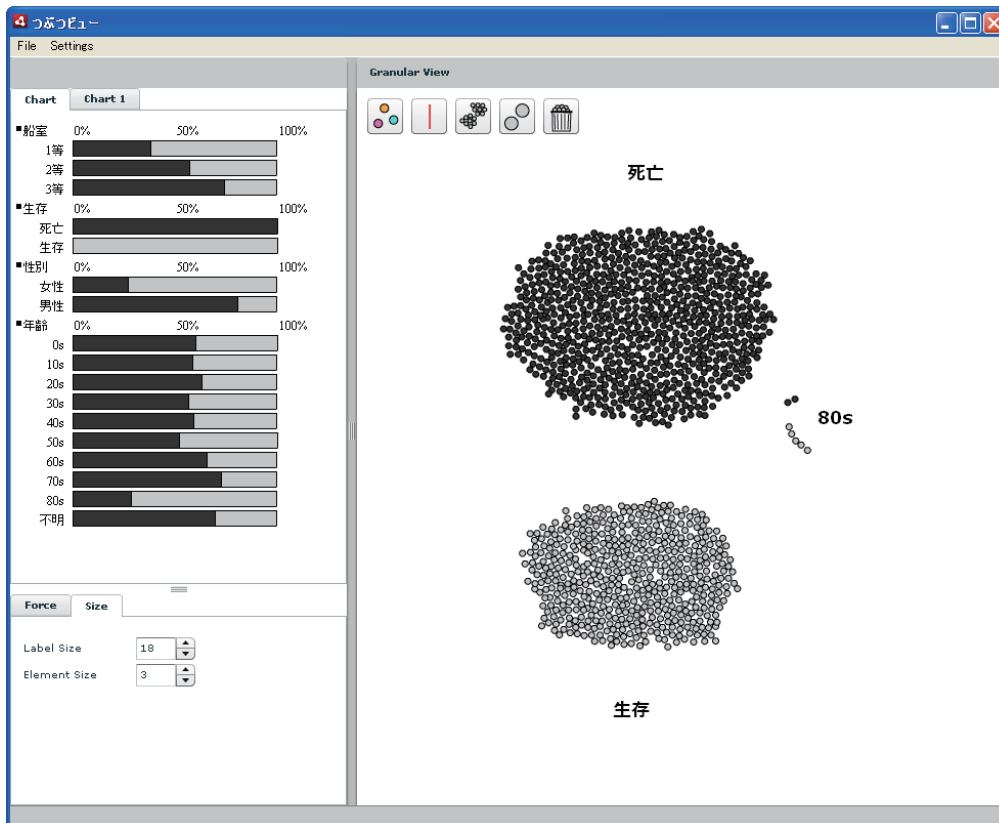


図 8 ツール概観

Fig. 8 An screenshot of the tool

何らかの方法で指定された部分をハイライトして表現する手法である<sup>10)</sup>。つぶつぶ表現において、着目したい要素のみを選択し、それらの持つ属性の値をグラフ表示することで Brushing を実現する。

### 5.2 レイアウト方法

要素やラベルのレイアウトには仮想的な力学モデルを用いている。要素同士には互いに反発しあう斥力を計算し、重ならないように描画している。また、ラベルで要素を引き寄せめる力はバネに類似した力を用いている。ラベルと要素の距離が遠い時には強い力を、また距離が短くなれば弱い力になっていく。また、線によるラベルの力制御は、ラベルと要素を結ぶ線分と力制御の線が交わると、ラベルの引力が効かなくなるように制御している。上記のレイアウト手法により、つぶつぶ表現における要素を柔軟に操作していくことが可能である。

### 5.3 実装

実装は ActionScript 3.0 を用いて AIR アプリケーションとして行った。図 8 にツールの概観を示す。本ツールは右の大きなメインパネルと、左のサブパネルから構成されている。メインパネルはデータがつぶつぶ表現によって表示され、ユーザーは主にこの画面で

作業を行う。左のサブパネルは上下 2 つに分かれており、グラフ表示や要素・ラベルの大きさ、レイアウトの力の強さなどの設定を行う。サブパネルの上下の表示内容はウィンドウメニューの設定から変更でき、タブで重ねることができる。図 8 ではサブパネルの上部にグラフ表示が、下部にサイズと力の設定が割り当てられた状態になっている。グラフ表示は単純集計した値を表示するモードと、メインパネルで選択した要素が持つカテゴリの集計された値を表示する 2 つのモードがあり、サブパネルの上下に割り当てることによって、見比べるということも可能である。

## 6. 利用例

本ツールを用いて実際にデータを分析する例を示す。使用するデータはタイタニック号の乗員乗客のデータである。属性は性別（男性、女性）、船室（1等、2等、3等）、生存（生存、死亡）、年齢（10歳ごとにカテゴリ化）の合計 4 つである。集計する属性としては扱わないが、名前やチケット番号などの個人に関するデータも含まれる。データのレコード数は 1,309 で

ある．ここでは，性別，船室，年齢によって生存率がどのように変わるかを分析することを想定して利用例を示す．

データを読み込むと，サブパネルのグラフ表示に各属性のカテゴリの単純集計された値が棒グラフで表示される．ユーザーはまずこのグラフ表示を見て，全体の傾向を概観する．グラフ表示のグラフ部分をクリックすると，そのグラフのカテゴリに一致する要素がつつぶ表現でメインパネル上に表示される．ここでは，生存と死亡のグラフをクリックする．すると，メインパネル上に生存と死亡の要素が表示される（図8）．次に，メインパネル上に表示された要素を生存と死亡のカテゴリ別に異なる色を付ける．ここでは生存のカテゴリを持つ要素に灰色を，死亡のカテゴリを持つ要素に黒色を割り当てた．図8左のグラフ表示を見ると，船室が上がるにしたがって生存率が増えており，性別では女性の方が男性に比べて生存率が高いことがわかる．さらに年齢では大きな差はないが，80代の乗客は他の年齢に比べて生存率が高いことがわかる．この原因として，80代の乗客は女性や船室の等級が高い人が多いということが考えられる．これを検証するため，80代のラベルによって80代のカテゴリを持つ要素を引き寄せ（この時の画面が図8である）．この状態から，80代の要素のみを選択し，これらの要素のみのグラフ表示を見ることによって，局所的な特徴の分析が行える．ここでは，特に女性や船室の等級が高い人が多いという傾向はないことがわかった．すなわち何か他の原因，もしくは80代のカテゴリを持つ要素は少ないため，偶然に生存率が上がったと考えられる．

次に，性別，船室，生存の3つの属性の関係を分析する．男性のカテゴリを持つ要素には青色を，女性のカテゴリを持つ要素には桃色を付け，生存と死亡のラベルによって引き寄せる．さらに，力制御の線を描き，船室の1等，2等，3等のラベルによって引き寄せ，6つの要素群に分ける（図9）．ここではつつぶ表現の要素の数を比較することで分析を試みる．女性の要素に着目して，船室の等級別に生存率を見てみると，全ての船室において男性に比べて生存率が高いことがわかる．さらに，船室の等級が高いと生存率が増えていることがわかる．すなわち，女性は全体的に救助を優先され，かつ船室の等級が高いとその優先度も上がっていたと考えられる．次に，男性の要素に着目してみる．2等と3等は生存の要素と死亡の要素の割合が同じぐらいであることがわかるが，1等に関しては他と比べて生存率が高いことがわかる．すなわち，男性に

関しては1等船室の人のみが優先されていたと考えられる．

## 7. 関連研究

カテゴリデータの可視化のアプローチは大きく2つに分けられる．一つはデータ中の値をそのまま可視化手法に適用するアプローチである．もう一方は統計手法をデータに適用し，算出された結果を可視化する方法である．

一つ目のアプローチはクロス集計表中のセル値を図形の大きさによって表現するという手法に基づいている．最も基本的な手法は棒グラフや円グラフなどのグラフ表現による可視化である．しかし，グラフ表現ではあくまで一つの図中に表現できる属性の関係に限られてくる．この問題を解決し，多くの属性を表現する手法が研究されている．Mosaic Display<sup>4)</sup>ではクロス集計表中のセルの値を四角形の大きさに対応させ，タイル状に並べることによって表現している．1次元上に複数の属性を表現するため，次元間での関係がわかりにくく，属性数が増えると視覚的に煩雑になる．Fourfold display<sup>6)</sup>ではクロス集計表のセル値を円の大きさに対応させることで表現している．Cattrees<sup>9)</sup>はTreemap<sup>8)</sup>でカテゴリデータを表現したものである．セル中の値をTreemapの一つのノードの大きさに対応させることで表現している．これらの特徴としては，空間効率よく多くの情報を提示できることである．しかし，属性数が多くなると，カテゴリと視覚的表現との間の関係がわかりにくくなるという問題がある．つつぶ表現はこれらの手法と同様に，ラベルと線によって多次元の情報を表現することが可能である．また，これらの手法では不可能な個々への参照が可能なのが特徴である．

二つ目のアプローチは，カテゴリデータに統計手法を適用し，カテゴリ間の類似度や相関を算出し，その結果を可視化する方法である．統計手法としてはカテゴリデータの多次元分析手法として古くから用いられている数量化 III 類<sup>16)</sup>や対応分析を利用したものが挙げられる．Rosarioらはカテゴリデータを量的データの可視化手法に適用するために，カテゴリ間の関係性を数量化し，量的データへと変換する手法を提案している<sup>11)</sup>．Johanssonらはカテゴリデータと量的データの混合データを量的データ可視化手法で表現するために，数量化の手法を提案し，parallel coordinatesでの可視化例を示している<sup>7)</sup>．parallel coordinatesはそもそも量的データを対象とした表現であるため，カテゴリデータに適用すると理解しづらい問題がある．また，

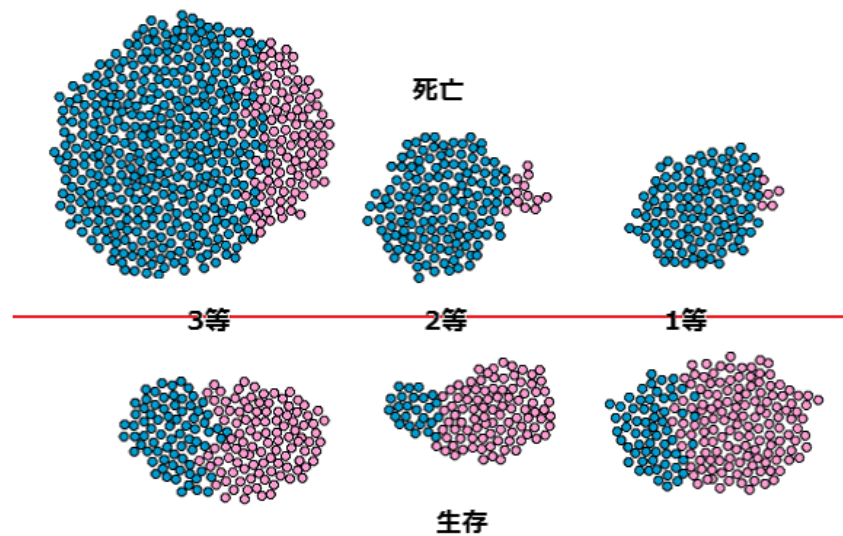


図9 タイタニック号データの可視化例  
Fig.9 An example of visualizing a titanic data

上記2つのアプローチを複合した表現もいくつか挙げられる。Cobweb diagram<sup>13)</sup>ではグラフ理論の網図表現のようにカテゴリ間の関係をエッジで結び、エッジの太さでセル値を表現している。Parallel Sets<sup>1)</sup>ではMosaic Displayとparallel coordinatesを複合した表現を提案している。これらの手法は表現方法が特殊であるため、直感的に理解しにくいという問題が挙げられる。

また、量的データの多変量解析の視覚的手法としてDust&Magnet<sup>15)</sup>(以下DnM)では、商品などの要素を一つ一つドットで表示し、要素の持つ属性の値の比較を直感的に行う手法を開発している。DnMでは磁石のメタファーを利用し、各要素の属性の値を磁石で要素を引き寄せる引力の強さに対応させる。すると、複数の要素を磁石で引っ張ることで配置が分かれ、要素の配置によってそれぞれの要素が持つ属性の値の相対的な特徴が分かる。また、似たような属性の値を持つ要素同士は近くに配置され、カテゴリ化される。つぶつぶ表現はDnMのように、データのレコードを一つ一つ視覚的に表示するというアプローチは同様であるが、DnMが要素間における属性の値の量的側面を可視化しているのに対して、つぶつぶ表現では要素を集合として見たときの、カテゴリの集会的側面を表現している。また、ツールの目的の違いから、インタフェースや機能面においても異なっており、グラフ表示、線による力制御やグループ化などはカテゴリデータの集会的側面を分析するためのつぶつぶ表現特有の機能である。

## 8. 考 察

### 8.1 つぶつぶ表現の利点

つぶつぶ表現の要素を一つ一つ表示することの利点について考える。タイタニック号データ利用例の後半(図9)では、要素の数から分析を行うことを試みた。このように、つぶつぶ表現では要素を一つ一つ表示することにより、絶対量を表現できるのが1つの利点であると考えている。従来のグラフ表現は量の割合の比較には適しているが、絶対量が把握しにくい。なぜなら、レコード数がいくつであっても、図形の長さや大きさの差として表現されてしまうためである。つぶつぶ表現ではデータ中の個体が一つのつぶとして独立に表現されるため、絶対量を直感的に把握することが可能である。

2つ目の利点は個々の参照である。要素を一つ一つ参照することで、タイタニック号のデータでは名前やチケット番号など、アンケートデータではコメントのようなユニークなデータと属性間の関係を分析することができる。初めの段階では要素数は多いが、ラベルによって要素をカテゴリに分けていくことによって、このような局所的な特徴の分析が有用になる。

### 8.2 直感性

一般的に、カテゴリデータの分析を行うには専門的な知識が必要なことが多い。また、関連研究においても、表現方法が特殊なものが多く、可視化表現を扱えるようになるまで時間を要することがある。本ツールではデータをオブジェクトの集合として直感的に理解できる。また、ラベルによる要素の配置制御や線によ

る力制御などの直感的な操作によって柔軟にドリルダウンすることが可能である。

## 9. ま と め

本論文ではカテゴリデータの多次元分析，局所的な分析を視覚的に行う手法，つぶつぶ表現を提案した。そして，つぶつぶ表現とグラフ表現の利点，欠点を互いに補い合うツールの設計方針を示し，実装を行った。今回はカテゴリデータを対象としていたが，つぶつぶ表現という概念は量的データにおいても表現可能であると考えられる。

コンピュータの処理速度の向上に伴い，よりインタラクティブなデータの分析手法が必要であると考えられる。その中で，つぶつぶ表現はラベルや線を用いたインタラクシオン手法といったインタラクティブで直感的，視覚的なデータ分析が行える。これは今後のデータ分析手法の発展の手がかりになると思われる。

## 参 考 文 献

- 1) Bendix, F., Kosara, R. and Hauser, H.: Parallel Sets: Visual Analysis of Categorical Data, *Proceedings of the IEEE Symposium on Information Visualization 2005 (INFOVIS'05)*, pp. 133–140 (2005).
- 2) Card, S.K., Mackinlay, J.D. and Shneiderman, B.: *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufmann Pub (1999).
- 3) Eick, S. and Wills, G.: High Interaction Graphics, *European Journal of Operational Research*, Vol.81, No.3, pp.445–459 (1995).
- 4) Friendly, M.: Mosaic displays for multi-way contingency tables, *American Statistical Association*, Vol.89, No.425, pp.190–200 (1994).
- 5) Friendly, M.: *Visualizing Categorical Data*, Sas Inst (2000).
- 6) Friendly, M.: Visualizing categorical data: Data, stories, and pictures, *Proceedings of the Twenty-Fifty Annual SAS Users Group International Conference* (2000).
- 7) Johansson, S., Jern, M. and Johansson, J.: Interactive Quantification of Categorical Variables in Mixed Data Sets, *Proceedings of IEEE International Conference on Information Visualisation (IV'08)*, pp.3–10 (2008).
- 8) Johnson, B. and Shneiderman, B.: Treemaps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures, *Proceedings of IEEE Information Visualization '91*, pp.275–282 (1991).
- 9) Kolatchm, E. and Weinstein, B.: Categorical trees: Dynamic visualization of categorical data using treemaps. ([http://www.cs.umd.edu/class/spring2001/cmsc838b/Project/Kolatch\\_Weinstein/index.html](http://www.cs.umd.edu/class/spring2001/cmsc838b/Project/Kolatch_Weinstein/index.html)) (2001).
- 10) Martin, A. and Ward, M.: High Dimensional Brushing for Interactive Exploration of Multivariate Data, *Proceedings of the 6th conference on Visualization '95*, pp.271–278 (1995).
- 11) Rosario, G.E., Rundensteiner, E.A., Brown, D.C., Ward, M.O. and Huang, S.: Mapping nominal values to numbers for effective visualization, *Proceedings of the IEEE Symposium on Information Visualization 2003 (INFOVIS'03)*, pp.80–95 (2003).
- 12) Shneiderman, B. and Plaisant, C.: *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, Addison Wesley (2004).
- 13) Upton, G. J.G.: Cobweb diagrams for multi-way contingency tables, *Journal of the Royal Statistical Society*, Vol. 49, No. 1, pp. 79–85 (2000).
- 14) Ware, C.: *Information Visualization: Perception for Design*, Morgan Kaufmann Pub (2004).
- 15) Yi, J.S., Ponder, R.M., Stasko, J. and Jacko, J.: Dust & Magnet: multivariate information visualization using a magnet metaphor, *Information Visualization*, Vol. 4, pp. 239–256 (2005).
- 16) 林知己夫：データ解析法の基本，放送大学教育振興会 (1988).