

表とパラレルコーディネートを組み合わせた 視覚的な分析ツール

結 城 崇^{†1} 三 末 和 男^{†1} 田 中 二 郎^{†1}

表とパラレルコーディネートを組み合わせた表現を提案するとともに、その表現に対する操作を備えた視覚的なデータ分析ツールを開発した。個々のデータの閲覧に適した表と、多次元データの概観把握に適したパラレルコーディネートを組み合わせることで、データの探索的な分析において、データの概観に基く手がかりを残したまま詳細に分析することが可能になった。購買履歴データの分析例を通して、ツールの機能を紹介するとともにその有用性を示す。

A visual data analysis tool combined table and parallel coordinate

TAKASHI YUKI,^{†1} KAZUO MISUE^{†1} and JIRO TANAKA^{†1}

We propose a representation combining tables and parallel coordinates, and developed a visual data analysis tool comprising this representation and operations for data analysis. Table representation is suitable for reading data in details. On the other hand, parallel coordinate is useful for understanding the overview of multidimensional data. Combining both representation techniques, it is possible to analyze data in details keeping track from the overview during an exploratory data analysis. We explain the functions and the utilities of our tool through an example of a purchase history data analysis.

1. はじめに

データから有益な情報を取得するために視覚的な分析ツールが開発されている。視覚的な

分析では、視覚的に表現された図に対してインタラクションを行い、分析を進めていく。その分析活動において、データの概観と詳細を行き来する場面が多く存在する。例えば、周期性や傾向などの概観に基づく情報を取得した時に、その情報の正しさや要因を確認するために、該当する部分について詳細なデータを調べる。データの概観と詳細の把握を支援するために、既存の分析ツールの多くが、概観及び詳細についてそれぞれに適した表現を並べて提供している。しかし、異なる表現方法の図同士を見比べることは困難である。そのため、概観と詳細を行き来する間に、それまで抱いていた図に対するイメージや知見が失われてしまうという問題がある。これは、分析活動を中断してしまう恐れがあることを示している。

本研究ではこの問題を解決するために、1枚の図上でデータの概観と詳細の把握を可能にすることを目的としている。目的を実現するために、表とパラレルコーディネートを組み合わせた視覚的な分析ツールを開発した。本ツールは、表とパラレルコーディネートの組み合わせを可能にする表現手法とその表現に対する操作を備えている。

第2章では、1枚の図上でデータの概観と詳細の把握を可能にする基本アイデアについて示す。第3章では、表とパラレルコーディネートを組み合わせる手法について示す。第4章では、提案手法に基づくツールと本表現に対する操作を示す。第5章では、本ツールの利用例を示し、ツールの機能を紹介すると共にその有用性を示す。第6章では、関連研究と本研究の位置づけについて示す。第7章では、研究のまとめについて示す。

2. 基本アイデア

視覚的な分析ツールにおいて、詳細なデータの提示に表がよく用いられている。データの概観の提示には、パラレルコーディネートがよく用いられている。1枚の図上で概観と詳細の把握を可能にするために、概観の提示に適したパラレルコーディネートと詳細の提示に適した表を組み合わせることを考えた。

2.1 表

表は、縦横の直線で区切られた領域(セル)にデータを並べることで、データ同士の関係を簡潔に表現する手法である。視覚的な分析ツールでは、多次元データを属性とレコードに基づいた形式で表現する。データの要素をセルとして表現し、縦横方向に隣接するようにセルを配置し、同レコードのセルは横方向に、同属性のセルは縦方向に位置を揃えてデータを表現する。データの要素を表すセルが隣接しかつ位置が揃っているため、ソート順に沿って個々のデータを閲覧及び選択しやすい。セル同士は重なることがなく、常にセルに書かれたテキストを眺めることができる。しかし、表から情報を取得するためには、セル中のテキストを読

^{†1} 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
Department of Computer Science, Graduate school of Systems and Information Engineering,
University of Tsukuba

まなければならぬ。そのため、データ分布を直感的に把握することは難しい。

2.2 パラレルコーディネート

パラレルコーディネート¹⁾は、多次元データの各次元を表す座標軸(以降、軸と呼ぶ)を平行に並べて表現する手法である。パラレルコーディネートでは、多次元データを以下の手順で表現する。まず、縦に伸びる軸上に各次元の持つデータに対応する点を配置し、軸を横方向に平行かつ等間隔に並べる。さらに、同じレコードである点同士を線で繋ぎ、1レコードを1本の線として表現する。平行に軸を並べるパラレルコーディネートでは、各次元のデータ分布を一覧することができる。さらに、隣り合う軸間の線から、2つの次元の相関を把握できる。例えば、ある軸の右隣の軸への線の角度が一定であれば、2つの属性の相関が強いことが分かる。しかし、軸の中でデータの値が近いものは点が重なってしまい、データ分布を正確に読み取れないことがある。

2.3 2表現の比較

表は、個々のデータ要素の閲覧しやすさから、詳細な分析に適している。しかし、データ分布は直感的に読み取りにくく、概観の分析には適していない。パラレルコーディネートは、多次元データの分布を直感的に読み取れることから、概観の分析に適している。しかし、データ要素が隠れてしまうことがあるため、詳細な分析には適していない。このように、表とパラレルコーディネートは、他方の特徴によって苦手な部分を補うことができる。この2表現を組み合わせることで、データ要素を閲覧しやすく、データ分布を直感的に把握できる特徴を併せ持った分析ツールが構築可能だと考えた。

表では、セルによってデータ要素を表し、縦方向の位置及び横方向の位置で関係を表現する。パラレルコーディネートでは、軸上の点によってデータ要素を表し、横方向の位置と点を繋ぐ線で関係を表現する。どちらの表現もデータ要素と要素間の2つの関係を表現している。これは各々の表現が提示する情報が本質的には同じであり、表現の互換性があることを示している。そのため、この2表現を組み合わせる事が可能だと考えた。

3. 提案手法

表とパラレルコーディネートを組み合わせる手法を提案する。データを表現している図中の任意の部分について、表もしくはパラレルコーディネートに自由に表現を変換させる手法である(図1)。本手法では、表とパラレルコーディネートの混ざり合った2表現の中間的な表現が存在する。提案手法は、表とパラレルコーディネートの組み合わせを可能にする表現手法と表現した要素の配置方法によって構成される。

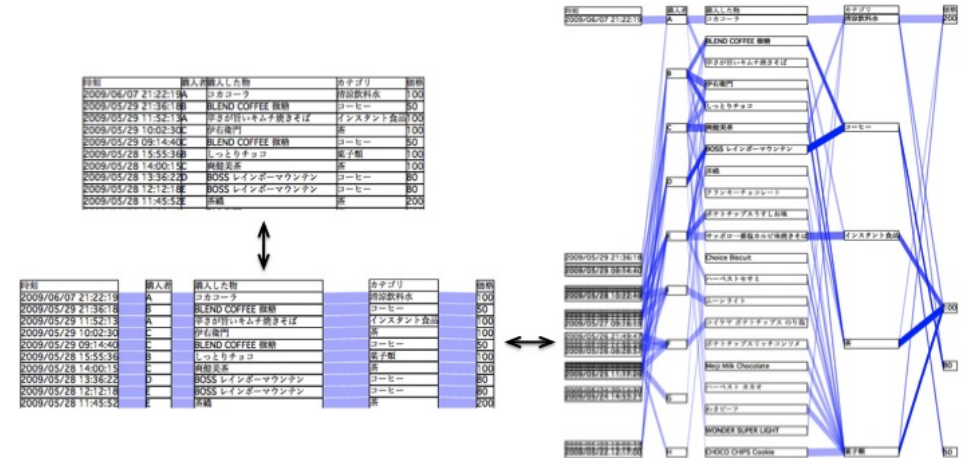


図1 表からパラレルコーディネートへの変換。
Fig.1 Transforming from table to parallel coordinate

3.1 表現方法

データ要素をテキストを埋め込んだ矩形のセルとして表現し、同じレコードに属するセル同士を線で繋いで表現する(図2)。パラレルコーディネートと同様に、横方向の一本の線がレコードに相当する。縦横の位置を揃えた時には表となり、個々のセルをソート順に沿って読みやすい。縦方向の位置に属性内の値をマッピングするとパラレルコーディネートとなり、多次元データの分布を直感的に把握できる(図3)。セル間の線から関連するデータを辿ることが可能である。また、線はセルの縦幅と同じ太さにし、さらに透明度を持たせる。これにより、データが密な所では線が重なって濃い色になり、データの密な部分を把握しやすくなる。

3.2 セルの配置

表とパラレルコーディネートの表現それぞれに対応したセルの配置方法について示す。横方向の配置方法は、同じ属性に所属するセルの位置を揃えて配置する。表形式とパラレルコーディネート形式どちらの表現についても、この方法で配置する。縦方向の配置方法は、表形式とパラレルコーディネート形式で異なる。表形式の配置では、縦方向に隣接するようにセルを配置する。パラレルコーディネート形式の配置では、ある属性についてのデータ分布を表すようにセルを配置する(図3)。パラレルコーディネート形式では、扱うデータの性質によって異なる配置方法を用いる。データが数値情報の場合には、セルの持つデータがセルの縦

| 時刻 | 購入者 | 購入した物 | カテゴリ | 価格 |
|---------------------|-----|-----------------|----------|-----|
| 2009/06/07 21:22:19 | A | ココア | 清涼飲料水 | 100 |
| 2009/05/29 21:36:18 | B | BLEND COFFEE 微精 | コーヒー | 50 |
| 2009/05/29 11:52:13 | A | 新さが旨いキムチ焼きそば | インスタント食品 | 100 |
| 2009/05/29 10:02:30 | C | 伊石衛門 | 茶 | 100 |
| 2009/05/29 09:14:40 | C | BLEND COFFEE 微精 | コーヒー | 50 |
| 2009/05/28 15:55:36 | B | しっとりチョコ | 菓子類 | 100 |
| 2009/05/28 14:00:15 | C | 爽健美茶 | 茶 | 100 |
| 2009/05/28 13:36:22 | D | BOSS レインボーマウンテン | コーヒー | 80 |
| 2009/05/28 12:12:18 | E | BOSS レインボーマウンテン | コーヒー | 80 |
| 2009/05/28 11:45:52 | F | 茶織 | 茶 | 200 |

図2 表とパラレルコーディネートを組み合わせた表現
Fig.2 A representation Combining table and parallel coordinate

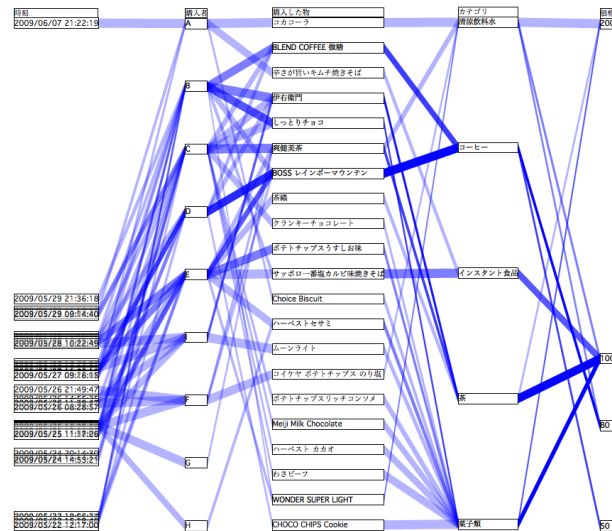


図3 パラレルコーディネート形式の表示。
Fig.3 A representation of parallel coordinate type.

方向の位置に対応するように配置する。属性内のデータの最大値が軸の上端、最小値が軸の下端に位置する。名前などの数値以外の情報の場合は、属性内の同一データを1カ所にまとめ、異なるデータについては等間隔に配置する。

3.3 提案手法の特徴

表とパラレルコーディネートの中間的な表現を提供することで、表現を変換した際に図から発見していた特徴を見失わずに概観と詳細を切り替えられる。例えば、パラレルコーディネートの状態でデータの概観に関する特徴を発見した時に、その部分を表に変換することで、その特徴を示す要因を詳細に調べることができる。また、表を用いて詳細な情報を閲覧した時に、閲覧中の部分をパラレルコーディネートに変換することで、データ全体における位置づけを把握できる。

4. ツールの開発

提案手法に基づいた視覚的な分析ツールを開発した。本ツールは、2表現を組み合わせるための表現方法とその表現に対する操作を備えている。

4.1 本表現に対する操作

本表現手法に対する操作として、データ選択、配置の変更、ハイライト、フィルタリング、グループ化を開発した。視覚的な分析活動は、データの基本的な情報(属性、属性内のデータ形式、属性数、レコード数)を読み取り、次にデータ全体から特徴的なデータを発見し、そのデータについて詳細に調べるといった流れで行われる。本ツールにおいては、まず図全体から特徴的な部分を見出し、データ選択をする。選択範囲について配置方法の変更や次元を減らすなどの詳細に分析するための操作を行う。これらの操作を用いて得られた図から必要な情報を読み解る。

4.1.1 データの選択

操作対象を決定するデータ選択は、分析の起点となる操作である。データ選択法として、矩形選択と遷移選択の2つをサポートする。矩形選択では、マウス操作で矩形を作成し、矩形と重なるセルを選択する。この方法では、図から発見した特徴及び軸のソート順に沿った範囲を選択しやすい。遷移選択では、セルへのマウスクリックで、セル単体→セルを含む軸(属性)→セルを含むレコード→表に基づくセルの周辺というように、属性とレコードに基づいた範囲を遷移させる(図4)。レコードや列を選択するような元のデータ(表)に基づく選択を素早く行うことができる。遷移順は、利用頻度の高い順に設計した。

4.1.2 配置の変更

選択した範囲に対して、セルのレイアウトを3.2節であげた表形式及びパラレルコーディネート形式に変換する操作である。配置を変化させる際は、アニメーションを用いて滑らかに位置を変化させる。アニメーションの効果として、変化前に注目していたセルについて、変



図 4 遷移選択の例
Fig. 4 A example of transitional selecting.

化中に目で追うことができ、変化前後の対応関係を把握できる。例えば、ある商品について時刻が密集している部分を詳細に分析しようと考えパラレルコーディネート形式に表現を変化させる時に、アニメーションによって注目していたセルが配置の変化後にどの位置に行くのかを把握できる。この例のように、配置を変化させても注目していた部分を見逃すことなく分析を進められる。

4.1.3 関係するデータの提示

ハイライト及びフィルタリングは、関係するデータを閲覧する操作である。ハイライトではセルの色によって、フィルタリングではセルの位置によって関係するセルを提示する。色による提示は、図に対して付加的な情報を加える時に図のレイアウトを変更せず情報を表現することができる。しかし、色には識別限界があり複数の着目点の提示に向かないこと、また複数の色が重なった時に識別しにくい。そのため、色による提示は一つの着目点の場合のみに用いる。一方で位置による提示は、図の全体像が崩れてしまう懸念はあるが、着目点が複数ある場合にもそれぞれを識別しやすい。そのため、着目点が複数の場合及び着目点を比較する場合は、位置によって関係するデータを提示する。

ハイライト

着目しているデータと関係するデータを色によって閲覧するための操作である(図5)。着目しているセルが青色にハイライトされ、そのセルと関係を持つ線が赤色でハイライトされる。これにより、着目しているセルの属するレコードを把握できる。図5では、青色にハイ

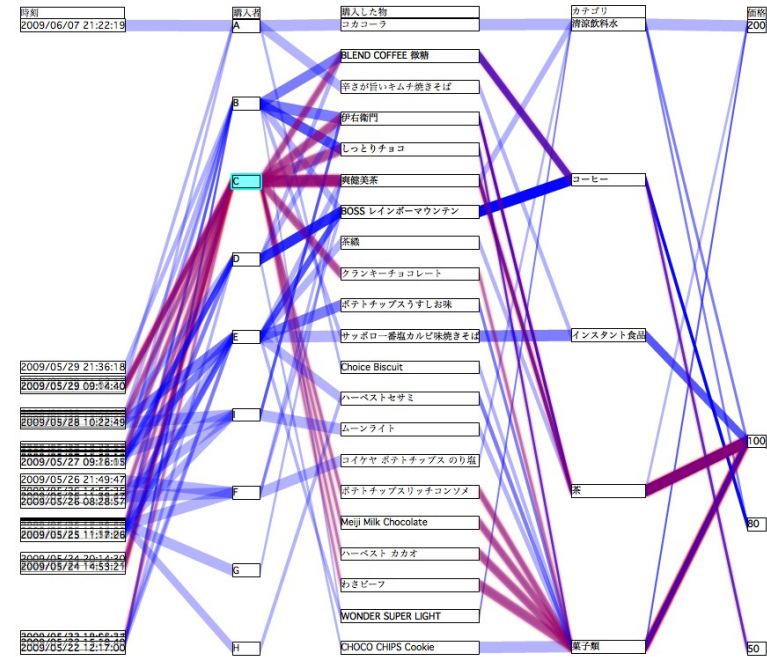


図 5 ハイライトの例
Fig. 5 A example of highlight

ライトされたセルに着目した時に、赤色の線によって全ての軸での分布を把握できる。セルへの色付けにより他の場所に視線を移しても、着目しているセルを見逃さずに分析が行える。また、赤色の線と青色の線の分布の違いを見ることで、データ全体における着目しているデータの位置づけを把握することができる。

フィルタリング

着目したデータについて、位置によって関係するデータを閲覧する操作である(図6)。着目しているセルをマウス操作によって移動させると、それと関係するセルが一定距離に保って位置をずらす。関係するセルが元の軸から抜き出されるように移動し、位置によって関係するセルだと把握できる。図6では、赤色のセルに対してフィルタリングを行い、そのセルと関係するセルが軸から左に離れた位置にずれている。これにより、一番左の軸の左側に赤色のセルと関係するセル、軸の付近には関係しないセルが並び、同時に閲覧できる。また、軸

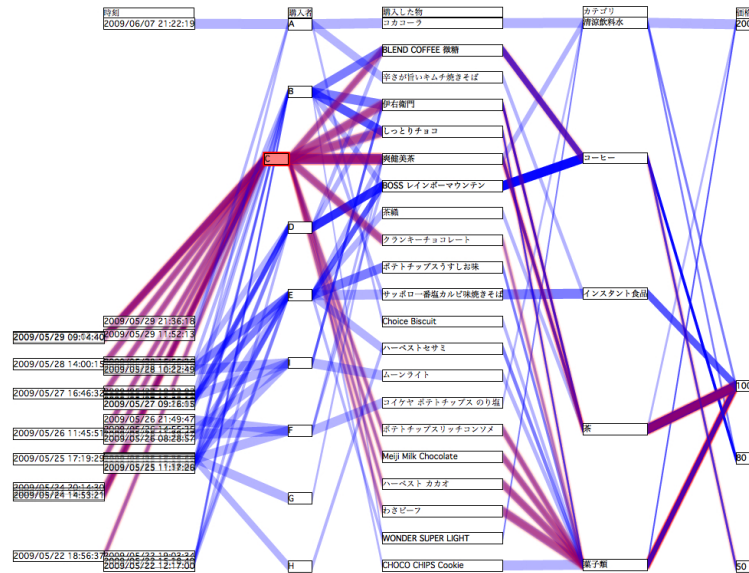


図 6 「時刻」へのフィルタリング
Fig.6 Filtering to "time".

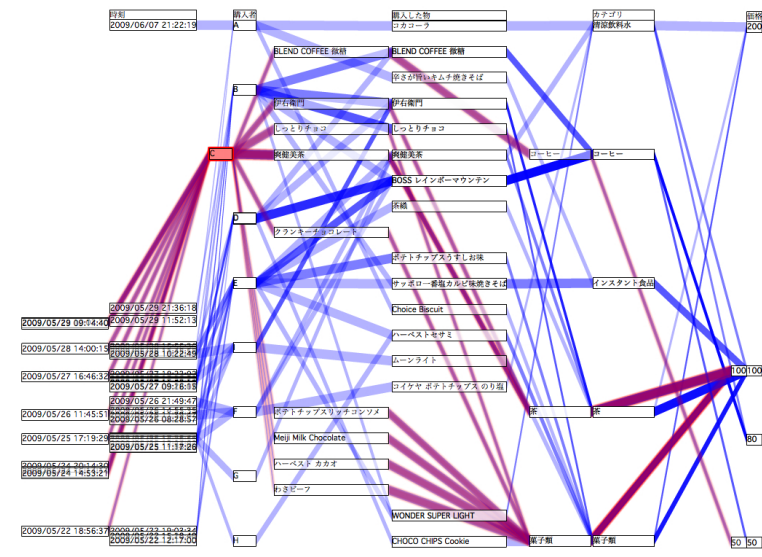


図 7 全属性へのフィルタリング
Fig.7 Filtering to all attribute.

に近い位置であれば、フィルタリング前の概観を残しつつ、分布を眺めることができる。つまり、着目点と関係するデータと関係しないデータと全てのデータを同時に把握できる。位置によって自身の調べたカテゴリを残しておくことで、後から見返した時にも把握がしやすい。複数の着目点についてフィルタリングを行うことで、着目しているデータ同士を比較できる。

フィルタリングの適応範囲として、1つの属性、全属性をサポートしている。1つの属性では、位置をずらすノードを限定することで、図全体の変化を少なくし概観を保ちながら、関係するデータを閲覧できる。データについてある程度特徴を把握でき、調べたい属性が分かっているような場面で利用する。全属性では位置をずらすノードを拡張することで、全属性における分布を把握できる。関係するデータ全てが一度に把握できるが、図の変化が大きく対応関係を追っていく。データの特徴をあまり把握できておらず、特徴的な部分を探す分析の初期段階での利用を想定している。

4.1.4 グループ化

関連の強い2軸を1つの軸のように扱うことで次元数を減らし、他属性との関係を明確に

する操作である。グループ化を行うと、選択した2つの軸が1つのグループとなり、1つの軸として扱われる。グループ化されたセル同士は常に隣り合い、縦方向の位置は一方の軸の位置を取る。図8では、「時刻」と「購入者」の2軸をグループ化することで、それらの属性のセルは隣接し、双方とも「時刻」の分布を表すように配置されている。「購入者」の次元を減らすことで全体の構造が明確になる。

5. ツールの利用例

著者の所属する研究室で運用されている購買システムの購買履歴データに対して本ツールを適応した結果について示す。データは2009年6月5日から2009年6月7日の期間に記録され、レコード数は100件、商品を購入したメンバーは10名、購入された商品数は21個である。購買履歴データは、「時刻」「購入者」「商品名」「商品カテゴリ」「金額」の5つの属性を持っている。「商品カテゴリ」は、商品に対して「菓子」や「茶」のような分類を手でつけた情報である。「購入者」については、個人名を隠すためにアルファベットに変換している。ショップの売り上げを向上させることを目的とし、メンバーや商品の購買

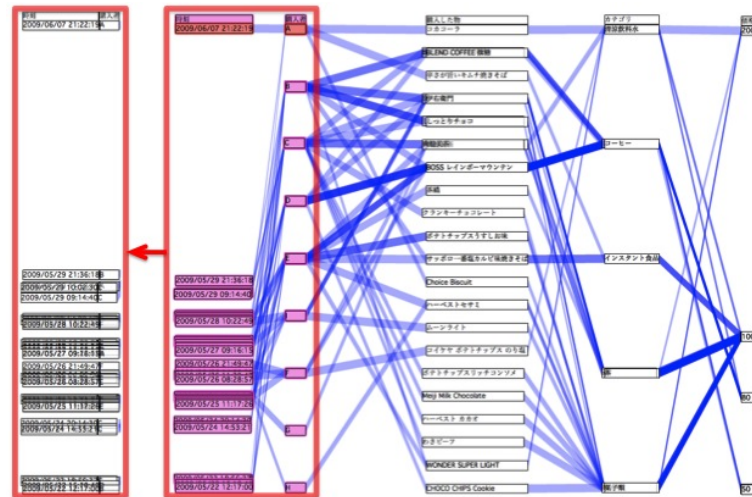


図 8 グループ化の例
 Fig. 8 A example of grouping.

傾向を調査する場面を想定して分析を行う。

まずデータを読み込むとデータが表として提示される。表から、「購入者」や「商品名」などのデータの持つ属性、属性内のデータ形式といったデータの基本的な情報が読み取れる。読み取った情報から「商品名」や「時刻」など気になった属性に関して分析を進める。次に、商品が購入された時刻の分布を見るために、「時刻」の属性をパラレルコーディネート形式に変化させる(図9)。分布を見ると、一定間隔で少し溝があり、これは朝 9:00-10:00 頃に商品が購入され、夜 23:00 頃まで商品が購入されていることを示している。さらに、1 日ごとの購入時間帯について調べるために線を見てみると、溝の間の中心辺りに濃い箇所がある日と濃さがほぼ一定の日があることが分かった。これは、昼頃に購入が集中している日と 1 日を通して平均的に購入された日があることを示す。このように「時刻」の分布から 1 日の周期と商品のよく売れる時間帯を読み取ることができた。分布内に特徴的な部分があれば、その部分についてさらに分析を進めることも可能である。

次に「時刻」以外の属性について分析を行おうと考え、全ての軸をパラレルコーディネート形式に変換する。全体を一望すると、エッジの濃さから「カテゴリ」の軸にある「お茶」と「コーヒー」が特によく購入されていることが分かる。そこでよく売れている商品の傾向を

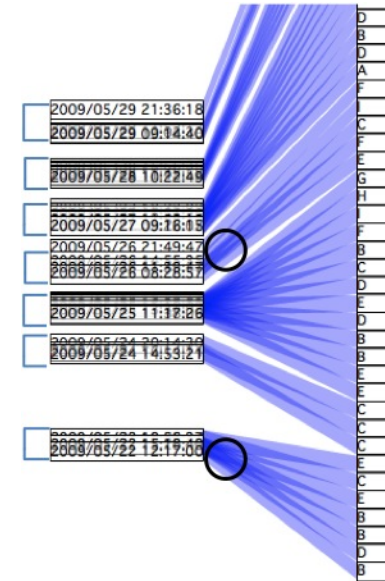


図 9 購買履歴データへの適応例。
 Fig. 9 A example of using this tool for the purchasing history data.

調べるために、購入時刻について比較する。分布を把握するために「お茶」について「時刻」をフィルタリングする。「お茶」を選択し左に移動させると、「お茶」が購入された「時刻」も左に移動し、お茶は毎日数本ずつ購入されているという分布が読み取れる。同様に「コーヒー」を右に移動させると、毎日は購入されないが、集中的によく売れているという分布が読み取れる。このように複数の着目点についてフィルタリングすることで、売れ方の傾向の違いを把握できる。さらに、「お茶」と「コーヒー」以外の軸に残っている時刻の分布と併せて見ると、データ全体の分布を読み解くことができる。これは複数のデータに着目した時も全体の情報を保存したまま、各々の分布を閲覧できることを示している。また、「お茶」「その他」について見てみると、データ全体では毎日多くの商品が売れているように見えていたが、実際は「その他」には 1 つしか商品が購入されていない日があり、「お茶」や「コーヒー」が全データの中で多くの売り上げを占めていることが分かる(図10)。

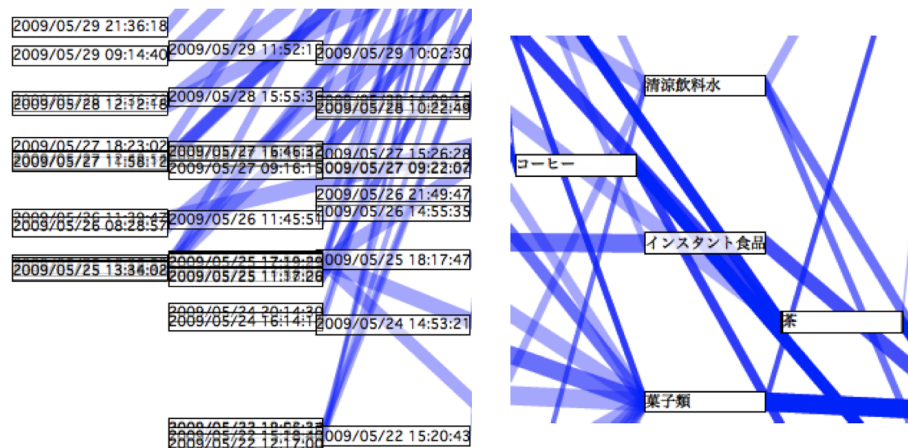


図 10 購買履歴データへの適応例。

Fig. 10 A example of using this tool for the purchasing history data.

6. 関連研究

6.1 視覚的分析ツール

データから有益な情報を取得するために、データを視覚的表現に変換し、その表現に対する操作で分析を行う視覚的分析ツールが開発されている。例えば表計算ソフト Excel *1 では、表を用いてデータ範囲を選択し、選択したデータ範囲について統計グラフに変換し、統計グラフからデータの特徴を視覚的に把握する。視覚的分析ツールでは、データの特徴を直感的に把握することができる。分析の過程においてさらに得たい情報を発見していくような探索的な活動を支援する場面で用いられることが多い。

6.1.1 パラレルコーディネートの拡張表現

パラレルコーディネートを用いた視覚的分析ツールが開発されている。Parallel sets²⁾ は、パラレルコーディネートのある軸について関連の強いデータが近くなるようにソート及び線を色分け、多次元データの概観把握を支援している。また、その中でもパラレルコーディネートと他の表現手法を組み合わせたデータ分析ツールが開発されている。Yuan らは、散布

図とパラレルコーディネートを組み合わせることで、局所的にデータが集中した多次元データの把握を支援している³⁾。Parallel Tag Clouds⁴⁾ は、タグクラウドとパラレルコーディネートを組み合わせることで、イベントの数などの量的なデータを視覚的に把握しやすくしている。本研究では表とパラレルコーディネートを組み合わせており、表の特徴である個々のデータの追やすさを持ち、詳細な分析を行うことに適している。

6.1.2 表の拡張表現

表の持つ表現力の向上を目的として、表現方法の拡張が行われている。TableLens⁵⁾ では、セルの中に棒グラフを埋め込むことで大規模な数値データを直感的に把握できる。Mathias らは、TableLens の属性内の数値情報を透明度によって色分けし、より詳細な情報の把握を可能にした⁶⁾。Telea らは、表とツリーマップを組み合わせることで、数値データとそれに紐づく階層データの把握を可能にした⁷⁾。本研究では、表とパラレルコーディネートを組み合わせることで、その特徴である属性間の相関と詳細な情報を同時に把握できる。

7. まとめ

表とパラレルコーディネートを組み合わせた視覚的分析ツールを開発した。本ツールは、2つの表現を滑らかに変化させるための表現方法と、その表現に対する操作を備えている。これにより、データの概観に基き手がかりを残したまま詳細に分析することが可能になった。購買履歴に対するツールの利用例を示し、本ツールの利用方法と有効性を示した。

参考文献

- 1) Inselberg, Alfred and Dimsdale, Bernard. "Parallel coordinates: a tool for visualizing multi-dimensional geometry," Proceedings of the 1st conference on Visualization '90, pp.361-378.
- 2) Bendix, F.; Kosara, R.; Hauser, H.; , "Parallel sets: visual analysis of categorical data," In Proceedings of the IEEE Symposium on Information Visualization 2005, pp. 133- 140.
- 3) Xiaoru Yuan, Peihong Guo, He Xiao, Hong Zhou, Huamin Qu, "Scattering Points in Parallel Coordinates," In Proceedings of the IEEE Symposium on Information Visualization 2009, pp. 1001-1008.
- 4) Collins, C.; Viegas, F.B.; Wattenberg, M.; , "Parallel Tag Clouds to explore and analyze faceted text corpora," IEEE Symposium on Visual Analytics Science and Technology 2009. , pp.91-98.
- 5) Rao, R. and Card, S. K. The table lens: merging graphical and symbolic repre-

*1 Microsoft Excel <http://office.microsoft.com/>

sentations in an interactive focus + context visualization for tabular information.
In Proceedings of the SIGCHI Conference on Human Factors in Computing 1994.
pp.318-322.

- 6) John, Mathias; Tominski, Christian; Schumann, Heidrun. Visual and analytical extensions for the table lens. Visualization and Data Analysis 2008. pp. 1-12.
- 7) A. Telea, "Combining Extended Table Lens and Treemap Techniques for Visualizing Tabular Data," in Proceedings of Joint Eurographics - IEEE VGTC Symposium on Visualization (EuroVis '06),pp.1-9.