

筑波大学大学院博士課程

システム情報工学研究科修士論文

表と平行ルコーディネートを組み合わせた
視覚的分析ツールの開発

結城 崇

(コンピュータサイエンス専攻)

指導教員 三末 和男

2011年3月

概要

現在、我々が生活を送る中で、様々な出来事がデータとして記録されている。これらのデータを分析することで、生活をより豊かにする有益な情報を取得できる。データの特徴を直感的に把握し、効果的に分析を行うための手段として、視覚的分析ツールが用いられており、開発が盛んに行われている。しかし、現在の視覚的分析ツールは複数のビューを備え、分析中にビュー同士を見比べる場面が多い。見比べる際に片方のビューから目を離さなければならず、図に対するイメージや知見が失われてしまう可能性がある。

この問題を解決する為に、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、表とパラレルコーディネートを組み合わせた表現手法と、その表現に対する操作を備えている。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、データの概観に基づ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。さらに、本ツールの視覚的分析への有用性を評価するために、表とパラレルコーディネートをそれぞれ別のビューとして備える分析ツールとの比較実験を行った。実験結果から、表現を滑らかに変化させることで、より素早く正確にデータ分析を行えることが明らかになった。

目次

第1章 序論	1
1.1 多次元データ	1
1.2 視覚的分析	1
1.3 視覚的分析における問題	2
1.4 本研究の目的	2
1.5 本研究の貢献	2
1.6 論文の構成	3
第2章 多次元データの視覚的分析ツール	4
2.1 従来の視覚的分析ツール	4
2.2 視覚的分析ツールの要件	5
2.3 多次元データ分析ツールで用いられる表現手法	6
2.3.1 表	6
2.3.2 パラレルコーディネート	6
2.3.3 散布図	7
第3章 提案手法	8
3.1 表現手法の設計	9
3.2 表現手法	9
3.3 セルの配置	10
3.4 提案手法の特徴	11
第4章 ツールの開発	13
4.1 本ツールの操作	13
4.1.1 データ選択	14
4.1.2 配置の変更	15
4.1.3 関連データの閲覧	16
ハイライト	17
シフト操作	18
インスタントビュー操作	20
4.1.4 グループ化	22
4.1.5 色付け	22
4.1.6 ビューの調整	23

第 5 章 ツールの利用例	24
5.1 購買履歴データの分析例	24
5.2 考察	28
5.2.1 本ツールでの分析の流れ	29
5.2.2 ツールの利点	29
第 6 章 評価実験	30
6.1 実験の目的	30
6.2 実験の概要	31
6.2.1 評価ツール	31
6.3 実験タスク	32
6.3.1 分析する属性数	32
6.3.2 分析タスクの種類	32
6.4 アンケート	33
6.5 結果と考察	34
6.5.1 正答率	34
6.5.2 タスク達成時間	35
6.5.3 各タスクでのツールの印象	38
6.5.4 ツールへの印象	39
6.6 実験のまとめ	40
第 7 章 関連研究	42
7.1 多次元データの視覚的分析ツール	42
7.1.1 概観に基づく分析ツール	42
7.1.2 詳細と概観に基づく分析ツール	43
7.2 多次元データの表現方法	43
7.2.1 パラレルコーディネートの拡張表現	43
7.2.2 表の拡張表現	44
第 8 章 結論	45
謝辞	46
参考文献	47

目次

1.1	多次元データを構成する要素.	1
2.1	従来の分析ツールでの分析の流れ.	5
2.2	パラレルコーディネートの例.	7
3.1	表とパラレルコーディネートの変換.	8
3.2	表とパラレルコーディネートを組み合わせた表現.	10
3.3	パラレルコーディネート形式の表示.	11
4.1	ツールの外観. (左) 表形式, (右) パラレルコーディネート形式での表示.	13
4.2	矩形選択.	14
4.3	遷移選択の例.	15
4.4	配置変更時のアニメーション.	16
4.5	ハイライトの例.	18
4.6	「時刻」へのシフト操作.	19
4.7	全属性へのシフト操作.	20
4.8	カテゴリ情報に対するインスタントビューの適応例.	21
4.9	数値情報に対するインスタントビューの適応例.	21
4.10	グループ化の例.	22
4.11	データ分布を表す色づけ.	23
5.1	「時刻」のパラレルコーディネート形式での提示.	25
5.2	「コーヒー」と「お茶」に対してシフト操作を行った例.	26
5.3	各価格に対してシフト操作を行った例.	26
5.4	「50円」の商品に対してインスタントビュー操作を行った例.	27
5.5	「コーヒー」に対してインスタントビュー操作を行った例.	27
5.6	「80円」の商品に対してインスタントビュー操作を行った例.	28
5.7	コーヒーの購買履歴に対して, 各価格についてシフト操作を行った例.	28
6.1	評価ツールの外観.	31
6.2	タスクの正答率.	34
6.3	属性数, 分析タスクごとの正答率.	35
6.4	被験者ごとのタスク達成時間.	36
6.5	タスク達成時間.	36

6.6	属性数, 分析タスクごとのタスク達成時間.	37
6.7	各タスク終了時に行ったアンケートの回答.	38
6.8	全タスク終了時に行ったアンケートの回答.	39
6.9	属性数, 分析タスクごとのアンケートの回答.	40

表目次

3.1 多次元データの代表的な表現手法の比較.	9
5.1 購買履歴データの例.	24
6.1 実験タスクの一覧.	33
6.2 タスク達成時間との相関.	38

第1章 序論

現在、我々が生活を送る中で、様々な出来事がデータとして記録されている。例えば、何か商品を購入すると購買データが記録される。これらのデータは、経営戦略の立案などの意思決定に利用される。データを分析し、ヒット商品の法則や年齢層による購買傾向といった知見を得ることで、これらの活動が支援される。データの特徴を直感的に把握し、効果的に分析を行うための手段として、視覚的分析ツールが用いられており、開発が盛んに行われている。

1.1 多次元データ

多次元データとは、複数の次元を持つデータのことである。1件のデータに対して複数の情報を付け加えることができる。多次元データは、データ中の次元を表す「属性」、データ1件を表す「レコード」、各レコード中の個々の属性に対応するデータを表す「データ要素」から構成される(図1.1)。例えば、商品の購買データは「購入時刻」や「購入者」や「商品名」などの次元から構成される多次元データである。多次元データを分析し、各属性のデータ分布や属性間の関係を把握することで、有益な情報を取得することができる。例えば、購買データから顧客の購買傾向を知り、販売戦略を考えることができる。

属性			
属性A	属性B	属性C	...
A'	B'	C'	...
A''	B''	C''	...
.

図 1.1: 多次元データを構成する要素。

1.2 視覚的分析

多次元データからの有益情報の取得を支援するために、視覚的分析ツールが開発されている。視覚的分析では、データが視覚的に表現され、図からデータの特徴を直感的に読み取ることができる。図に対するインタラクションによって、得たい情報に適した図に変化させ、分析を進めていく。また、視覚

的分析ツールを用いた分析活動では、データの概観と詳細を行き来する場面が多く存在する。データの概観とは、データの全体的な傾向を指し、データ分布を読み解くことで得られる。データの詳細とは、元データの持つ詳細な情報を指す。元データの持つ全ての属性及びレコードがこれに当たる。例えば、周期性や傾向といった概観に基づく情報を取得した時、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。さらに、詳細なデータを調べる中でデータに対する気づきがあった場合、再び概観に戻り分析を行う。このように、一連の分析の中で、概観と詳細を相互に把握しあうことが重要である。

1.3 視覚的分析における問題

従来の視覚的分析ツールでは、複数のビューを提供することで、概観と詳細を提示している。また、個々のビューでは提示する情報に適した表現方法が用いられ、ツール上で複数の表現方法が用いられることが多い。異なる表現方法の図同士を見比べることは困難であり、ビューを見比べる時に一方のビューから目を離さなければならない。そのため、従来ツールでは概観と詳細を行き来する間に、分析中に得られたイメージや知見が失われてしまう可能性がある。概観と詳細を切り替えた時にイメージが失われてしまった場合には、もう一度同じ分析を行わなければならない、分析活動が中断されてしまう。

1.4 本研究の目的

本研究の目的は、図に対するイメージや知見を保ちながら継続的に分析が行える視覚的分析ツールの開発である。これを実現するために、視覚的分析においてデータの概観と詳細の滑らかな切り替えを可能にする。このアプローチとして、詳細の把握に適した表と概観の把握に適したパラレルコーディネートを組み合わせた表現とその表現に対する操作を備えた視覚的分析ツールを開発した。本ツールでは、表とパラレルコーディネートを滑らかに切り替えることができ、データの概観と詳細を自由に行き来することができる。

1.5 本研究の貢献

本研究の貢献は、詳細と概観を滑らかに切り替え、双方を自由に行き来できる視覚的分析ツールを開発したことである。分析活動の中で詳細と概観を自由に行き来できることで、分析活動を妨げることなく知見を十分に活用しながら、分析を行えるようになった。

もう一つの貢献は、表とパラレルコーディネートを組み合わせた多次元データの表現手法及び操作を開発したことである。表とパラレルコーディネートは、それぞれ多次元データの詳細と概観を提示する代表的な表現手法であり、視覚的分析ツールでも多く利用されている。本手法は両表現手法の特徴を合わせ持っており、従来の操作方法をそのまま利用でき、学習コストを抑えることができる。

1.6 論文の構成

第2章では、従来の視覚的分析ツールについて整理し、視覚的分析ツールの要件をまとめる。第3章では、データの概観と詳細の自由な行き来を可能にする、表とパラレルコーディネートを組み合わせた手法について示す。第4章では、表とパラレルコーディネートを組み合わせた表現と本表現に対する操作を備えた視覚的分析ツールについて示す。第5章では、購買履歴データを用いた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、表現の組み合わせ方による分析への影響を明らかにするために行った評価実験について示す。第7章では、視覚的分析における本研究の位置づけについて示す。第8章では、研究の結論を述べる。

第2章 多次元データの視覚的分析ツール

従来の視覚的分析ツールについて整理し、現状の問題点から多次元データ分析に必要な要件をまとめる。

2.1 従来の視覚的分析ツール

視覚的分析ツールでは、大規模なデータや複雑な関係を持ったデータについて、データの特徴を直感的に把握できるようにデータを視覚的表現へと変換する。従来の視覚的分析ツールでは、データ内の周期性や偏りといった、概観に基づく特定の情報を強調するように変換を行う。概観とは、データの全体的な傾向のことを指し、データ分布から読み解くことができる情報のことである。分析の最終段階として、得られた知見の正当性や妥当性を検証するために、詳細を確認する必要がある。視覚的表現への変換時に、特定の概観を強調するために、詳細が図に表されず失われてしまうことがある。詳細とは、システムに入力される元データを指し、属性とレコードに基づく形式で書かれた文字列である。多次元データにおいては、データ要素が詳細に当たり、データ要素の分布や関係が概観に当たる。このため、従来の分析ツールでは概観を提示する概観ビューと詳細を提示する詳細ビューの複数のビューを備える場合が多い。

多次元データ分析に用いられている表計算ソフト Excel¹では、表を詳細ビューとして用い、統計グラフを概観ビューとして用いている。表では元データ全てが提示され、表内の任意の部分について範囲を選択する。そして、選択した部分について統計グラフを用いて概観を提示する。データについて調べたい事柄がある場合、棒グラフや折れ線グラフ等の得たい情報に適した統計グラフを作成する。他の分析ツールでも同様に、把握したい事柄によってビューを切り替えながら、分析が行われる。図 2.1 は、ある購買データについての分析の流れを示している。表から発見した特徴的な傾向を示す部分について、統計グラフへと変換し概観を調べている。図から気づきを得てさらに詳しく調べたい時、表もしくは統計グラフへと図を切り替えて分析を進めていくため、分析中に詳細と概観の行き来が頻繁に行われる。

このように、複数のビューを用いることで、図から情報を読み取る、表現を変換するといった本来の分析活動に加え、ビューを見比べる作業を行わなければならない。従来の分析ツールでは、概観に基づく特定の情報について直感的に把握することができるが、最終的にデータから知見を得る段階まで考えた時に、詳細を確認するためにビューを見比べなければならない、分析活動が複雑になってしまう。

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



図 2.1: 従来の分析ツールでの分析の流れ。

2.2 視覚的分析ツールの要件

2.1 節で述べた従来ツールでの分析の複雑さを軽減する為に、1つのビュー上で詳細と概観の把握を可能にする。これを実現するために、概観と詳細それぞれに適した表現手法を滑らかに変換させることを考えた。表現を滑らかに切り替えることで、変化前後の図の対応関係を自然に把握することができる。さらに、概観と詳細について、従来ツールと同等の分析活動をサポートしなければならない。詳細と概観を自由に行き来しながら分析を行うための要件を以下にまとめた。

1. 詳細情報を常に閲覧できる。
2. データ分布を直感的に把握できる。
3. 属性及びレコードへの所属関係を把握できる。
4. 個々のデータ要素を常に識別できる。

まず、分析活動を行うためには、詳細と概観のそれぞれを把握できる必要がある。要件1は、詳細の把握を支援するための要件である。分析中は常に詳細情報を閲覧できるようにする。分析中に、詳細情報を閲覧したい場面は多くあり、図を変化させることなく詳細情報にアクセスできるようにする。多次元データにおいて詳細情報はデータ要素に当たり、データ要素が常に図に表現されるようにする。

要件2は、概観の把握を支援するための要件である。概観の直感的な把握を支援するために、視覚的表現の位置関係によってデータ分布を提示する。データ分布は、視覚的分析においてデータの特徴を把握する為に重要な情報である。Bertin[1]がまとめた視覚的表現の種類の中で、位置関係による情報提示がデータ分布を最も把握しやすい視覚表現である。このことから、データ分布を位置関係によって表現することで概観の直感的な把握を支援できると考えた。

要件3は、多次元データを詳しく分析するための要件である。多次元データの構成要素の関係を把握することで、多次元データを多角的に分析することができる。データ要素が所属するレコード及び属性を理解することで、着目したデータの全属性での分布を把握できる。任意のデータ要素について、着目したデータ要素の属性及びレコードへの所属関係が把握可能でなければならない。

要件4は、詳細と概観を切り替えた時の、図の対応関係の把握を支援するための要件である。詳細と概観の切り替え前後で、データ要素の表現方法を統一することで、対応関係の把握を支援する。これは、多次元データの最小単位であるデータ要素について、常に統一した表現を用いることで、図が変化しても個々のデータ要素を識別できると考えたためである。データ要素に統一した表現を用いることで、データ要素の位置が変更されても対応関係を把握できる。これらの要件を満たすことで、表現を滑らかに変化させ、詳細と概観を自由に行き来しながら分析を行える。

2.3 多次元データ分析ツールで用いられる表現手法

多次元データ分析ツールで用いられる代表的な表現手法として、表とパラレルコーディネートと散布図がある。表は、詳細の提示によく用いられ、時刻表や成績表など広く利用されている表現である。パラレルコーディネートと散布図は、多次元データ分析ツールでは概観の提示によく用いられている。パラレルコーディネートでは、次元ごとのデータ分布を一覧できる。散布図では、2つの次元の関係を直感的に把握できる。

2.3.1 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、データ同士の関係性を表現する手法である。視覚的分析ツールでは、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。データの基本的な情報(属性名や属性数)の読み取りやすさと個々のデータ要素を閲覧しやすい特徴を持つ。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まなければならない、データ分布を直感的に把握することは難しい。視覚的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。

2.3.2 パラレルコーディネート

パラレルコーディネート [2] 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は、多次元データを以下の手順で描く。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。図2.2では、縦に伸びる太い線が軸、横方向に伸びる灰色の各線がレコードに対応し、灰色の線と軸が交わる位置で各次元のデータ分布が表されている。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。データの概観の提示に、パラレルコーディネートがよく用いられている。

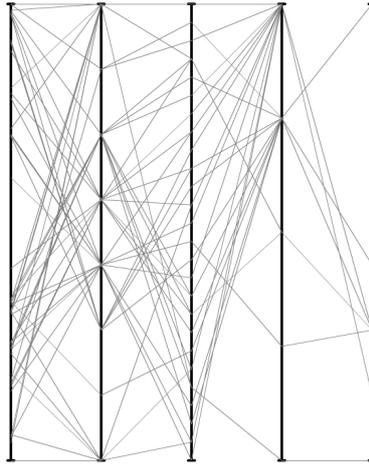


図 2.2: パラレルコーディネートの例.

2.3.3 散布図

散布図は、縦と横の2つの座標軸にデータ中の次元を対応させ、データが存在する位置に点を配置する表現手法である。点の位置関係から、直感的にデータ分布を把握することができる。散布図では、座標軸が属性を表現し、点がレコードを表現している。ただし、散布図の点は、元データのレコードから座標軸となっている2属性だけについて表現する。また、データ要素は図中には明確に表現されない。視覚的分析ツールでは、データから任意の2次元を選択し、選択された2次元のデータ分布を散布図を用いて提示する。

第3章 提案手法

第2.2節の要件を満たす分析ツールを構築するために、表とパラレルコーディネートを組み合わせた手法を開発した。本手法では、データを表現した図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表現形式を滑らかに変換させる(図3.1)。

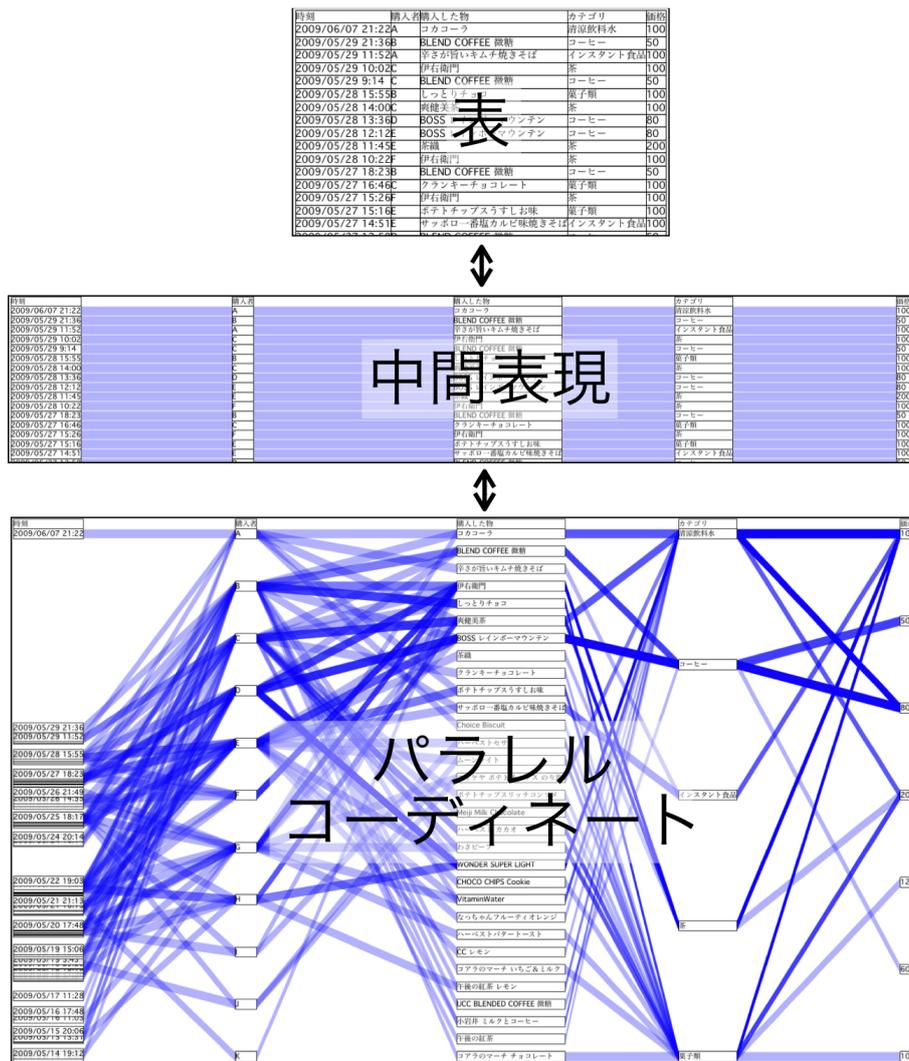


図 3.1: 表とパラレルコーディネートの変換。

取得したい情報に合わせて、データ要素を閲覧しやすい表、もしくはデータ分布を直感的に把握できるパラレルコーディネートに自由に変換できる。また、表現形式が変更する時にはアニメーションを用いて滑らかに図が変化する。このようにすることで、変化前後の図の対応関係が把握しやすくなり、図に対するイメージや知見が失われにくくなる。本手法では、1つのビュー内で表現が混ざり合うことがあり、2表現の中間的な状態が存在する。この中間的な表現によって、表及びパラレルコーディネートへと滑らかに変換することができる。本手法は、表とパラレルコーディネートを組み合わせた表現手法とデータ要素の配置方法によって構成される。

3.1 表現手法の設計

概観及び詳細への滑らかな切り替えを実現するために、概観と詳細それぞれに適した表現手法を組み合わせることを考えた。2.3節にて述べた視覚的分析ツールで用いられる代表的な表現手法について、表現から読み取れる特徴を比較した(表3.1)。表現手法の特徴から、データ要素同士が重なることがなく、詳細情報の提示に適した「表」を詳細に適した表現手法として用いる。ここで、「表」と組み合わせる表現手法を選択するために、各表現手法が表現する情報についてまとめた。「表」では、矩形にテキストが埋め込まれた「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セルの関係を表現する。「パラレルコーディネート」では、軸上の「位置」によって「データ要素」を表し、「横方向の位置」と軸間の「線」で関係を表現する。「散布図」では、「データ要素」は表現されず、「座標軸」で属性を表す、「点」でレコードを表現する。「表」と「パラレルコーディネート」は、データ要素と要素間の2つの関係を表現している。これは各々の表現の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。このことから、2表現を組み合わせることが可能であると考えた。概観に適した表現手法として、「表」と表現する情報が似た「パラレルコーディネート」を用いる。

表3.1: 多次元データの代表的な表現手法の比較。

表現手法	データ要素	属性	レコード	特徴(詳細)	特徴(概観)
表	セル	縦の並び	横の並び	○	×
パラレルコーディネート	軸上の位置	軸	線	×	○
散布図	×	座標軸	点(一部)	×	○

3.2 表現手法

表とパラレルコーディネートを組み合わせるために、どちらの表現形式においてもデータ要素に同一の表現を用いる。常に統一した表現形式を用いることで、個々のデータ要素を識別でき、要件4を満たすことができる。両表現形式で統一してデータ要素を表現するために、データ要素を表と同様の「セル」として表現する。表形式ではセルを用いたことから、そのままの形で扱うことができる。パラレルコーディネート形式では、軸上の位置に対応させてセルを配置する。さらに、任意のデータ要素について属性及びレコードへの所属関係を常に把握できる必要がある。任意のセルについてレコード

への所属関係を提示するために、同じレコードに属するセル同士は線で繋ぐ(図 3.2)。これは、パラレルコーディネート形式のデータ要素を繋ぐ線に倣った形式である。パラレルコーディネートと同様に一本の線がレコードに相当し、セルに繋がる線を辿ることでレコードを把握できる。縦方向の位置が離れていても、同じレコードのセルを特定することができる。このようにすることで、1つのビュー内で表形式とパラレルコーディネート形式が混ざり合った時にも、データ要素と属性及びレコードを把握することができる。セルを繋ぐ線によってレコード、横方向の位置関係によって属性を把握でき、要件 3 が満たされる。

データ要素に統一した表現を用いたことで、セルの配置方法の変更によって、表形式とパラレルコーディネート形式を切り替えることができる。ある属性についてセルの縦の位置を揃えると表形式となり、個々のセルの詳細情報を閲覧しやすい。セルの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形式になり、セル中のテキストを読まずともデータ分布を把握できる(図 3.3)。本表現では、同じレコードのセルが離れた位置にある時にも、関連するデータを辿ることができる。また、セルを繋ぐ線はセルの縦幅と同じ高さにし、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。

時刻	購入者	購入した物	カテゴリ	価格
2009/06/07 21:22	A	ココアコーラ	清涼飲料水	100
2009/05/29 21:36	B	BLEND COFFEE 微糖	コーヒー	50
2009/05/29 11:52	A	辛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	C	伊右衛門	茶	100
2009/05/29 9:14	C	BLEND COFFEE 微糖	コーヒー	50
2009/05/28 15:55	B	しっとりチョコ	菓子類	100
2009/05/28 14:00	C	爽健美茶	茶	100
2009/05/28 13:36	D	BOSS レインボーマウンテン	コーヒー	80
2009/05/28 12:12	E	BOSS レインボーマウンテン	コーヒー	80
2009/05/28 11:45	E	茶織	茶	200
2009/05/28 10:22	F	伊右衛門	茶	100

図 3.2: 表とパラレルコーディネートを組み合わせた表現。

3.3 セルの配置

セルの配置方法によって、表形式もしくはパラレルコーディネート形式に切り替わる。横方向については、同じ属性に所属するセルの位置を揃えるようにする。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、位置を揃えて配置する。縦方向については、表形式とパラレルコーディネート形式の配置方法がある。表形式では、セル同士が縦方向に隣接するようにする配置され、セル同士が重ならないように、一定の順番に沿って並ぶ。パラレルコーディネート形式の配置では、セルをある属性のデータ分布を表すようにする(図 3.3)。実際の分析シーンでは、多様なデータが扱われているため、データの性質によって適した配置方法を用いる。データが数値情報の場合は、セルの持つデータが縦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を軸の上端、最小値を軸の下端とし、各セルがデータに対応するように配置する。数値以外の情報の場合は、同じデータを持つセ

ルを1カ所にまとめ、異なるデータを持つセル同士を等間隔に配置する。図3.3では、左端と右端の軸を数値情報として並べ、他の軸を数値以外の情報として並べている。

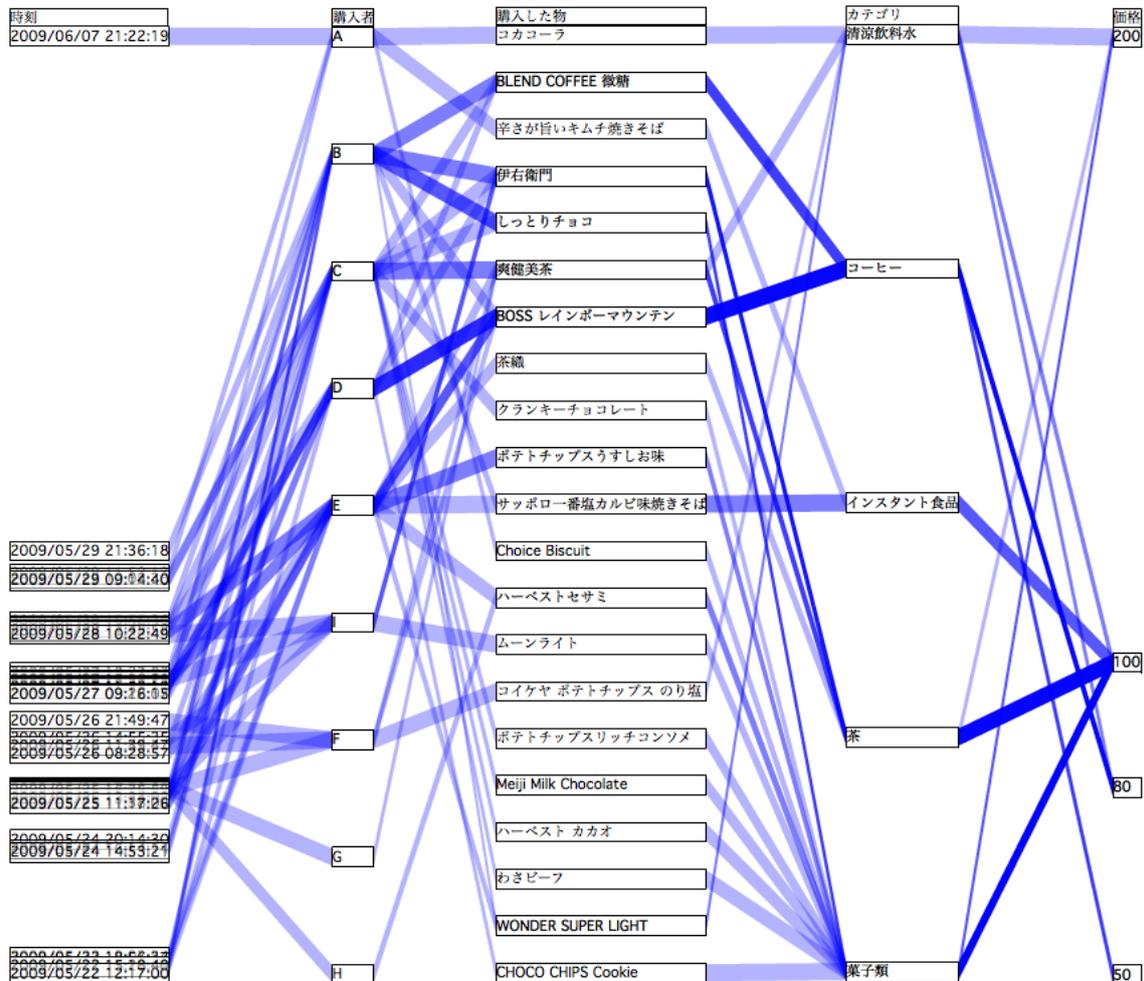


図 3.3: パラレルコーディネート形式の表示。

3.4 提案手法の特徴

提案手法では、詳細もしくは概観に適した表現形式に自由に切り替えることができる。表形式では個々のデータ要素を閲覧しやすく、詳細の分析に適している。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形式では、多次元データの分布を直感的に読み取れ、概観の分析に適している。しかし、データ要素同士が重なってしまうことがあり、詳細な分析には適していない。分析場面に合わせ、適した表現形式に変換することで、円滑に分析を進めることができる。詳細な分析を行いたい場面では表形式を用いることで、詳細な情報を閲覧でき

る。概観に基づく分析を行いたい場面ではパラレルコーディネート形式を用いることで、データ分布を直感的に把握できる。詳細と概観それぞれを把握でき、要件 1,2 を満たすことができる。

本手法では、表とパラレルコーディネートの中間的な表現が存在する。2 表現の中間的な表現形式によって、表現を滑らかに切り替えることができる。これにより、図から得られた知見を見失わずに詳細と概観を行き来することができる。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。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。さらに、分析中に気になった部分があれば、図の全体像を残しつつ、さらに分析を進めることが可能である。

第4章 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(図 4.1)。本ツールは、表と平行ルコーディネートを組み合わせた表現とその表現に対する操作を備えている。操作について、表と平行ルコーディネートそれぞれに対する既存の操作を行えるようにした。

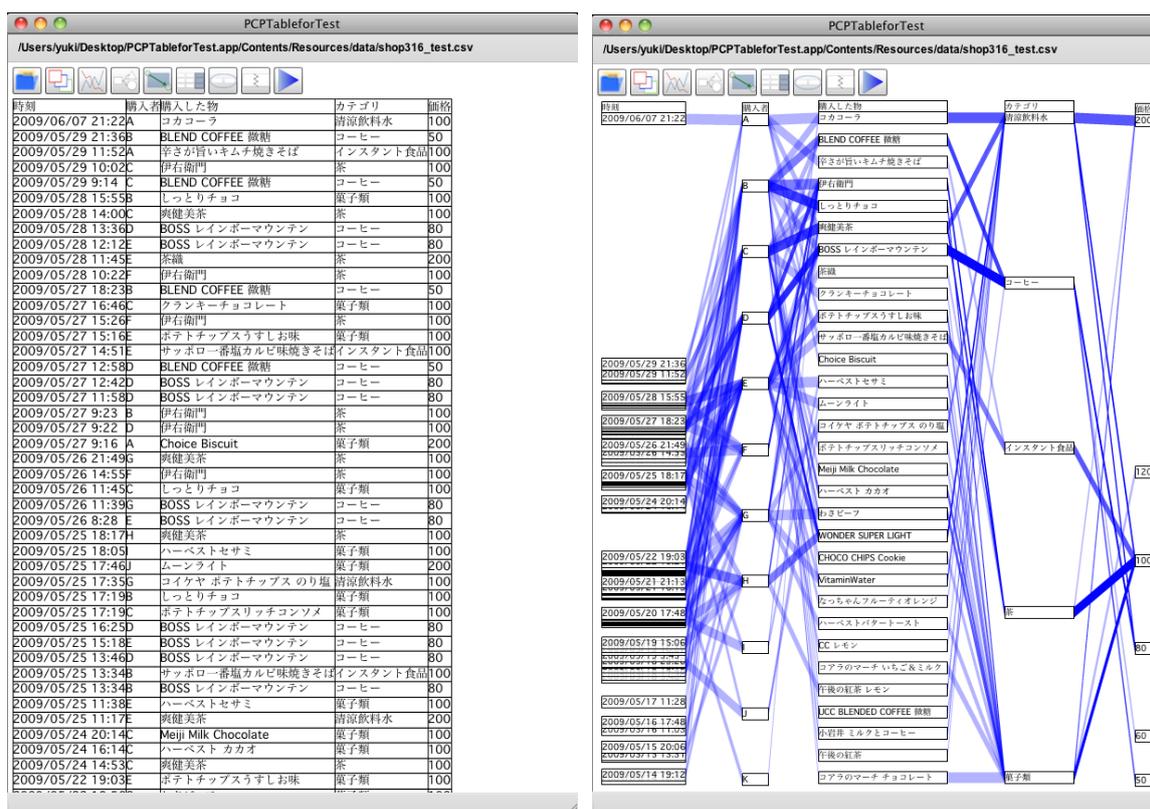


図 4.1: ツールの外観。(左) 表形式, (右) 平行ルコーディネート形式での表示。

4.1 本ツールの操作

本ツールの備える操作として、データ選択、配置の変更、関連データの閲覧、グループ化、色づけ、ビューの調整がある。データ選択、関連データの閲覧は、表と平行ルコーディネートに対する既存の操作を統合した操作である。本ツールでは、分析場面に合わせて表形式もしくは平行ルコーディネート

形式に表現を変化させられる。ツールの学習コストを抑えるために、本ツールでは表と平行コーディネートに対する既存の操作を行えるようにする。そのために、表及び平行コーディネートに対する既存の操作を比較し、操作の統合について検討した。配置の変更は任意の部分で、表形式もしくは平行コーディネート形式に変換する操作である。グループ化、色づけ、ビューの調整は、多次元データ分析のための操作である。

4.1.1 データ選択

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。表では、セルに対して選択操作を行い、属性及びレコードに基づいてデータを選択する。平行コーディネートでは、軸及び軸間の線に対して選択操作を行う。軸上の位置や線の角度から読み取ったデータ分布に基づいて選択する。本ツールでは、セルに対する範囲選択法と元データに基づく選択法を備えることで、両手法での選択操作を可能にする。両表現の操作を備えることで、表と平行コーディネートの選択法に似た感覚でデータ選択を行える。

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。図 4.2 では、購買履歴データ内の「購入時刻」のデータ分布から、中央から下にかけての一連のデータ(5月24日から5月29日までの期間)を選択している。

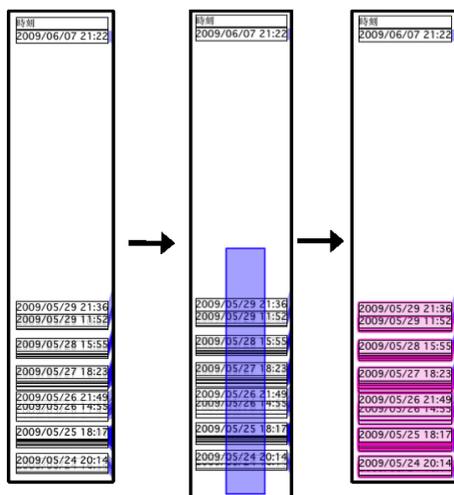


図 4.2: 矩形選択.

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へと、属性とレコードに基づいた範囲を遷移して選択する(図 4.3)。レコードや属性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。遷移する順番については、本ツールでの利用が多い順になっている。

時刻	購入者	購入した物	カテゴリ	価格
2009/06/07 21:22A		コココーラ	清涼飲料水	100
2009/05/29 21:36B		BLEND COFFEE 微糖	コーヒー	50
2009/05/29 11:52A		辛さが旨いキムチ焼きそば	インスタント食品	100
2009/05/29 10:02C		伊右衛門	茶	100
2009/05/29 9:14 C		BLEND COFFEE 微糖	コーヒー	50
2009/05/28 15:55B		しっとりチョコ	菓子類	100
2009/05/28 14:00C		爽健美茶	茶	100
2009/05/28 13:36D		BOSS レインボーマウンテン	コーヒー	80
2009/05/28 12:12E		BOSS レインボーマウンテン	コーヒー	80
2009/05/28 11:45E		茶織	茶	200

時刻	購入者	購入した物	カテゴリ	価格
2009/06/07 21:22A		コココーラ	清涼飲料水	100
2009/05/29 21:36B		BLEND COFFEE 微糖	コーヒー	50
2009/05/29 11:52A		辛さが旨いキムチ焼きそば	インスタント食品	100
2009/05/29 10:02C		伊右衛門	茶	100
2009/05/29 9:14 C		BLEND COFFEE 微糖	コーヒー	50
2009/05/28 15:55B		しっとりチョコ	菓子類	100
2009/05/28 14:00C		爽健美茶	茶	100
2009/05/28 13:36D		BOSS レインボーマウンテン	コーヒー	80
2009/05/28 12:12E		BOSS レインボーマウンテン	コーヒー	80
2009/05/28 11:45E		茶織	茶	200

時刻	購入者	購入した物	カテゴリ	価格
2009/06/07 21:22A		コココーラ	清涼飲料水	100
2009/05/29 21:36B		BLEND COFFEE 微糖	コーヒー	50
2009/05/29 11:52A		辛さが旨いキムチ焼きそば	インスタント食品	100
2009/05/29 10:02C		伊右衛門	茶	100
2009/05/29 9:14 C		BLEND COFFEE 微糖	コーヒー	50
2009/05/28 15:55B		しっとりチョコ	菓子類	100
2009/05/28 14:00C		爽健美茶	茶	100
2009/05/28 13:36D		BOSS レインボーマウンテン	コーヒー	80
2009/05/28 12:12E		BOSS レインボーマウンテン	コーヒー	80
2009/05/28 11:45E		茶織	茶	200

時刻	購入者	購入した物	カテゴリ	価格
2009/06/07 21:22A		コココーラ	清涼飲料水	100
2009/05/29 21:36B		BLEND COFFEE 微糖	コーヒー	50
2009/05/29 11:52A		辛さが旨いキムチ焼きそば	インスタント食品	100
2009/05/29 10:02C		伊右衛門	茶	100
2009/05/29 9:14 C		BLEND COFFEE 微糖	コーヒー	50
2009/05/28 15:55B		しっとりチョコ	菓子類	100
2009/05/28 14:00C		爽健美茶	茶	100
2009/05/28 13:36D		BOSS レインボーマウンテン	コーヒー	80
2009/05/28 12:12E		BOSS レインボーマウンテン	コーヒー	80
2009/05/28 11:45E		茶織	茶	200

図 4.3: 遷移選択の例

4.1.2 配置の変更

選択中のセルに対して、3.3 節であげた表形式及びパラレルコーディネート形式の配置に変更する操作である。データ分布を把握したい時にはパラレルコーディネート形式を用い、詳細な情報を知りたいときには表形式を用いる。配置を変更する時に、アニメーションを用いて滑らかに位置を変化させる。図 4.4 では、左端の表形式の配置から右端のパラレルコーディネート形式に変換している。アニメーションによって、変化前後の図の対応関係を把握しやすくなる。例えば、ある商品の売り上げ時刻が集中していることを発見し、該当する部分のデータ分布を閲覧したいと考え、パラレルコーディネート形式に表現を変化させる。変化の際にアニメーションによって注目している部分を把握しながら、次の分析へと進むことができる。配置の変更時には、軸の上端と下端の位置が保たれるように移動する。つまり、軸の上下の位置は一定であり、配置変更によって最大値と最小値のマッピングのみが変更される。このようにすることで、軸全体の位置が保たれ、配置の変更時に視線の動きを少なくできる。

は、表現形式を一部分だけ変更して、関連するセルを提示する。

セルや線に色を付けることで、図のレイアウトを変更せずに、付加的な情報を表現することができる。しかし、本手法では複数の色が重なった時に色の濃さを読み取りにくくなってしまうため、着目点が1つの場合にのみハイライトを用いる。一方、セルの位置関係によって付加的な情報を提示する場合、着目点が複数ある場合にもそれぞれを識別しやすい。本ツールでは、着目点が複数の場合及び着目点同士を比較したい場合には、位置によって関連データを提示する。図に対するイメージを維持するために、ユーザ操作を用いた場合にのみ、セルの位置によって関連データが提示される。

ハイライト

着目しているデータと関連するデータを色によって閲覧するための操作である(図4.5)。表では、選択中のセルが点線や太線で囲われ、選択部分分かるようにハイライトする。パラレルコーディネートでは、選択中の線の色の変更や選択していない部分の透明度を高くすることで、選択部分をハイライトする。本ツールにおいては、あるセルが選択された時に、表と同様にセルに色をつけ、セルと関連する線についてパラレルコーディネートと同様に色を付ける。これにより、表とパラレルコーディネートのハイライト操作と同等の関連データの閲覧が行える。

着目しているセルが青色にハイライトされ、そのセルと関連する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。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。図4.5では、青色にハイ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。着目したセルと同レコードのセルがどのように分布しているか把握することができる。

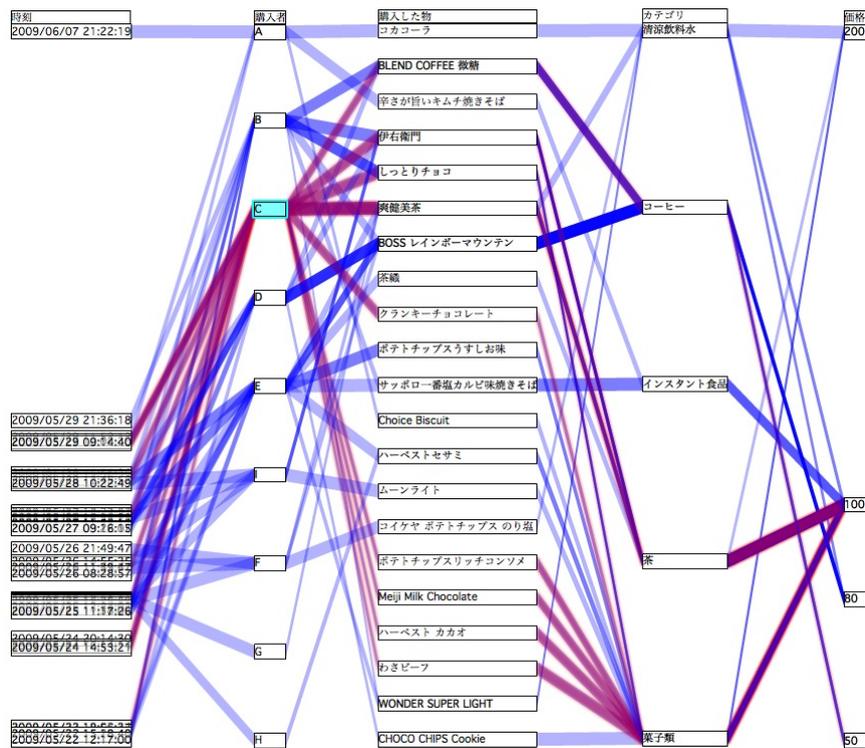


図 4.5: ハイライトの例

シフト操作

着目したデータについて、位置によって関連するデータを閲覧する操作である (図 4.6)。着目しているセルをマウス操作によって移動させると、関連するセルも同様に位置がずれ、所属する属性に対応する軸から関連するセルだけが抜き出される。この時に、パラレルコーディネートの性質を保つために、セルの横幅と移動距離の比を保つように関連するセルを移動させる。このようにすることで、選択中のセルの左端を軸の右端に合わせた場合、全ての関連するセルが軸の右端に位置するようになる。関連するセルが元の軸から抜き出されるように移動し、元々の軸との位置関係から関係するセルを把握できる。図 4.6 では、赤色のセルに対してフィルタリングを行い、そのセルと関連するセルが軸から左に離れた位置にずれている。一番左の軸の左側に赤色のセルと関係するセル、軸の直下には関係しないセルが並び、データ分布を同時に把握できる。また、シフト操作を行った時に軸に近い位置にセルを置くことで、シフト操作前の概観を崩さずに、セルと関連するデータの分布を把握することができる。つまり、着目点と関連するデータ分布と関係しないデータ分布に加え、元々のデータ分布を同時に把握できる。セルの位置によって自身の調べた情報が残り、後から見返した時に自然に履歴が残される。複数の着目点についてシフト操作を行うことで、着目しているデータ同士を比較できる。

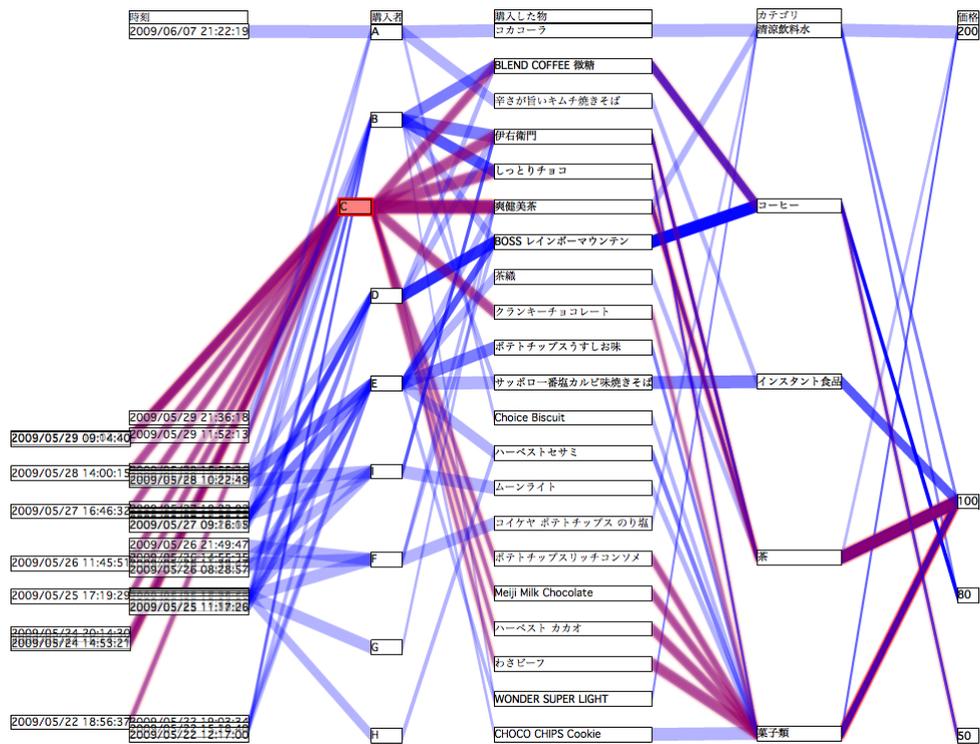


図 4.6: 「時刻」へのシフト操作

シフト操作の適応範囲として、「1つの属性」、「全属性」をサポートしている。「1つの属性」では、セルの位置をずらす軸を1つだけに限定することで、図全体の変化を少なくし概観を保ちながら、関連するデータを閲覧できる。データの特徴をある程度把握し、調べたい属性が分かっている場面で利用する。「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いにくい。データの特徴をあまり掴めていないような分析の初期段階での利用を想定している。

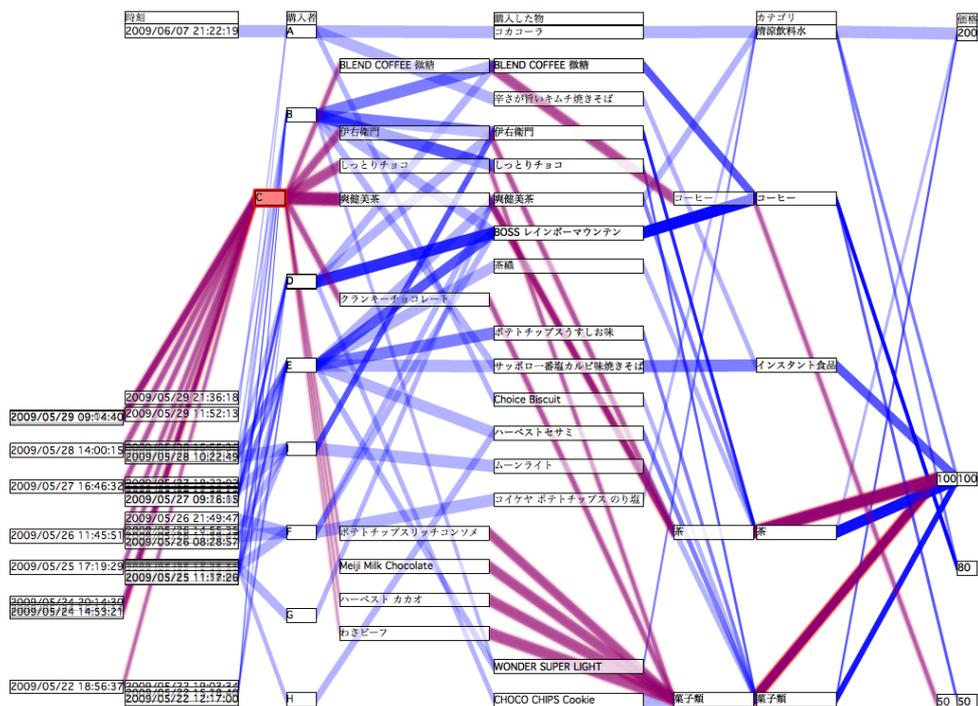


図 4.7: 全属性へのシフト操作

インスタントビュー操作

関連するデータを閲覧するために、任意の部分の表現形式を変換する操作である。シフトキーを押しながらセルにマウスホバーすることで、選択したセルと関連するセルが、現在用いている表現形式と異なる形式に変換される。パラレルコーディネート形式で表現されている部分では表形式に変換され、表形式で表現されている部分ではパラレルコーディネート形式に変換される。これにより、データ分布から得られたデータの特徴について、詳しく調べることができる。また、詳細情報から発見したデータの連続性や特異性について、データ分布を調べることもできる。図 4.8 では、左図のパラレルコーディネート形式の軸から、赤色のセルにインスタントビュー操作を行い、右図において操作を行ったセルと同じデータを持つセルを表形式で提示している。このように、概観から得られた情報について、図の全体像を保持しながら詳細へと切り替えることができる。

インスタントビュー操作の適応範囲が、データの種類によって異なる。対象セルの持つデータが数値以外の情報の時は、次元内の同一データを変換対象とする。数値情報では、属性内に同一のデータがないセルが多いことが考えられるため、着目したセルと近い値を持つセルを変換対象とする。図 4.9 では、数値情報についてインスタントビュー機能を行い、着目したセルと近いデータを持つセルが、表形式で提示されている。

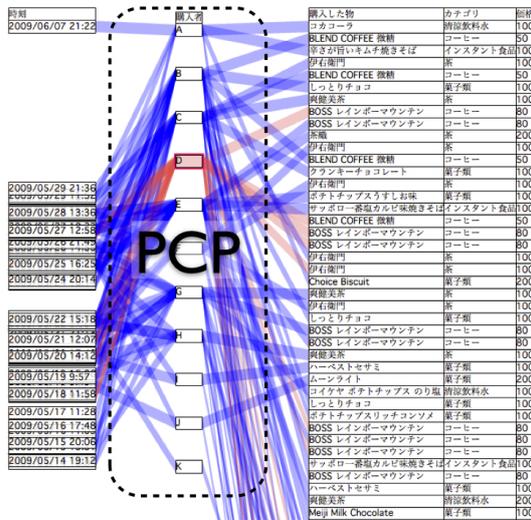


図 4.8: カテゴリ情報に対するインスタントビューの適応例.

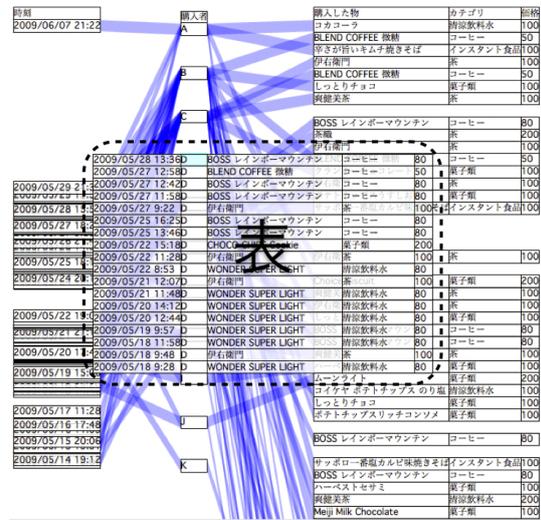


図 4.9: 数値情報に対するインスタントビューの適応例.

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。グループ化したい属性を選択しグループ化を選択すると、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4.10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」の属性を減らすことで全体の構造を明確にすることができる。

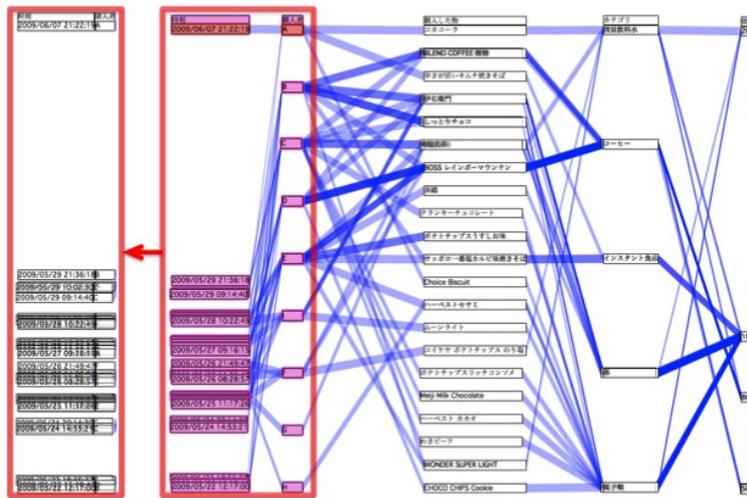


図 4.10: グループ化の例

4.1.5 色付け

着目したデータを目立たせるために、選択したセルについて色を付ける操作である。セルを選択し色付けを選択すると、セルのテキストを囲む枠に色が付けられ、セルと接続する線がセルと同じ色になる。線が接続するセルで異なる色が用いられている場合には、線の端から端に向かってグラデーションを用いて色を付ける。着目した部分に色をつけておくことで、配置変更前後により明確に関係を把握することができる。

また、データ分布を表すように色付けを行う機能をサポートしている。分析タスクやユーザの嗜好に合わせて、「特定の色の濃さ」、「赤と緑」、「異なる色」によって色づけを行える。「特定の色の濃さ」では、次元内の最大値を持つセルを選択した色にし、値が小さくなるにつれて薄い色になる。図4.11左では、各次元について異なる色を選択し、色の濃さによってデータ分布が提示されている。「赤と緑」では、次元内の最大値に近いほど赤色になり、最小値に近いほど緑になる。図4.11右では、全次元に対して赤と緑で配色し、軸の上に位置するセルがより赤く、軸の下に位置するセルがより緑に色付けられている。「異なる色」では、それぞれのセルに付けられる色がなるべく識別しやすいように色づけを行う。軸の中間付近の線を眺めることで、グラデーションによる色の変化から、軸間の相関を把握し

やすい。セルが表形式の並びの時に色を付けると、表の配置を維持しつつデータの傾向を読み取れる。

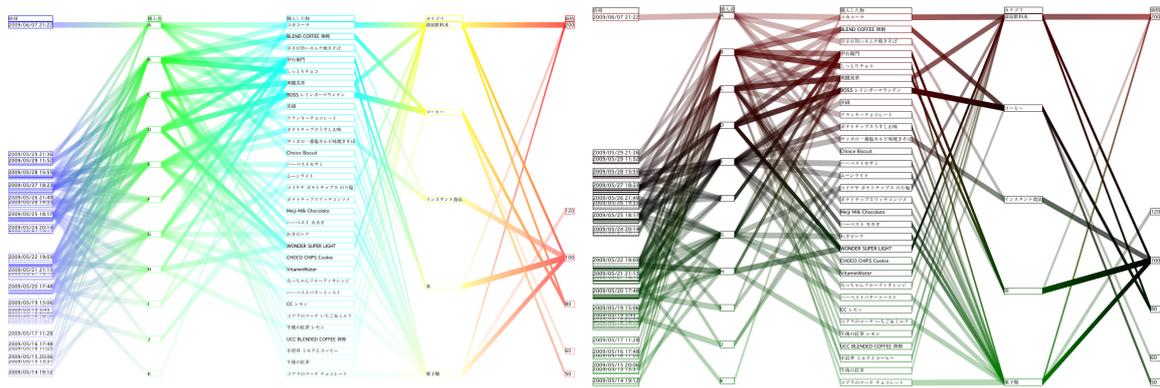


図 4.11: データ分布を表す色づけ。

4.1.6 ビューの調整

ビューに対してパン及びズーム操作を行える。ビューに対してドラッグ&ドロップを行うことでパン操作を実行でき、ビューの位置を見たい情報に合わせて調整できる。ビュー上でマウスホイールを行うことでズーム操作を実行できる。ズームアウトすることで、データの全体像を把握することができる。ズームインすることで、気になる部分について詳細に眺めることができる。これらの操作によって任意の部分に着目して分析を進めることができる。

第5章 ツールの利用例

本ツールでの分析の流れと各操作の利用場面を紹介する。

5.1 購買履歴データの分析例

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5月22日から2009年6月7日の期間に記録された。レコード数は100件、商品を購入したメンバーは11名、購入された商品数は21個である(表5.1)。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つの属性を持ち、1レコードが1つの商品の購買履歴に当たる。「商品カテゴリ」は、商品に対して「菓子」や「茶」のような分類を表す属性である。「購入者」については、元データの各個人に対応させたA-Kのアルファベットで表している。売り上げの向上を目的とし、メンバーや商品の購買傾向を調査する場面を想定して分析を行う。

表 5.1: 購買履歴データの例.

時刻	購入者	商品名	商品カテゴリ	金額
2009/06/07 21:22	A	コカコーラ	清涼飲料水	100
2009/05/29 21:36	B	BLEND COFFEE 微糖	コーヒー	50
2009/05/29 11:52	A	辛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	C	伊右衛門	茶	100
2009/05/29 09:14	C	BLEND COFFEE 微糖	コーヒー	50

まず、データを読み込むとデータが表形式で提示される。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ていく。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5.1)。時刻の軸のセルの位置を見ると、ほぼ一定の間隔で空白があることが発見できる。空白の直後のセルを見ると21:00-23:00頃の購入時刻であり、直前のセルは9:00-10:00頃の購入時刻である。このことから、毎朝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が分かる。さらに、1日の固まりごとに線の濃さを見てみると、固まりの中心辺りに濃い箇所があるものと濃さがほぼ一定のものがあることを確認できる。このことから、昼頃に購入が集中した日と1日を通して平均的に購入された日があることが分かる。「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。セルの位置関係から大まかなデータ分布を把握し、線の濃さからより詳しくデータ分布



図 5.2: 「コーヒー」と「お茶」に対してシフト操作を行った例。

また、著者は、以前からよく売れる商品に関する仮説として、「安いものほどよく売れる」のではないかと考えていた。そこで、この仮説を確かめるために分析を行う。各々の商品の売り上げ数の違いを比べるだけでは、価格と売れ方の相関が把握できない。そこで、各々の価格について「購入時刻」の分布を比較する。まず、「価格」の軸の上方に位置する、価格の高い順にセルを大きく左に移動させる(図 5.3)。「購入時刻」と「価格」の軸が、左から順番に 200 円から 50 円までの順番に並ぶ。各価格の購入時刻の分布を見ると、「購入時刻」の左から 2 番目の軸の全体にセルが存在することから、「100 円」の商品が多く売れていることが分かる。左から 1 番目と 3 番目の軸を見るとセルの位置が似ており、「200 円」と「80 円」の商品の購入時刻が似ていることが分かる。最も右の軸にはセルが少なく、「50 円」の商品があまり売れていないことが分かる。「50 円」の商品は 5 月 27 日以前には購入されておらず、「安いものがよく売れる」という傾向は見られなかった。

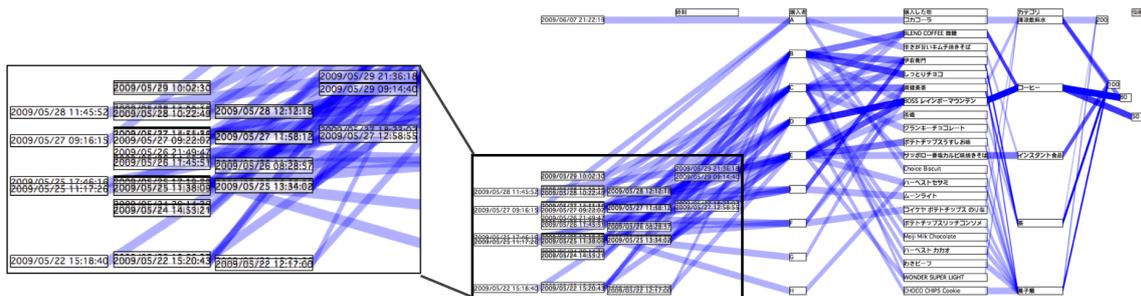


図 5.3: 各価格に対してシフト操作を行った例。

仮説が確かめられなかった原因として、何かの要素が影響を与えているのかもしれないと考え、50 円の商品についてさらに調べてみる。「50 円」の商品に対してインスタントビュー操作を行い、50 円の商品の詳細を閲覧すると、全てが「コーヒー」であることが分かった(図 5.4)。

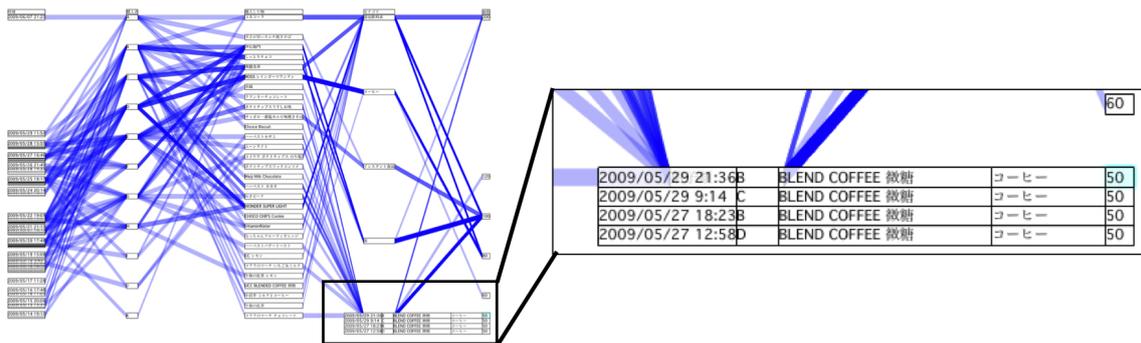


図 5.4: 「50 円」の商品に対してインスタントビュー操作を行った例.

コーヒーは個人によって好みが変わる商品であると考えられるため、「安いものがよく売れる」という傾向が見られなかったのかもしれない。そこで、「コーヒー」について詳しく調べるために、インスタントビュー操作を行う (図 5.5)。表形式の中段の辺りで「BOSS レインボーマウンテン」連続しており、コーヒーの中でよく売れた商品であることが分かる。さらに、コーヒーには 50 円と 60 円と 80 円の商品があることが分かる。

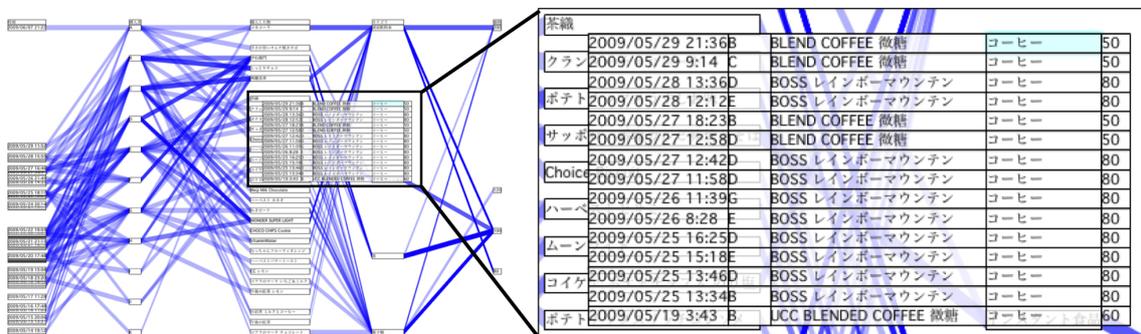


図 5.5: 「コーヒー」に対してインスタントビュー操作を行った例.

ここでコーヒーの購買傾向を把握するために、コーヒーの各価格について調べてみる。50 円のコーヒーについては既に調べているため、60 円と 80 円のコーヒーについて詳細に調べていく。60 円の商品は全データ中で 1 件だけだったため、ハイライト操作を行う。次に、80 円の商品は購買数が多く、ハイライト操作だけでは正確に把握しきれないため、より詳しく調べるためにインスタントビュー操作を行う (図 5.6)。インスタントビューでは時刻順にセルが並べられており、「BOSS レインボーマウンテン」や「WONDER SUPER LIGHT」が隣り合って並んでいることから、同じ商品が連続して購入されていることが分かる。このように、インスタントビュー操作やハイライトを用いて、調べたい事柄に関する情報について詳しく調べることができる。

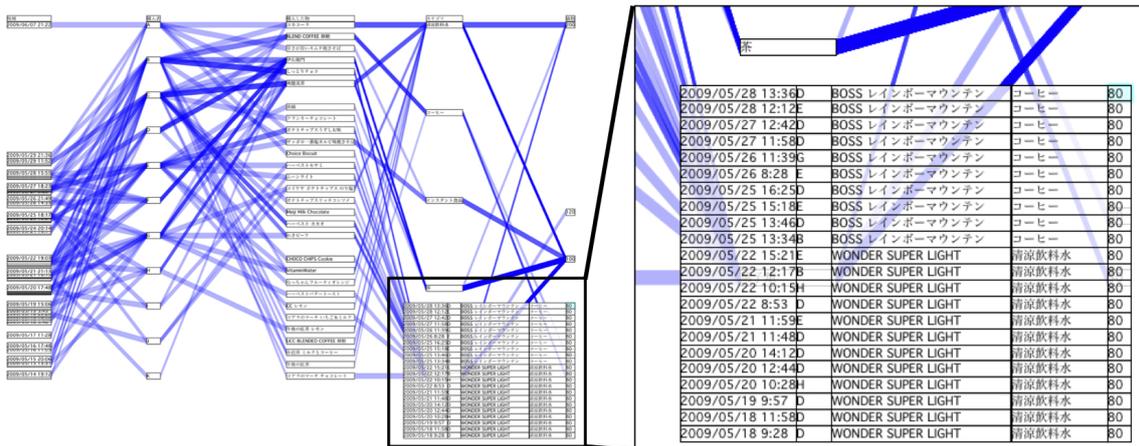


図 5.6: 「80 円」の商品に対してインスタントビュー操作を行った例。

「コーヒー」の購入履歴の中で各価格の購入時刻の比較を行う。まず、「コーヒー」について全属性へのシフト操作を行う。「価格」の軸から「コーヒー」と関係するセルだけが抜き出され、「コーヒー」の価格を把握することができる。次に、コーヒーの各価格について、価格の高い順に右に移動させる(図 5.7)。購入時刻の軸が、左から順にコーヒー以外、50 円、60 円、80 円と並ぶ。各価格の分布を見比べると、80 円のコーヒーが 5 月 25、26 日によく売れ、その少し後の期間になると 50 円のコーヒーが売れ出している。コーヒーに関しては「安いものがよく売れる」のではなく、ある一定以上の品質の商品(60 円、80 円)がよく売れると類推できる。このことから、50 円のコーヒーはあまり人気がなく、60 円以上のコーヒーを多く納入すれば、売り上げが向上すると考えられる。

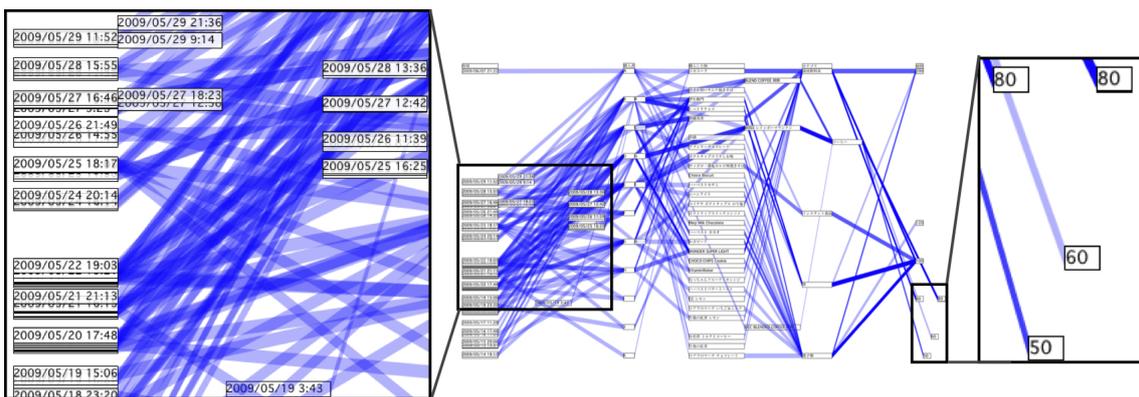


図 5.7: コーヒーの購買履歴に対して、各価格についてシフト操作を行った例。

5.2 考察

分析結果から得られた本ツールでの分析の流れとツールの利点について示す。

5.2.1 本ツールでの分析の流れ

本ツールを用いた分析活動では、始めに表からデータの基本的な情報(属性、データ形式、属性数、レコード数)を読み取る。データの基本的な情報を理解した後に、データ中で気になった属性や特徴的なデータを発見し、そのデータについて詳しく調べていく。図から発見した特徴的な部分を選択し、選択範囲について配置の変更や色付けなどのデータ分布をより詳しく把握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に対してさらに深く分析を行い、必要な情報を読み取っていく。データから知見が得られた場合には、インスタントビュー操作もしくは表形式へと配置変更を行い、詳細情報を確認する。確認時に知見に合わないデータを見つけた場合には、他の属性との関係を閲覧するなど、多角的にそのデータについて調べることができる。

5.2.2 ツールの利点

本ツールの利点として、セルの位置関係と線の濃さによって、大まかなデータ分布とより詳しいデータ分布を把握できることである。特に分布に偏りがあるデータの場合、線が重なった部分について線の濃さから密度を読み取れ、分布を把握しやすい。セルの位置から特定の範囲のデータの有無を把握し、線の濃さから密集した部分の密度を把握することができる。

もう一つの利点として、シフト操作により複数の項目について分布の比較を行える。特に、比較したい項目数が増えた場合に有効である。これは、シフト操作において付加的な情報の提示にセルの位置を用いたことで、個々のデータ分布を識別しやすいためだと考えられる。

さらに、本ツールでは、セルの配置変更やインスタントビュー操作などのセルの位置に影響を与える操作を行いやすい。アニメーションと中間的な表現によって、表とパラレルコオーディネートを滑らかに切り替えられ、図の変化に対する抵抗感が少ないと考えられる。

第6章 評価実験

データ分析に対する本ツールの有効性を評価するために、表とパラレルコーディネートをそれぞれ別々のビューとして備える評価ツールとの比較実験を行った。

6.1 実験の目的

実験の目的は、本ツールのデータ分析への有効性の検証である。これを確かめるために、表とパラレルコーディネートの組み合わせた方による視覚的分析への効果及び影響を明らかにする。表とパラレルコーディネートが滑らかに切り替わる本ツールと、それぞれの表現を別のビューとして備える評価用ツールとの比較実験を行う。この実験で、1つのビューを用いた分析(本ツール)と複数のビューを用いた分析(評価ツール)の違いについて、以下の仮説を立てた。

1. 1つのビューを用いた分析では、分析への疲労を感じにくい。
複数のビューを用いた分析ではビューを頻繁に見比べる必要があり、見比べる活動によって疲労を感じやすいと考えられる。1つのビューを用いた分析では、表現形式を滑らかに切り替えられ、変換前後の対応関係を把握しやすいと考えられる。
2. 1つのビューを用いた分析では、複数のビューを用いた分析と比べて、情報を正確に読み取れない可能性がある。
1つのビューを用いた分析では、1つのデータについて同時に詳細と概観を閲覧することができない。複数のビューを用いた分析では、データの詳細と概観が提示されており、正確に情報を読み取ることができると考えられる。
3. 単純な分析場面では、複数のビューを用いた分析の方がタスクを素早く達成でき、複雑な分析タスクでは、1つのビューを用いた分析の方がタスクを素早く達成できる。
1つのビューを用いた分析では、表現形式の変換にアニメーションを用いており、表現形式の変換に時間がかかる。複数のビューを用いた分析では、既に図が提示されており、ビュー間の対応関係を把握するのみで、変換の必要はない。そのため、変換が少ない分析タスクにおいては、複数のビューを用いた分析の方がタスクを素早く達成できると考えられる。変換時間がタスク達成時間に大きく影響を与えてしまうことが考えたためである。また、多くの変換が必要になる分析タスクでは、変化前後の対応関係を把握しやすい1つのビューを用いた分析の方がタスクを素早く達成できると考えられる。

6.2 実験の概要

本ツールと表とパラレルコーディネートのビューを並べたツール(評価ツール)を用いて、実際のデータ分析を模したタスクを行ってもらった。被験者は、コンピュータサイエンスを学ぶ大学生及び大学院生6名である。被験者には、タスクを行う前にツールの使用方法を説明し、実際のタスクに似せた練習用のタスクを行ってもらい、ツールに慣れてもらった。実験に用いたデータは、5.1節で示した購買履歴データである。全タスク終了後に、ツールに対するアンケートに回答してもらった。

6.2.1 評価ツール

表とパラレルコーディネートの組み合わせによる効果を測定するために、表とパラレルコーディネートをそれぞれ別のビューとして備えるツールを用意した(図6.1)。評価ツールは、左側に表、右側にパラレルコーディネートのビューを備えている。ビュー間のデータの対応関係の把握を支援するために、Linking and Brushing[3]を実装している。Linking[4]とは、複数のビューでハイライトの状態を同期させ、異なるビューにある視覚的表現の対応関係の把握を支援する手法である。評価ツールでは、ハイライト及び選択状態が2つのビューでリンクし、データ要素に対する色づけが同期する。片方のビューで操作を行うと、もう一方のビューでも対応するデータに対し統一した色付けを行う。Brushing[5]とは、任意のデータ範囲を選択することでビュー内に選択部分のデータだけを提示し、着目したデータをハイライトする手法である。評価ツールでは、パラレルコーディネート上でデータ選択を行うと、表でも選択したデータのみが提示される。これは、パラレルコーディネートを用いてデータの概観を提示し、表を用いて詳細情報を提示する従来ツールと同様の形式である。また、パラレルコーディネートの各軸の配置方法については、本ツールと同等の配置方法を用いることができる。

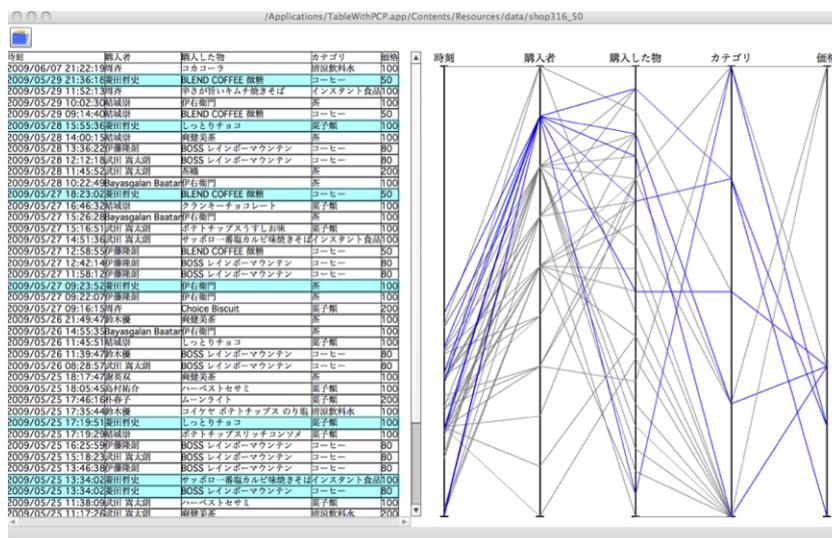


図 6.1: 評価ツールの外観。

6.3 実験タスク

実験タスクは、全 22 個を用意した (表 6.1)、全タスクについて明確な答えが存在するように設計した。さらに、実験タスクを「分析する属性数」と「分析タスクの種類」について設計を行った。それぞれのツールを用いて、タスクを 11 個ずつ行ってもらい、また、被験者を 2 つのグループに分けた。片方のグループには本ツールを用いて 1-11 のタスク、評価ツールを用いて 12-22 のタスクを行ってもらった。もう一方のグループには、用いるツールの順番を入れ替えてタスクを行ってもらった。

6.3.1 分析する属性数

分析する属性数とは、タスク達成のために分析が必要な属性数である。分析する属性数が増えることで分析の複雑さが増し、タスク達成により時間がかかる。分析タスクの複雑さを調整するために、1-3 までの属性数について各 3 個のタスクを用意した。

6.3.2 分析タスクの種類

データ分析の基礎的なタスクとして、データを「特定」するタスクとデータ同士を「比較」するタスクがある。特定とは、データ中から分布やラベルを通して一つのデータを発見するタスクである。比較とは、複数のデータについてのデータ分布を比較するタスクである。「特定」は、さらに「テキスト」と「分布」のタスクの種類に分類した。「テキスト」とは、特定のデータを視覚的に表現されたラベルから発見するタスクである。評価ツールにおいて、表とパラレルコーディネートのデータの対応関係の把握が必要になることの多いタスクである。表でラベルを確認し、ラベルと対応するパラレルコーディネート上の線を探すという分析の流れが多くなる。「分布」とは、特定のデータをセルや点の分布から発見するタスクである。パラレルコーディネートから読み取った分布に対応するラベルを表から探す必要がある。「テキスト」と「分布」では、ビューを見比べる回数が異なると考えられる。

表 6.1: 実験タスクの一覧

タスク番号	質問	属性数	分析タスク
1	最も高い価格は何円ですか?	1	特定 (分布)
2	清涼飲料水は商品カテゴリの中で何番目によく売れていますか?	1	特定 (テキスト)
3	6番目によく売れた商品は何ですか?	1	特定 (分布)
4	200 円の商品の中で最もよく売れた商品は何ですか?	2	特定 (テキスト)
5	3番目によく売れた商品カテゴリの中で、最もよく売れた商品は何ですか?	2	特定 (分布)
6	ハーベストカカオを最もよく買った購入者は誰ですか?	2	特定 (テキスト)
7	最も安い価格の商品の中で最もよく売れた商品を、最もよく買った購入者は誰ですか?	3	特定 (分布)
8	伊右衛門だけを買った購入者の中で購入数が少ない購入者が商品を最もよく買った「日」はいつですか?	3	特定 (テキスト)
9	最もよく売れた商品カテゴリの中で最も良く売れた商品を最もよく買った購入者は誰ですか?	3	特定 (分布)
10	購入者 C と D はどちらが定期的に商品を購入していますか?	2	比較
11	茶と清涼飲料水ではどちらが定期的に商品を購入していますか?	2	比較
12	BOSS レインボーマウンテンは商品の中で何番目に売り上げが多いですか?	1	特定 (テキスト)
13	最も古い売り上げ時刻はいつですか?	1	特定 (分布)
14	Hさんは購入者の中で商品の購入数が何番目に少ないですか?	1	特定 (テキスト)
15	2番目によく売れた商品を最もよく買っている購入者は誰ですか?	2	特定 (分布)
16	Eさんの購入履歴の中で最も多い商品カテゴリは何ですか?	2	特定 (テキスト)
17	2番目によく売れた商品カテゴリの履歴の中で最もよく売れた商品は何ですか?	2	特定 (分布)
18	200 円の商品を最もよく買った購入者が最もよく買う商品は何ですか?	3	特定 (テキスト)
19	2番目によく売れた商品の属する商品カテゴリを最もよく買った購入者は誰ですか?	3	特定 (分布)
20	コーヒーの中でよく売れた商品が最もよく売れた日はいつですか?	3	特定 (テキスト)
21	購入者 E と F はどちらが定期的に商品を購入していますか?	2	比較
22	200 円の商品と 80 円の商品はどちらか定期的に売れているか?	2	比較

6.4 アンケート

各タスク終了時と全タスク終了時にツールに対する印象について回答してもらった。各タスク終了時には、使用したツールで「タスクを行いやすかったか」を5段階で回答してもらった。全タスク終

了時に以下の質問に対して、(1) そう思わない-(5) そう思うの 5 段階で回答してもらった。

- すぐにツールに慣れることができた。
- 直感的に分析を行えた。
- データ分析を行いやすい。
- 分析による疲労を感じにくい。
- データ分析の機会にこのツールを使いたい。

6.5 結果と考察

6.5.1 正答率

各タスクごとの正答率を図 6.2 に示す。正答率は、各タスクを行った被験者の内、正答した被験者の割合である。タスク 22 を除いた全タスクについて、本ツールの方が正答率が高いもしくは同等の結果となった。評価ツールでは、正答率が 0.3 程度のタスクがいくつか見られる。本ツールでは、全タスクにおいて正答率が 0.5 以上である。評価ツールで正答率が低くなっている要因は、表と平行ルコーディネートの対応関係を把握する時に、ビュー間でデータの対応関係を誤まって認識してしまったためだと考えられる。また、評価ツールでの実験中にタスクを達成できないと判断した被験者も見られた。

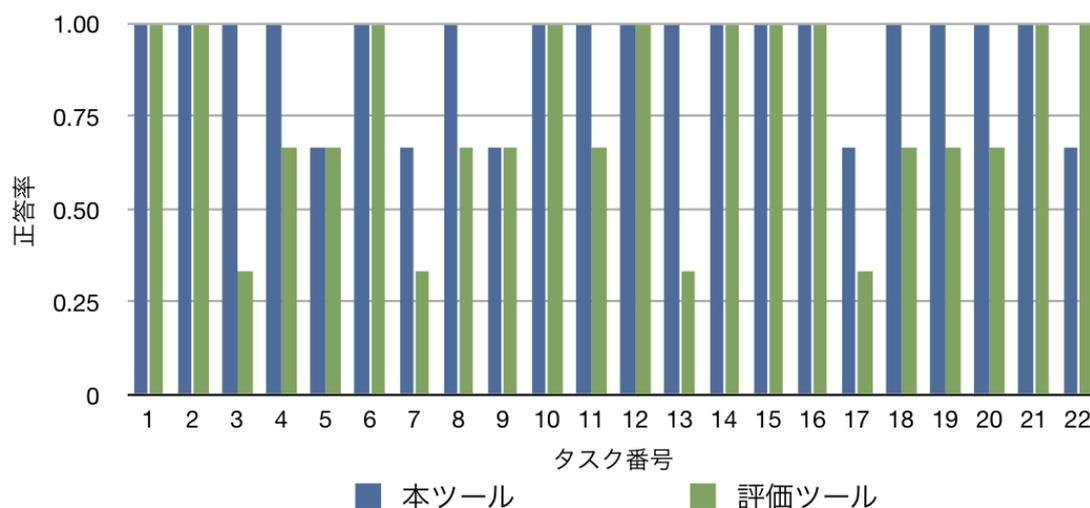


図 6.2: タスクの正答率。

次元数と分析タスクごとの正答率(図 6.3)を調べてみた。分析タスクがラベルの時に、次元数の増加に伴って本ツールと評価ツールの正答率の差が大きくなっていることが分かる。複雑な分析場面で

はビューを見比べ、異なる表現間のデータの対応関係を把握する場面が増える。評価ツールで正答率が低かった原因として、表現を見比べる際のラベルの対応関係の誤認識が考えられる。このことから、本ツールでは視覚的表現に対するラベル付けが効果的であったと考えられる。

分析タスクが分布の時には、評価ツールの正答率が低い結果となった。この原因として、評価ツールではパラレルコーディネートと表の対応関係を正確に把握できなかったことが考えられる。分布を読み解くタスクでは、始めにデータ分布を把握し、それから該当する部分について詳細を閲覧するため、必ず表現を見比べる必要がある。さらに、この結果は、本表現がパラレルコーディネートよりもデータ分布を把握しやすいことを示唆している。これは、セル間の線に太さと透明度を持たせ、データが密集した部分を把握できたためだと考えている。

分析タスクが比較の時には、両ツールがほぼ同じ正答率である。評価ツールでは、色によってデータ分布の比較を行い、本ツールでは、シフト操作を用いることで位置関係からデータ分布を比較を行っていた。今回の実験では、色と位置関係による情報提示について差が見られなかった。しかし、これは今回のタスクで扱ったデータ分布が見比べやすく、線の色だけで比較が行えてしまったためだと考えている。

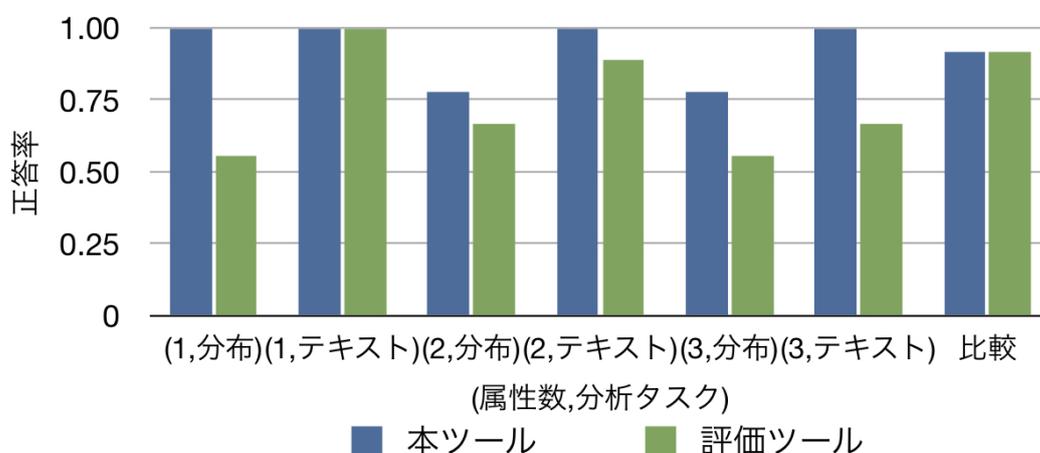


図 6.3: 属性数、分析タスクごとの正答率。

6.5.2 タスク達成時間

図 6.4 に被験者ごとのタスクの達成時間を示す。序盤のタスク (1-5 番) では被験者によって達成時間の差が大きいことが分かる。このことから、ツールの初期段階の習熟度には個人差があることが確認できる。しかし、実験を進めるに従い、徐々に被験者ごとのタスク達成時間の差が少なくなっている。ただ、比較タスク (10,11,21,22 番) では、被験者によって大きく差があることを確認できる。

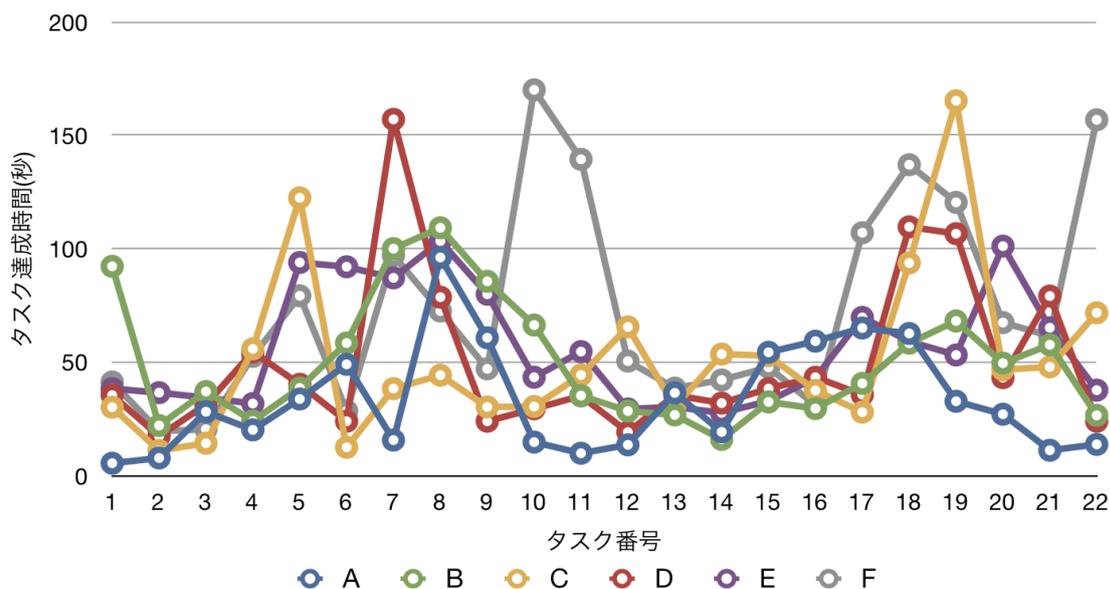


図 6.4: 被験者ごとのタスク達成時間.

各タスクの達成時間を図 6.5 に示す。本ツールでは、実験を進めるに従ってタスク達成時間が短くなっている。タスクを行うことで、本ツールでの分析方法を学習し、速度向上に繋がったと考えられる。評価ツールでは、タスクを追うごとに達成時間が長くなっているタスク(18,19)が見られる。

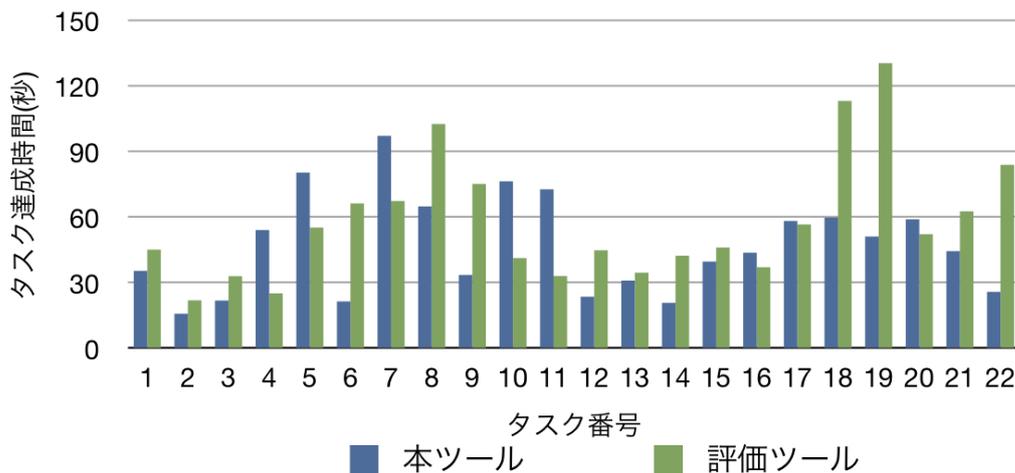


図 6.5: タスク達成時間.

分析する属性数と分析タスクに据とのタスク達成時間を図 6.6 に示す。属性数 2 かつ分析タスクが分布の時以外は、本ツールの方がタスク達成時間が短かった。特に属性数が大きい時に、タスク達成時間の差が大きくなっている。これは、複雑な分析場面では詳細と概観の切り替えが多くなり、表現間の対応関係の把握に時間がかかったためだと考えられる。属性数 2 かつ分析タスクが分布の時に達成時間が逆転した原因として、本ツールでは分布を表するためにアニメーションを用いているため、配置の変更に時間がかかってしまったためだと考えている。

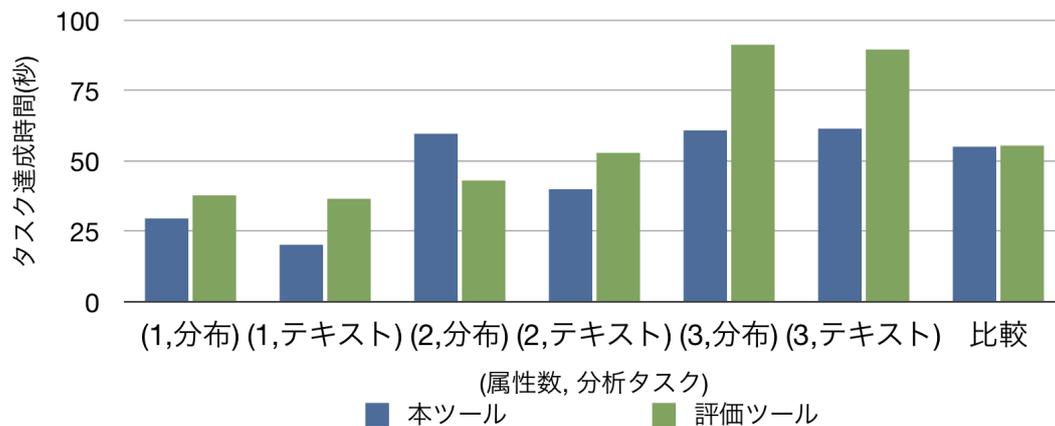


図 6.6: 属性数、分析タスクごとのタスク達成時間。

多次元データ分析のボトルネックとなる活動を特定するために、タスク達成時間とタスク中の操作数及び交差数との相関を求めた。操作は両ツールで共通する操作のことを指し、軸の並び順の変更がこれに当たる。操作とは、軸の並び方を変更する操作を表す。交差とは、評価ツールにおいて操作対象が表と平行コーディネートで切り替わることを指す。操作数及び交差数とタスク達成時間との相関係数を求めた結果を表 6.2 に示す。操作数について本ツールのよりも評価ツールの方が相関が低い。このことから、本ツールではより多くの時間を操作に費やせていたことが分かる。視覚的分析では、インタラクションによって図を変化させながら分析を行うため、時間当たりの操作の多さはより深く分析を行えたことを示す。本ツールのインスタントビュー機能の実行回数よりも評価ツールの交差数の方が相関が強かった。評価ツールでの交差及び本ツールでのインスタントビュー機能は、表と平行コーディネートを切り替えることに当たる。本ツールでは分析時間の多くを表現の切り替えに費やしており、評価ツールでは図を見比べることに時間を費やしていたと考えられる。このことから、表現が滑らかに切り替わることで、ビューの見比べによる分析の複雑さを軽減したことが示された。

表 6.2: タスク達成時間との相関.

	本ツール	評価ツール
操作数 (共通部分)	0.415200280963633	0.329101703240659
インスタントテーブル機能	0.560222846837796	null
交差数	null	0.665395308465565
操作数+交差数	null	0.667786649446345
次元数	0.424832666174342	0.600538009171038

6.5.3 各タスクでのツールの印象

各タスク終了後にアンケートを行ったタスクの行いやすさについての回答結果を示す (6.7). 1が「タスクを行いにくかった」の最も悪い評価, 5が「タスクを行いやすかった」の最も良い評価を表す. 本ツールと評価ツールの回答に対してt検定を行った. 5%の水準で優位差が見られたものは, グラフ間に二股の矢印を付けている. 次元数が3の時に, 本ツールで優位に高い評価が得られた. この結果から, 本ツールがより複雑な分析場面で好まれたことが分かった. 次元数3のタスクでは, 複数の次元について分析する必要がある, 本ツールでは表と平行ルコーディネートが滑らかに切り替わり, 図に対するイメージや知見を保持しつつ継続的に分析を行えたためだと考えている.

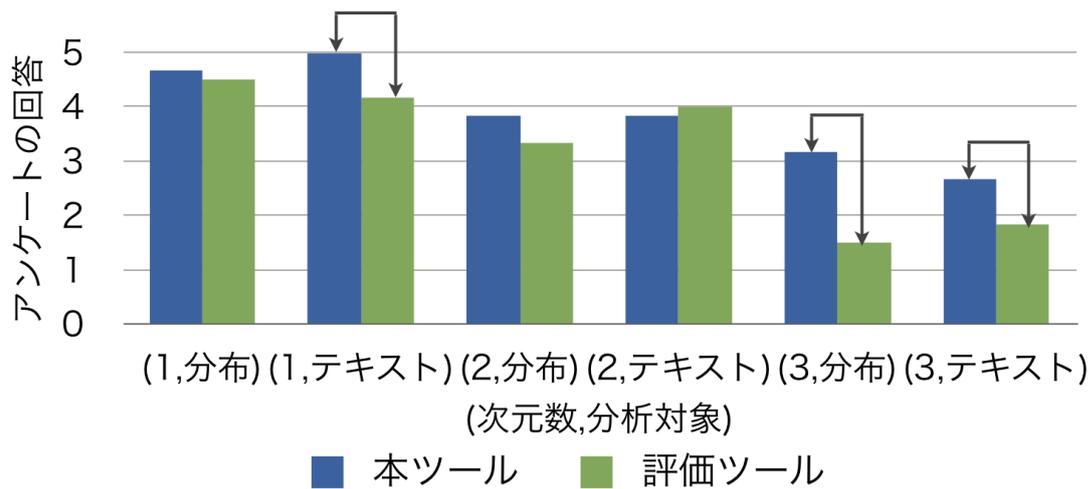


図 6.7: 各タスク終了時に行ったアンケートの回答.

6.5.4 ツールへの印象

タスク後に行った各ツールへの印象についてのアンケートの回答結果を示す。回答結果は1が最も悪く、5が最も良い評価である。図 6.8 に、各個人の回答結果について示す。被験者は A-F のアルファベットで表している。

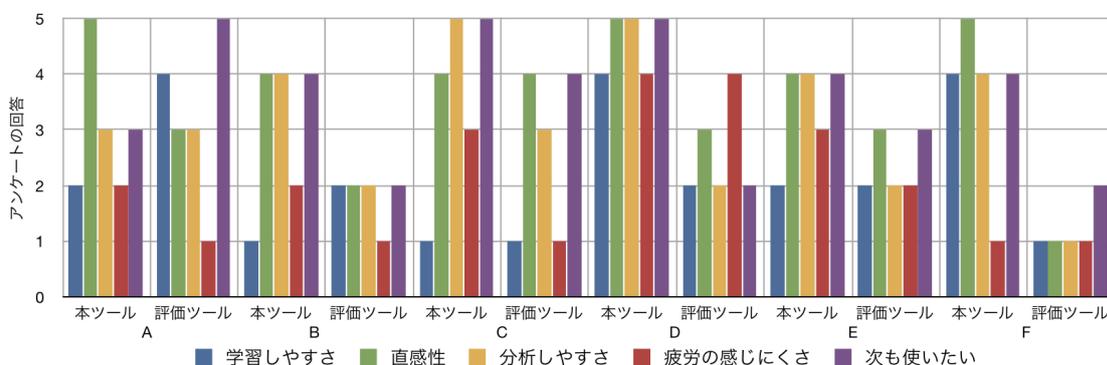


図 6.8: 全タスク終了時に行ったアンケートの回答。

図 6.9 にて、被験者ごとに平均した回答結果を示す。「直感性」、「分析のしやすさ」、「次も使いたい」という項目については、4 を上回る結果が得られた。各ツールに対する回答結果の差の有無を判断するために、項目ごとに t 検定を行った。5% の水準で優位差が見られた項目には、図 6.9 中に二股の矢印がついている。直感性、分析のしやすさ、疲労しにくさについて優位差が見られた。直感性に優位差が見られたことから、データへのイメージとツールが提供する図が似ていることが考えられる。イメージに近い図が提供されることで、より早くかつ深くデータを分析することができる。この要因として、本手法ではデータ要素が常に統一して表現されており、図の変化前後の対応関係を把握しやすいためだと考えている。分析しやすさに優位差が見られたことから、データ分析に必要な操作が適切に用意できていたと考えられる。データの特定や分布の把握の操作を実行でき、得たい情報が適切に表現されていたと考えられる。疲労しにくさに優位差が見られたことから、本ツールでは疲労を感じにくく、複雑なデータの分析を行いやすいと考えられる。この要因は、異なる表現の図を見比べることによって疲労を感じているためだと考えられる。評価用ツールでは、タスク中にパラレルコーディネートと表を見比べることが多く、分析中に表現の異なる図の切り替えが頻繁に行われていた。本ツールでは、アニメーションとドラッグ操作によって滑らかに図が切り替わっていく。学習コストに優位差が見られなかった要因として、既存の表現(表とパラレルコーディネート)を提示するツールと学習コストにあまり差がないことが言える。「次も使いたい」について優位差が見られなかった原因は、ツールの具体的な利用シーンを思い浮かべられなかったためだと考えている。

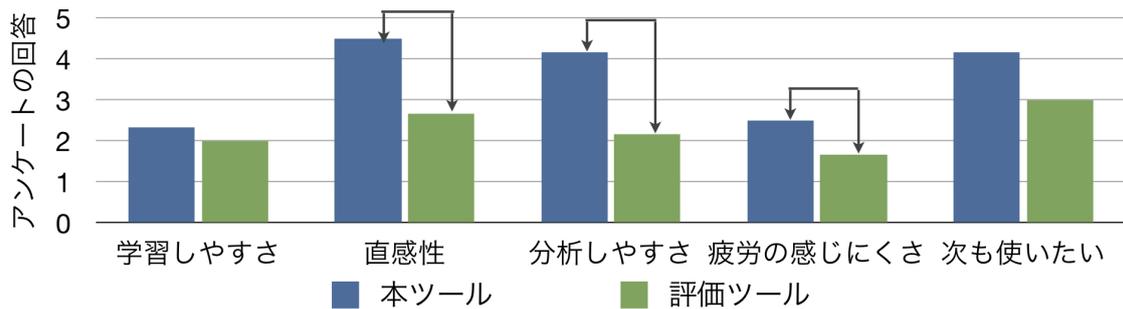


図 6.9: 属性数, 分析タスクごとのアンケートの回答.

6.6 実験のまとめ

実験結果から, 仮説に関して以下のことが示された. 仮説 1 について, 実験後のアンケートの回答から本ツールでは疲労を感じにくいことが示され, 1つのビューを用いた分析が疲労を感じにくくさせることを確認できた. 仮説 2 について, タスク正答率から全タスクにおいて本ツールでのタスク正答率の方が高かったことが示され, 仮説とは逆に 1つのビューを用いた分析の方が情報を正確に読み取れることが確認できた. この結果は, 複数のビューを用いた分析でビューを見比べる際に, 対応関係を誤って認識してしまうためだと考えられる. 仮説 3 について, タスク達成時間から, 属性数が 2 かつ分析タスクが分布の時以外は, 本ツールの方がタスクを素早く達成できたことが示され, 1つのビューを用いた分析の方が素早く分析を行えることが確認できた.

実験結果から示された本ツールの利点を以下にまとめる.

- 情報を正確に読み取れる.
- 複雑な分析場面において, すばやく分析を行える.
- あまり疲労を感じさせない.
- 複雑な分析に向いている.

タスクの正答率が高かったことから, 情報を正確に読み取ることができたと言える. この要因として, データ要素に同一の表現を用いることで, 変化前後のデータの対応関係を正確に把握できたためだと考えられる. タスク達成時間から, 本ツールではすばやく分析を行えることが分かった. アニメーションによる時間経過に関わらず, タスクの達成時間が短かった. しかし, アニメーションによる分析への効果を示す研究もあり, 評価ツールの軸の並び替え時にアニメーションを用いたことによる影響の可能性があり, 更なる評価が必要である.

タスク後のアンケートへの回答結果から, 本ツールでは疲労を感じにくいことが示された. 疲労を感じにくいことの要因として, 異なる表現を見比べることにより疲労を感じやすく, 表現が滑らかに

変化することが疲労感の軽減に繋がったためと考えられる。次元数が多い(複雑な分析)の時に、良い主観評価が得られたことから、複雑な分析に向いていると考えられる。

分析タスクが切り替わった時に図を少し前の状態に戻したいというコメントがあった。ツールの利便性を高めるために、アンドゥ・リドゥ機能を備える必要がある。また、本手法では、データ分布を位置を用いて表現しているため、軸の表現している情報を把握しなければならず、量的情報を直感的には読み取りにくい。量的情報の場合にセルのサイズにマッピングすることも考えられるが、現在の属性ごとにセルの横幅を一定にする方法でなければ、表としての機能が失われてしまうかもしれない。

第7章 関連研究

7.1 多次元データの視覚的分析ツール

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。表計算ソフト Excel¹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。Guoら [6][7]は、位置情報を持つ多次元データについて、様々な表現のパネルを組み合わせることでデータ内のクラスターと詳細情報の把握を支援している。SellTrend[8]では、概観から発見した特徴をさらに詳しく調べるために、統計グラフの任意の部分を選択しその部分の全体像を提示する。これらの多次元データ分析ツールは、複数のビューを備え、着目する情報によって適したビューを用いて分析が行われる。Temporal Summaries[9]は、時系列データについてデータ数に着目した分析を支援している。本研究では、詳細情報としてデータ数ではなく、多次元データのデータ要素に着目している点で異なる。

7.1.1 概観に基づく分析ツール

データの概観に基づいて分析を行う視覚的分析ツール Tableau²や Polaris[10]が開発されている。これらの分析ツールでは、概観に基づく特徴だけを用いて分析を行い、最終的に得られた結果について詳細情報を閲覧する。しかし、実際の分析場面では、詳細情報を閲覧したい場面が多く、異なるビューで提示される詳細情報を閲覧することで分析が中断されてしまう恐れがある。Heerら [11]は、統計グラフをアニメーションを用いて滑らかに変化させることで、様々な側面からのデータの把握を可能にした。Rolling the Dice[12]では、散布図の次元の切り替えをアニメーションを用いて滑らかに行えるようにし、図に対するイメージや知見を保持しながら多角的に分析を行うことが可能になった。The FlowVizMenu[13]では、パラレルコーディネートと散布図が滑らかに切り替わることで、任意の2次元のデータ分布と相関を同時に把握できる。これらの研究では、概観の提示方法を滑らかに切り替えることで、図に対するイメージや知見を保ちながら、多角的に分析を行うことができる。本研究では、概観と詳細を滑らかに切り替えることができ、実際の分析場面で必要となるデータの詳細を把握することができる。

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

²Tableau <http://www.tableausoftware.com/>

7.1.2 詳細と概観に基づく分析ツール

詳細と概観に基づく多次元データ分析を支援するツールが開発されている。FromDaDy[14]は、選択操作とドラッグ操作によって、データの概観から詳細を抜き出して、概観の関係を保持したまま詳細情報を把握することができる。本手法では、データの概観の中に詳細を埋め込むことができ、双方を同時に把握することができる。GeneShelf[15]は、折れ線グラフと棒グラフを組み合わせることで、量的情報と時系列情報を同時に把握できるようにした。本研究では、詳細と概観が切り替わった際もデータ要素に統一した表現が用いられ、データ要素の対応関係をより単純に把握できる。また、詳細と概観を同時に提示する手法としてFocus+Context[16]が研究されている。Focus+Contextとは、データの一部についての詳細情報(Focus)とデータ全体における位置付け(Context)を同時に提示する手法である。SignalLens[17]では、電気信号の波形に対するFocus+Context手法を開発した。Joachimら[18]は、概観と詳細を滑らかに切り替えるビューの変形方法を開発し、図の全体像を維持しながら任意の部分拡大することを可能にした。これらの研究では詳細を視覚的表現の最小単位として捉えており、多次元データにおける最小単位のデータ要素及びその関係には対応していない。

7.2 多次元データの表現方法

多次元データの特徴の直感的な把握を目指し、パラレルコーディネートや表を拡張した表現手法が開発されている。また、それ以外の多次元データの表現手法も開発されており、多次元データのレコード内のグループを最小単位として表現するMatchMaker[19]がある。多次元データのデータ要素を色付きのセルとして表現し、レコードの一部に当たる矩形を連結させて並べ、さらに同じレコードである矩形同士を曲線で繋ぐ。セルの色と線の繋がり方から、多次元データ内のクラスタ間の関係を直感的に把握することができる。Diversity Map[20]は、各属性のデータ範囲をセルとして表現し、次元内でのデータ数に対応させてセルに色付けを行い、データ数の分布の把握を支援した。本研究では、多次元データの詳細であるデータ要素に着目し、詳細を常に提示しておくことが可能である。

7.2.1 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多次元データの視覚的分析ツールが開発されている。Parallel sets[21]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の色分けを行い、多次元データの概観の把握を支援している。Pargnostics[22]は、パラレルコーディネートの可読性を向上させるために、軸の並び順と軸の逆転を最適化する手法を開発した。Zhouら[23]は、データ内のクラスタについてパラレルコーディネートの軸間の線に曲線を用い、同じクラスタの線を束ねることでクラスタを表現した。Ellisら[24]は、パラレルコーディネートの閲覧を支援するために、線の色によるフィルタを開発した。

また、パラレルコーディネートと既存の表現手法を組み合わせた視覚的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[25]。Parallel Tag Clouds[26]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

表現は、表と平行コーディネートを組み合わせており、表の特徴を合わせ持つことから詳細な分析を行うことに適している。さらに、本表現は平行コーディネートの特徴も持っており、これらの手法を適応することも可能である。

7.2.2 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。TableLens[27]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[28]。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[29]。本研究では、表と平行コーディネートを組み合わせることで、平行コーディネートの特徴である属性間の相関と、表の特徴である詳細情報を同時に把握できる。

第8章 結論

本研究では、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を滑らかに切り替え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。アニメーションと2表現の中間表現を用いて、表現を滑らかに切り替えられることで、1つのビュー上でデータの概観と詳細の自由な行き来を可能にした。これにより、概観及び詳細に基づく手がかりを残したまま、さらに分析を進めることができる。

評価として、表とパラレルコーディネートをそれぞれ別のビューとして備える評価ツールとの比較実験を行った。実験結果から、本ツールでは素早く正確に分析を行うことができることが示された。さらに、アンケートの回答から、表現が滑らかに切り替わる本ツールでは疲労を感じにくいとの結果が得られた。これは、表現が滑らかに切り替わり、異なる表現間のデータの対応関係を追うことなく分析を行えたためだと考えられる。

本研究により、詳細と概観を自由に行き来しながら、多次元データを分析することが可能になった。これにより、図に対するイメージや知見を保持しながら、1つのビュー上での継続的にデータを分析することができる。

謝辞

本研究を行うに当たり、指導教員である三末和男准教授には丁寧な指導と助言を頂き、心から感謝しております。時に厳しい言葉をかけられることもありましたが、その言葉をバネにより良い研究を目指し、熱意を持って研究に取り組むことができました。無事に論文を執筆できたのは先生のご指導のおかげです。本当に有り難うございました。

田中二郎教授には、研究に必要な設備を提供して頂き、快適な研究室生活を送ることができました。また、研究を行うに当たっての心構えや効果的なプレゼンテーションの仕方などの研究を行うための基礎をご教授頂きました。しっかりと地に足をつけて学生生活を送ることができました。本当に有り難うございました。

高橋伸准教授には、ゼミでの発表の場において有意義な意見を頂き、研究を円滑に進めることができました。志築文太郎講師には、ゼミでの発表の場はもちろんのこと、学外発表の機会に私の発表で座長を務めていただき、大変お世話になりました。落ち着いて発表を行うことができ、成功を取めることができました。本当に有り難うございました。

インタラクティブプログラミング研究室の皆様には公私共に大変お世話になりました。迷惑をおかけしてしまうことも多々あったかと思いますが、私のことを暖かい目で見守って頂き、時に手を差し伸べて頂きました。特に NAIS チームの皆様には、ゼミでの議論だけでなく、同じ道を歩む仲間として苦楽を共にし互いに切磋琢磨しながら、一生に残る深い思い出を作ることができました。心から感謝しております。

また、学生生活を送るに当たり、多大な援助をして頂いた家族に深く感謝しております。家族の支えなしには、この充実した学生生活を送ることはできなかったと思います。最後に、学生生活でお世話になった全ての方々に深く感謝しております。本当に有り難うございました。

参考文献

- [1] Bertin J, "Semiology of Graphics." University of Wisconsin Press, 1983.
- [2] Inselberg Alfred, Dimsdale Bernard, "Parallel coordinates: a tool for visualizing multi-dimensional geometry." Proceedings of the 1st conference on Visualization, pp.361-378, 1990.
- [3] Daniel A. Keim, "Information visualization and visual data mining." IEEE Transactions on Visualization and Computer Graphics, Vol. 8, No. 1, pp. 1-8, .2002.
- [4] S. Eick, G. Wills, "High interaction graphics." European Journal of Operational Research, Vol. 81, No. 3, pp. 445-459, 2000.
- [5] Allen Martin, Matthew Ward, "High dimensional brushing for interactive exploration of multivariate data." In Proceedings of the 6th conference on Visualization, pp. 271-278, 1995.
- [6] Diansheng Guo, "Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data." Visualization and Computer Graphics, IEEE Transactions on, vol.15. no.6, pp.1041-1048, 2009.
- [7] Diansheng Guo, Jin Chen, Alan M, MacEachren, Ke Liao, "A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP)." IEEE Transactions on Visualization and Computer Graphics, pp. 1461-1474, 2006.
- [8] Zhicheng Liu, John Stasko, Timothy Sullivan, "SellTrend: Inter-Attribute Visual Analysis of Temporal Transaction Data." IEEE Transactions on Visualization and Computer Graphics, pp. 1025-1032, 2009.
- [9] Taowei David Wang, Catherine Plaisant, Ben Shneiderman, Neil Spring, David Roseman, Greg Marchand, Vikramjit Mukherjee, Mark Smith, "Temporal Summaries: Supporting Temporal Categorical Searching, Aggregation and Comparison." IEEE Transactions on Visualization and Computer Graphics, pp. 1049-1056, 2009.
- [10] Chris Stolte, Diane Tang, Pat Hanrahan, "Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases." IEEE Transactions on Visualization and Computer Graphics, pp. 52-65, 2002.
- [11] Jeffrey Heer, George Robertson, "Animated Transitions in Statistical Data Graphics." IEEE Transactions on Visualization and Computer Graphics, pp. 1240-1247, 2007.
- [12] Niklas Elmqvist, Pierre Dragicevic, Jean-Daniel Fekete, "Rolling the Dice: Multidimensional Visual Exploration using Scatterplot Matrix Navigation." IEEE Transactions on Visualization and Computer Graphics, pp. 1141-1148, 2008.

- [13] Christophe Viau, Michael J. McGuffin, Yves Chiricota, Igor Jurisica, "The FlowVizMenu and Parallel Scatterplot Matrix: Hybrid Multidimensional Visualizations for Network Exploration." *IEEE Transactions on Visualization and Computer Graphics*, pp. 1100-1108, 2010.
- [14] Christophe Hurter, Benjamin Tissoires, Stephane Conversy, "FromDaDy: Spreading Aircraft Trajectories Across Views to Support Iterative Queries." *IEEE Transactions on Visualization and Computer Graphics*, pp. 1017-1024, 2009.
- [15] Bohyoung Kim, Bongshin Lee, Susan Knoblach, Eric Hoffman, Jinwook Seo, "GeneShelf: A Web-based Visual Interface for Large Gene Expression Time-Series Data Repositories." *IEEE Transactions on Visualization and Computer Graphics*, pp. 905-912, 2009.
- [16] Card S.K, Mackinlay J.D, Shneiderman B, "Readings in Information Visualization: Using Vision to Think." Morgan Kaufmann Publishers, San Francisco, California, pp. 1-34, 1999.
- [17] Robert Kincaid, "SignalLens: Focus+Context Applied to Electronic Time Series." *IEEE Transactions on Visualization and Computer Graphics*, pp. 900-907, 2010.
- [18] Joachim Bottger, Michael Balzer, Oliver Deussen, "Complex Logarithmic Views for Small Details in Large Contexts." *IEEE Transactions on Visualization and Computer Graphics*, pp. 845-852, 2006.
- [19] Alexander Lex, Marc Streit, Christian Partl, Dieter Schmalstieg, "Comparative Analysis of Multidimensional, Quantitative Data." *IEEE Transactions on Visualization and Computer Graphics*, pp. 1027-1035, 2010.
- [20] Tuan Pham, Rob Hess, Crystal Ju, Eugene Zhang, Ronald Metoyer, "Visualization of Diversity in Large Multivariate Data Sets." *IEEE Transactions on Visualization and Computer Graphics*, pp. 1053-1062, 2010.
- [21] Bendix F, Kosara R, Hauser H, "Parallel sets: visual analysis of categorical data." In *Proceedings of the IEEE Symposium on Information Visualization*, pp. 133- 140, 2005.
- [22] Aritra Dasgupta, Robert Kosara, "Pargnostics: Screen-Space Metrics for Parallel Coordinates." *IEEE Transactions on Visualization and Computer Graphics*, pp. 1017-1026, 2010.
- [23] H. Zhou, X. Yuan, H. Qu, W. Cui, B. Chen, "Visual clustering in parallel coordinates." *Computer Graphics Forum*, pp.1047-1054, 2008.
- [24] G. Ellis, A. Dix. "Enabling automatic clutter reduction in parallel coordinate plots." *IEEE Transactions on Visualization and Computer Graphics*, pp.717-724, 2006.
- [25] 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, 2009.
- [26] 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, 2009.

- [27] Rao R, Card S.K, "The table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information." In Proceedings of the SIGCHI Conference on Human Factors in Computing, pp.318-322, 1994.
- [28] John Mathias, Tominski, Christian, Schumann Heidrun. "Visual and analytical extensions for the table lens." Visualization and Data Analysis 2008. pp. 1-12, 2008.
- [29] A. Telea, " Combining Extended Table Lens and Treemap Techniques for Visualizing Tabular Data. " in Proceedings of Joint Eurographics - IEEE VGTC Symposium on Visualization, pp.1-9, 2006.