

データベースのビジュアルな 検索と分析 (OLAP)

■ 経営分析とデータビジュアル化技術の融合 ——

近年のますます多様化、複雑化するビジネス市場で競争優位を確保するために、データウェアハウスを核にした、情報の戦略的有効活用が重要視されている。

仮説検証型のOLAPや仮説創造型のデータマイニングは、このような情報活用ニーズに解を与えるものとして導入が進みつつあるが、一方で分析スキルの不足が壁になっていることも事実である。

そのため、データをさまざまな切り口で要約することによって、情報を適確かつ容易に把握し、新しい発想を促すことのできるデータビジュアライゼーションの技術が注目され始めている。

本稿では、データビジュアライゼーションの技術の1つである、平行座標 (Parallel Coordinates) に着目し、多次元分析における有用性を説明する。

(株) 富士通ソフトウェア生産技術研究所第一開発部
加藤 博己

■ OLAPとデータビジュアライゼーション ——

企業活動には、いわゆるPDCA (Plan, Do, Check, Action) と呼ばれるビジネスサイクルがある。まず、企業や部門のビジネスゴールに向けた戦略を策定 (Plan) し、日々の業務の中で絶えず施行 (Do)、検証 (Check) を繰り返すのである。

近年の市場や社会の変化は、さらに早く、複雑になってきている。苦勞して策定したきわめて効果的な販売戦略もすぐに他社に追従されてしまうのである。このような状況の中で、近年、現場でのタイムリな分析的活用が重要視され始めている。

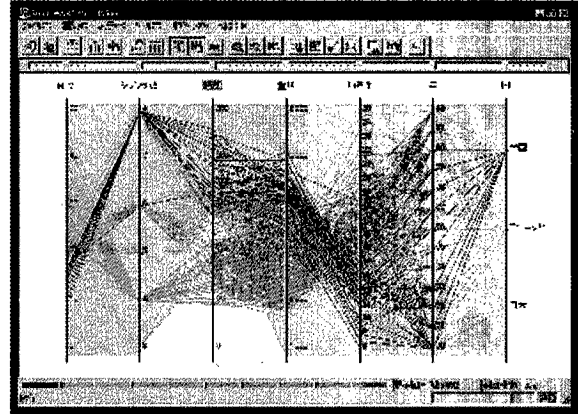
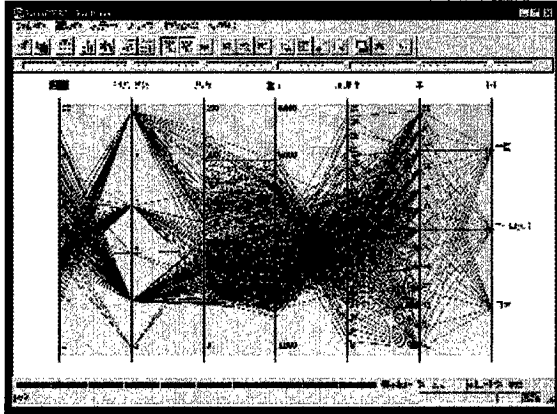
OLAPツールは、その容易な操作性と運用性で、現場での情報の管理/分析ツールとして定着し始めている。OLAPは、一般には多次元表の形で、データを種々の切り口で加工/整理しながら情報を検証するための道具である¹⁾。OLAPのドリリングやスライス&ダイスといった操作は、人間の仮説検証型の思考パターンに馴染みやすい。OLAPは現場の業務 (ワークフロー) に密着した帳票イメージで情報を整理し、分析的活用を実現

するものである。そのため、現場での非専門家にとって受け入れやすい分析ツールであろう。

しかし、一方、「人」「モノ」「金」が流れるビジネスの世界では、通常、100次元の管理すべき切り口があるといわれている。年度別/期別/月別といった時間の切り口、店舗別/支店別/地域別といった場所の切り口など、非常に多くの切り口のバリエーションでビジネスが管理される。したがって、日々の業務の中では、より多くの切り口を効率的にレイアウトした、複雑な帳票が一般的に用いられる。この複雑な帳票から問題点を適切に把握し、ビジネスの意思決定に結び付けるには、経験と熟練が不可欠である³⁾。

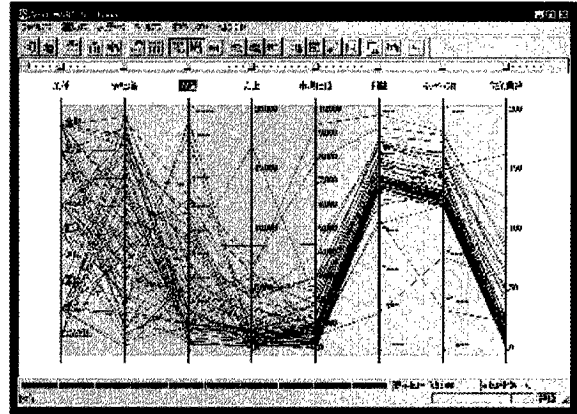
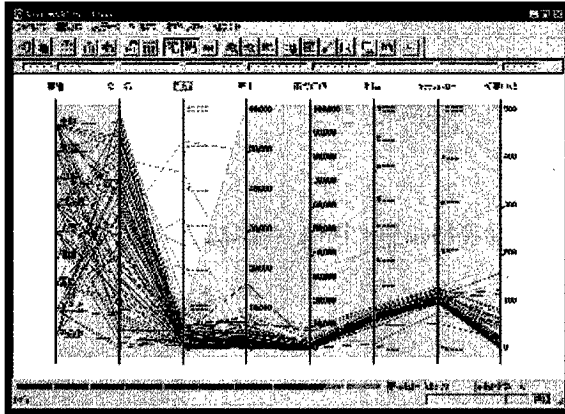
このようなスキルやリテラシに依存する部分をデータビジュアル化技術によって補完しようとする試みも数多くなされている。データビジュアライゼーションは誰もが直感的に情報を把握でき、より効果的な情報活用ができる点が期待されている²⁾。

平行座標 (Parallel Coordinates) 表現のデータビジュアライズ技術は、多次元の情報一度に表現し、簡単な操作機能を実装することによって、直感的なデータ検索や傾向把握を可能にするものである。さらに、科学的な



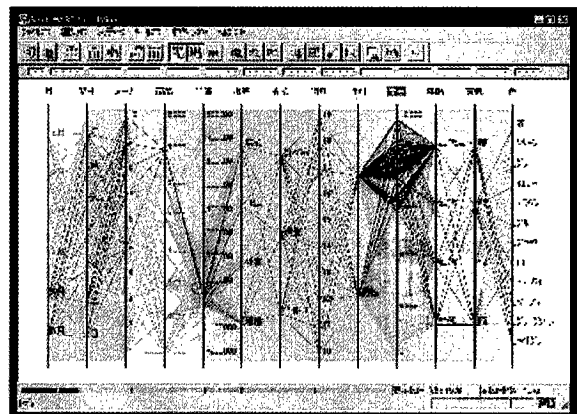
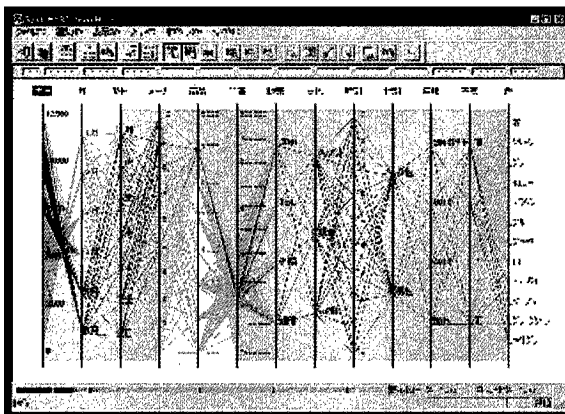
馬力の高い部分にフィルタリングすると、その車の特徴が浮き彫りになります。

図-1 フィルタリング



資本金の低い部分でズームすると、密集していた低資本金のデータが拡大されます。

図-2 ズーム



スペクトルバーで購入顧客のプロフィールをマグニファイすると、傾向が浮き彫りになります。

図-3 マグニファイ

分析手法に共通する分析過程のブラックボックス化がなく、分析者の専門的スキルを必要としない。これらの特徴はまさに、現場での情報活用を効果的に実現できる道具としてきわめて有望である。

■ 平行座標とは ——

平行座標 (Parallel Coordinates) とは、複数の切り口 (一般に軸あるいは次元と呼ばれる) を平行に並べ、各データを連続したひとつながりの折れ線で表現したグラ

フィカル表現である。平行座標では、一度に複数の次元の情報を表現することができ、データ全体の概観を直感的に捉えることができる。

平行座標上では、OLAPによる多次元分析と同様に、いくつかの分析手法を適用することができる。主な分析手法として、以下の3つがある。

1) フィルタリング

軸ごとに着目すべき範囲を限定する操作である。これによって、特定の軸に対する範囲限定を所与の条件として、不特定任意の軸間の関連性の強さを視覚的に捉えることができる。図-1にフィルタリングの操作例を示す。

2) ズーミング

特定のデータの情報を拡大する操作である。これによって、データが集中している部分だけを拡大して分析することや特異データを除去して分析することが容易になる。ズーミングは、分析者にとっての“ノイズ”を除去し、情報の浄化（クレンジング）を可能にする。図-2にズーミングの操作例を示す。

3) マグニファイ

一般的に、ビジネス分野におけるデータは多くのカテゴリによって、元々要約されている場合が多い。そのため、視覚的な情報量が小さく、特徴的な情報を得ることができない。

マグニファイ操作は、粒度の高い任意の軸によって、カテゴリカルな軸の視覚的な情報量を大きくすることによって、隠れた情報をひきだすことができる。図-3にマグニファイの操作例を示す。

■ ビジュアライジングOLAP

ここでは、ビジュアルに表現した多次元データに対し、種々の操作を行うことによって、傾向や特性を直感的に把握できるビジュアライジングOLAPについて説明する。

一般に、データに埋もれた問題点を明確にしたり、その原因を探るためには、広い視野から問題点を概観し、注目すべき点を探したりすることが必要となる。多くの場合、ここで人の発想や洞察が重要になる。発想や洞察は、直感的な理解と連想から生まれる。したがって、多次元に渡る情報全体を直感的に把握し、着目すべき点を即座に掘り下げたり、別の着目点に即座に移動したりする、いわゆる、フライスルーが可能なビジュアルな操作イメージが必要となる。

平行座標によるビジュアル表現を用い、かつ、フィルタリング、ズーミング、マグニファイ操作を組み合わせることによって、上記の要件を十分に満足するビジュアライジングOLAPを実現することができる。

ビジュアライジングOLAPには、2つのデータビジュアライジング方法がある。1つは、いくつかの切り口で集約されたセグメントごとにビジュアライズするもので、セグメントビジュアライゼーションと呼ぶ。2つ目は、生のデータをそのままビジュアライズするディテールビジュアライゼーションである。

セグメントビジュアライゼーションでは、セグメントごとのパターンの類似性や特徴を認識するのに役立つ。また、ディテールビジュアライゼーションは、生のデータの集合（線の集中度）によって、軸間の相関や特異データを探るのに役立つ。図-4にセグメントビジュアライゼーションの例を、図-5にディテールビジュアライゼーションの例を示す。図-4は、地域と顧客層を限定すると、購入商品のパターンが見えてくることを示す。ここでは、都心部のミドル男性の商品購入パターンが分かる。また、図-5は、5月～6月の売れ筋商品と、その購買層が強調されることを示す。この購買層をフィルタリングして、外部のファイルに抽出することにより、さらに、ダイレクトメールなどの施策に即座に活用することができる。

ビジュアライジングOLAPは、他の統計解析やデータマイニングによって得られた解を軸の形で連動することによって、より幅広い分析が可能になる。たとえば、クラスタ分析と連動して、新たに生成された分類軸を平行座標に加えることによって、各グループの特性を直感的に捉えることができる。

図-6には、保険保証金額、払い込み金額、解約月、解約期間、および契約者の年齢の5つのデータにより、保険解約顧客をクラスタ分類した結果を示す。ここでは各クラスタの特性が色によって直感的に把握することができる。たとえば、4クラスタに分類した軸上で色分けを行うと、第1クラスタ（青）は契約月数が短く、解約月が上半期に集中している顧客層、第2クラスタ（水色）は保険保証金額、払い込み金額の大きな顧客層、第3クラスタは契約期間の長い顧客層であることが分かる。

また、多変量解析の主成分（因子）分析と連動して、新たに生成された主成分（因子）軸でマグニファイすることで、データの直感的なセグメンテーションを行うことが可能になる。

図-7では、車のスペックデータ（重量、加速性、燃費、馬力など）の第1主成分を新たな軸として、重量と加速性の間にマグニファイしている。この結果、主成分値2の近辺で顕著に弁別されていることが分かる。ここで、主成分値2以上のデータだけをフィルタリングすることにより、高重量、高馬力のいわゆる高級車を表すデータが弁別される。

このように、ビジュアライジングOLAPでは、他の種々の手法によって生成された新しい情報を付加的に表示することにより、情報の表現力を高め、人の発想を支

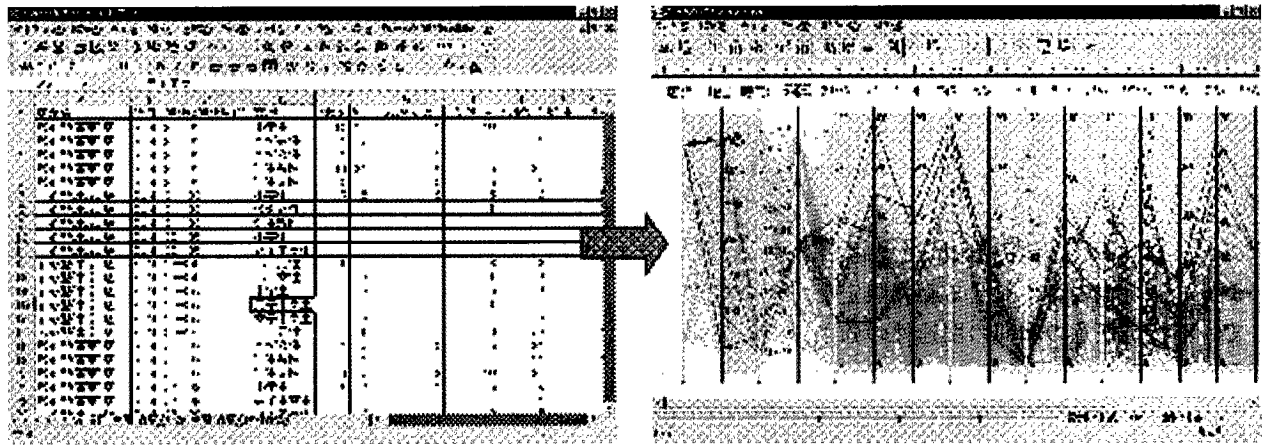


図-4 セグメントビジュアライゼーション

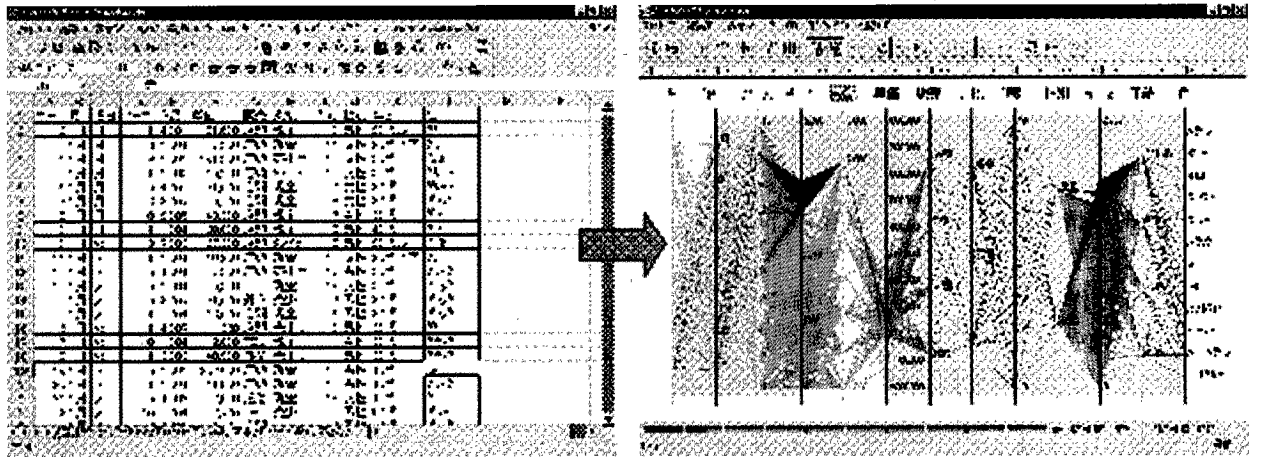


図-5 デイテールビジュアライゼーション

援する新しいGUIとして柔軟に拡張していくことが可能である。

■ ビジュアライジングOLAPの適用例 —

ビジュアライジングOLAPの適用分野として以下がある。

- 顧客購買特性の把握
複数の切り口から、売れ筋商品の購買層や関連購買商品を直感的に把握する。
- 営業力増大に向けた営業員特性の把握
複数の切り口から、営業員の販売傾向、得意分野を直感的に把握する。
- 品質管理における関連要因の把握
工程ごとの品質状況と特性要因の間の関連性を直感

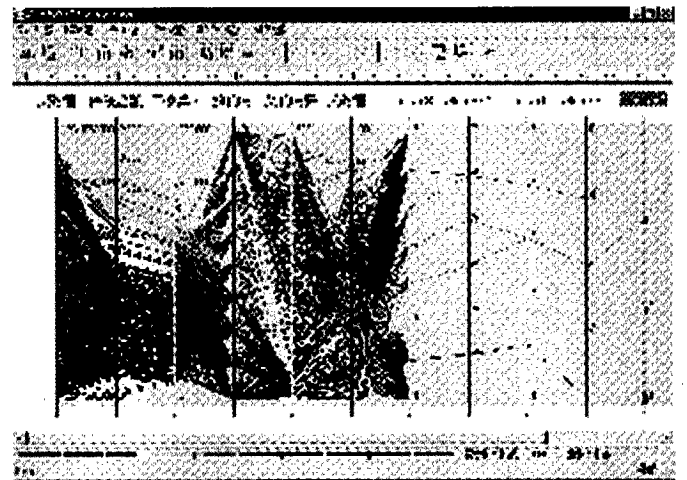


図-6 クラスタ分析結果による特性把握

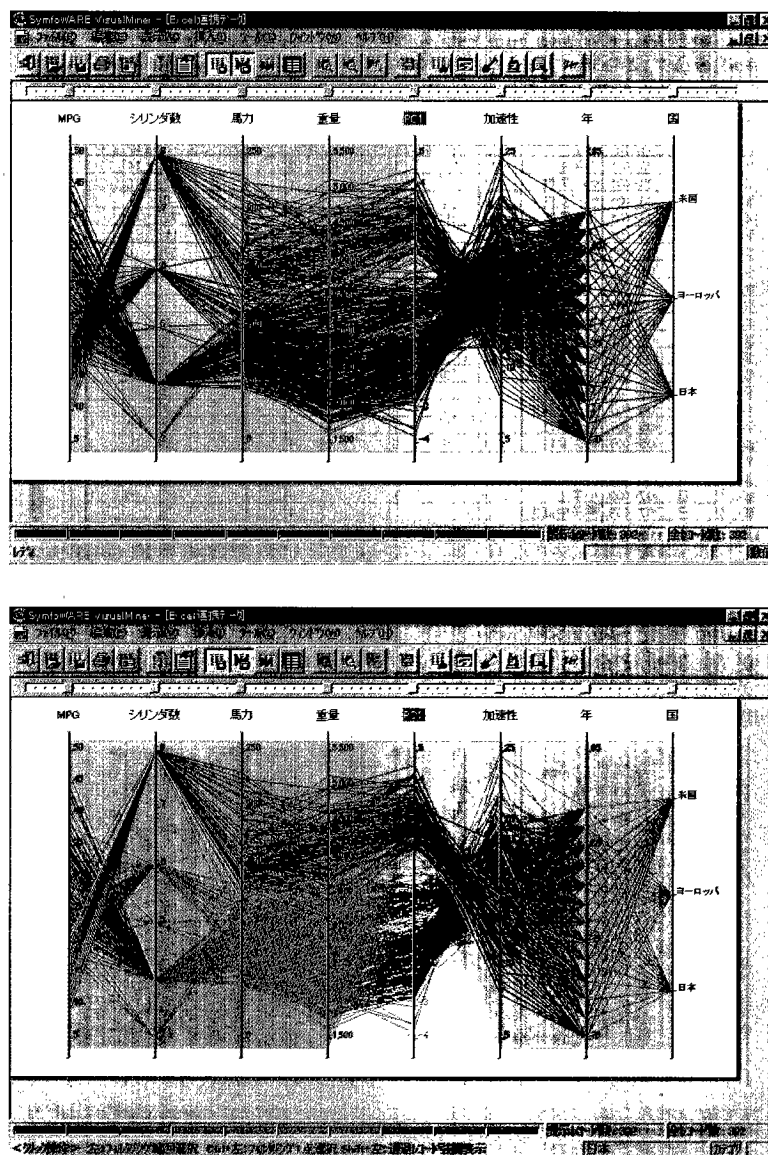


図-7 主成分軸によるデータの弁別

的に把握する。

■ 購買特性分析例

販売明細データを顧客層別に集計したデータのセグメントビジュアライゼーションを用いて、購買特性分析を行ってみる。

全体の傾向をみて、着目したい部分を絞り込む(図-8)。たとえば、「菓子」の売上げの高い部分に着目してみる。その結果、選択範囲内のデータだけが強調表示される。ミドルの男性/女性が多いという傾向が分かる(図-9)。

さらに、「東京都心部」に着目してみる。その結果、都心部ではミドル男性が、菓子やアイスクリームをよく買っている傾向が直感的に分かる。また同時に、買

う時間帯から「残業食」であることも想像がつく(当然、ビールは売れていない)(図-10)。

このように、ビジュアライジングOLAPを使用すると、たくさんの切り口からなる複雑なデータから、着目すべき顧客セグメントを認識したり、売上げの傾向や特性などを容易に把握することができ、販促施策のための商品レイアウトの検討に利用することができる。

参考文献

- 1) Inmon, W. H.: Building the Data Warehouse, Wiley-QED, Chapt7 EIS and Data Warehouse (1995).
- 2) Dilly, R. (The Queen's Univ. of Belfast) : Data Mining - An Introduction, Chapt2 Data Mining Functions, Chapt3 Data Mining Techniques (1995).
- 3) 加藤他: 戦略的データマイニングシステム, FUJITSU.

店名	曜日	時間帯	客層	スナック	菓子	嗜好品	アイス cream	弁当	ビール	びん・缶飲料	書籍	衣料
1	都心型	日	23-01	小学生	13	15	1	10	8	63	14	
2	都心型	日	23-01	ヤング男性	1	6			1	22	2	
3	都心型	日	23-01	ヤング女性	1	3	1			8		
4	都心型	日	23-01	シル男性	14	21	6	25	28	72	11	
5	都心型	日	23-01	シル女性	3	2	5	6	1	11	2	
6	都心型	日	01-03	小学生	7	15	2	15	12	57	6	2
7	都心型	日	01-03	ヤング女性	1			1	1	2		
8	都心型	日	01-03	シル男性							4	
9	都心型	日	03-05	小学生								1
10	都心型	日	03-05	女子中高生						1		
11	都心型	日	03-05	ヤング男性	5	5	2	8	22	16	12	
12	都心型	日	03-05	ヤング女性	2	3		1	9	13	2	1
13	都心型	日	05-07	小学生		3	1	9	9	4	1	
14	都心型	日	05-07	男子中高生			1	1	4	3	2	
15	都心型	日	05-07	女子中高生		3			7	2		
16	都心型	日	05-07	ヤング男性	1	6	3	6	27	27	9	1
17	都心型	日	05-07	ヤング女性	1	6	3	6	24	26	10	3
18	都心型	日	05-07	シル男性		1			11	3	1	
19	都心型	日	05-07	シル女性	4	6	2	10	25	18	7	
20	都心型	日	07-09	小学生	6	7	1	5	9	15	6	1
21	都心型	日	07-09	男子中高生					4			
22	都心型	日	07-09	女子中高生					5			

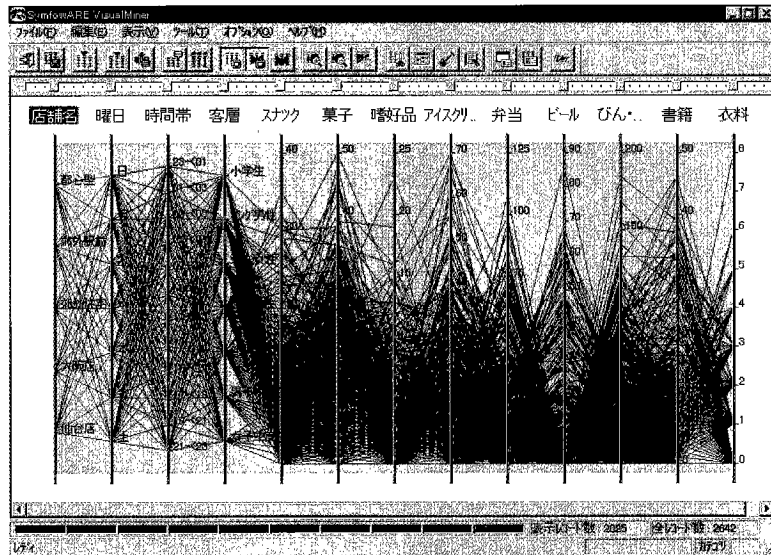


図-8 客層別商品別販売実績とビジュアライズ

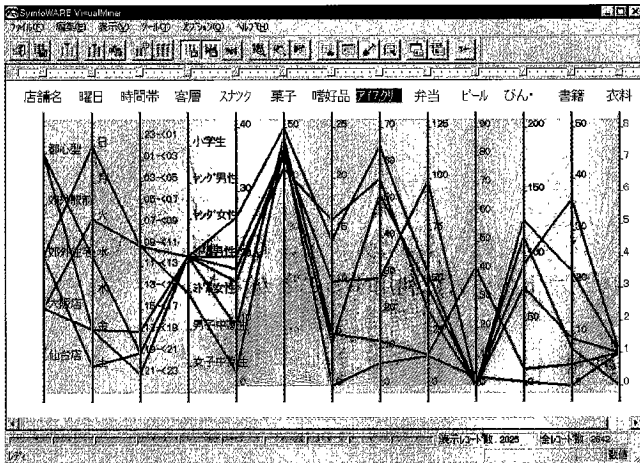


図-9 特定商品でフィルタリング

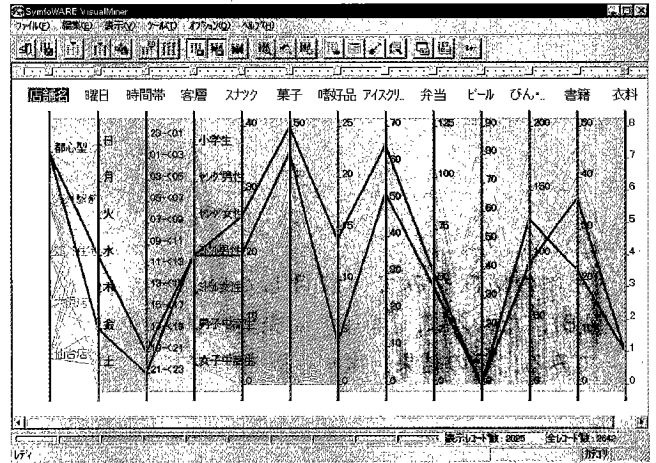


図-10 特定地区でフィルタリング

(平成12年2月28日受付)