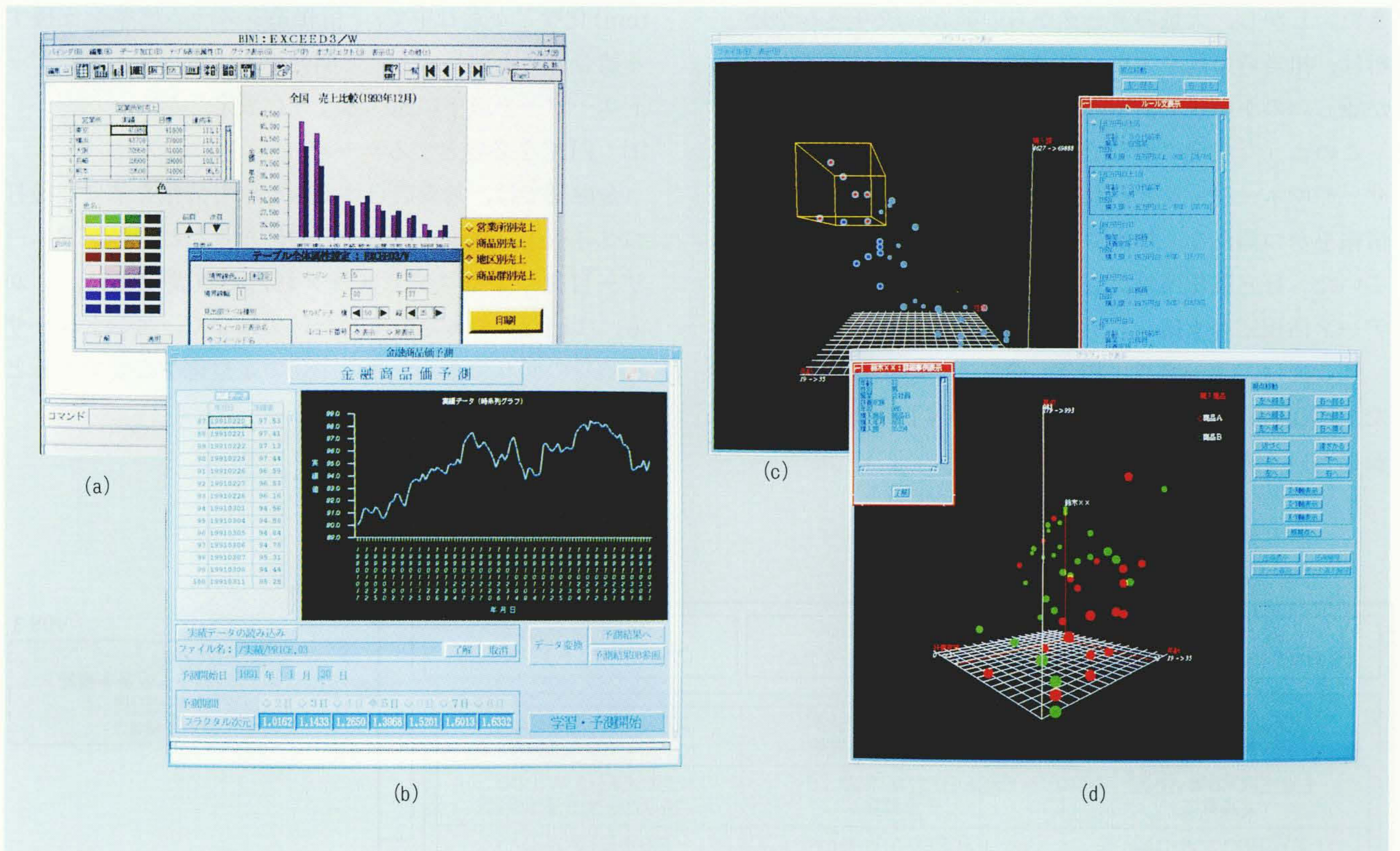


企業内大量データの戦略的活用を支援するデータ活用総合支援システム“EXCEED3”, “DATAFRONT”

—金融情報系における適用—

Total Support Tools for Strategic Data Use

山崎紀之* Noriyuki Yamasaki 高橋ヨリ* Yori Takahashi
山中一朋* Kazutomo Yamanaka 金子篤志* Atsushi Kaneko



データ活用総合支援システムの画面例

ビジュアルなヒューマン インタフェースにより、エンドユーザーが簡単に利用できる総合操作環境を提供する。(a), (b)はEXCEED3によるユーザーアプリケーション画面例を、(c), (d)はDATAFRONTによるルールモデルと多次元グラフ表示の画面例をそれぞれ示す。

金融業界をはじめとしてコンピュータの高度利用が進み、顧客のデータベースには貴重な情報が大量に蓄積されている。これらの情報をより有効に、また戦略的に活用するためのツールが求められてきた。その利用者は、データ分析の専門家ではなく、各部門のエンドユーザーである。

このような状況を踏まえて開発してきたEUC (End User Computing) ツールがEXCEED (Executive Management Decision Support System) であり、DATAFRONT (Data Frontier) である。EXCEED

は、ホストコンピュータの時代から活用されてきた製品であり、時代の変化に対応して現在ではパソコン(パーソナルコンピュータ)でも利用できるようになっている。DATAFRONTは、大量データの中から、その特徴を簡単な操作で、より高速に抽出できるデータマイニングツール(Data Mining: 価値ある情報の発掘の意)である。

エンドユーザーはこれらのツールを利用することにより、作業の効率向上と価値ある情報の発見による戦略的な業務アプローチを図ることが可能になる。

* 日立製作所 ソフトウェア開発本部

1 はじめに

BPR(Business Process Re-engineering)の推進や、勘定系・情報系の業務の自動化に伴い、企業内の情報システムには各種データが大量に蓄積されてきている。これらのデータを利用して、個人をターゲットにしたダイレクトマーケティングやテレマーケティングが実施されてきた。しかし、大量のデータからの顧客の絞り込みや分析は、担当者の勘やノウハウに頼っていて、システム化が難しいのが現状である。

このような状況に対して、大量のデータをエンドユーザーやマネージャなどだれもが簡単に処理でき、有効な情報をおつ漏れなく抽出できるシステムの提供が望まれていた。日立製作所は、これらの要望にこたえるためのデータ活用総合支援システムを開発してきた。

ここでは、データ活用総合支援システムとして開発した、EXCEED3シリーズとDATAFRONTのねらいと特長について述べる。

2 データ活用総合支援システム「EXCEED3シリーズ」

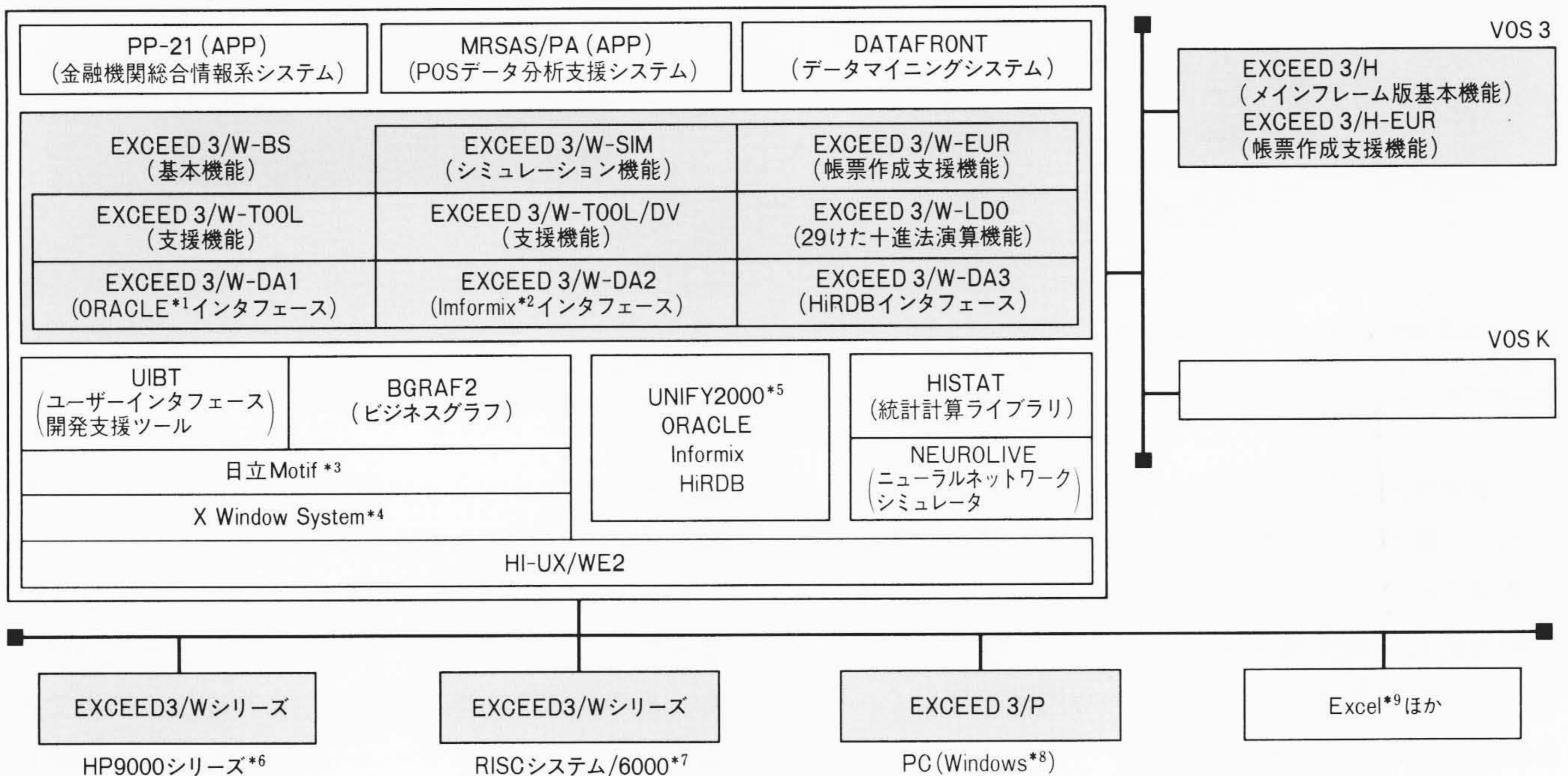
2.1 ターゲットユーザー

EXCEED3シリーズのシステム概要を図1に示す。EXCEED3シリーズはデータ利用を総合的に支援することを指向してきた。特にEUCやCSS(Client Server System)化などが進む中で、「情報系システム構築を支援する競合力ある製品」として開発するにあたり、ターゲットユーザーを以下に述べるように設定した。

(1) EUCを必要とするユーザー

金融業では、営業店自体で顧客取引情報を基に今後打つべき販売具体策の検討が必要になるなど、エンドユーザー自身によるコンピュータ利用が重要である。しかも、1回だけの非定型業務だけではなく、エンドユーザーの身の回りにある一連の業務の定型化が必要であると考えられる。

(2) ネットワークを利用するユーザー



注：略語説明ほか APP (Application Program Product), PC (Personal Computer), VOS K (Virtual-storage Operating System Kindness)
 *1 ORACLEは、米国Oracle Corporationの登録商標である。
 *2 Informixは、米国Informix Software, Inc.の商品名称である。
 *3 Motifは、Open Software Foundation, Inc.の商標である。
 *4 X Window Systemは、米国the Massachusetts Institute of Technologyが開発したソフトウェアである。
 *5 UNIFY2000は、米国Unify Corp.の商品名称である。
 *6 HP9000は、米国Hewlett-Packard Companyの商品名称である。
 *7 RISCシステム/6000は、日本アイ・ビー・エム株式会社の商品名称である。
 *8 Windowsは、米国Microsoft Corp.の商標である。
 *9 Excelは、米国Microsoft Corp.の商品名称である。

図1 EXCEED3シリーズのシステム構成
 EXCEED3シリーズは、ホストからパソコンまでをカバーする一貫したシステムである。

CSSをはじめとする情報システム環境の進歩により、これまで企業が基幹業務系システムに整備・蓄積してきた大量データを利用できるようになった。このため、メインフレーム、部門コンピュータ、ワークステーション、パソコンで構成されるネットワーク(LAN, WAN: Wide Area Network)を介した広域複合システム環境で利用できる必要がある。

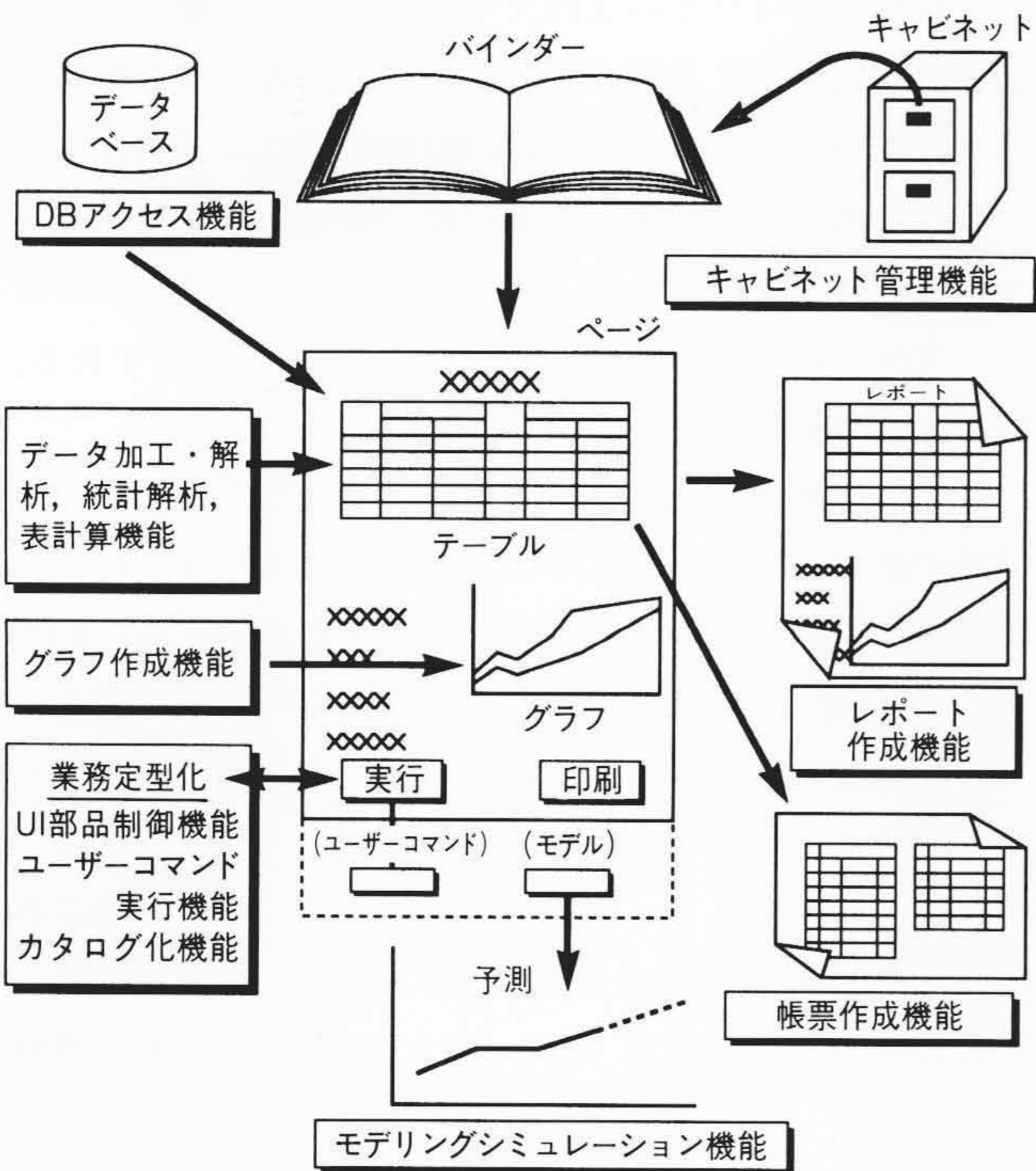
(3) オープン環境を利用するユーザー

CSS環境でも、国際(業界)標準仕様準拠インフラストラクチャ、特にGUI(Graphical User Interface)、通信、データベース、分散ファイルシステム、流通パソコンソフトへの対応が必要である。同時にマルチベンダ環境、著名データベース、他社ワークステーション、パソコン環境で利用できる必要がある。

2.2 EXCEED3シリーズの特長

EXCEED3シリーズでは、大量のデータを処理して目的に合った表の作成やシミュレーションに利用したり、グラフなどに視覚化して必要な資料が簡単に作成でき、手間のかかっていた帳票作成も大幅に効率化できる。

基幹業務系システムに蓄積されている大量データの活用、CSSをはじめとする分散システムとの連携、そして利



注：略語説明 DB(Database), UI(User Interface)

図2 統合されたデータ活用支援システム

データベース検索更新からデータ加工・解析、そしてグラフ・レポート・帳票まで、データ活用業務を総合的に支援する。

用者の要求にフレキシブルにこたえるアプリケーション開発へと、バランスのとれた情報活用環境を備えている。

これらの特長について以下に述べる(図2参照)。

(1) エンドユーザー向けユーザーインターフェース

キャビネット、机、バインダー、ページなど、ユーザーにわかりやすい作業環境を提供する。

(2) 定型業務化支援

試行錯誤的な業務履歴を利用し、定型的な業務アプリケーションが作成できる。

(3) 帳票作成支援

ワープロ感覚で手軽に、かつフレキシブルに帳票の作成や変更が行える帳票作成機能を提供する。

(4) オープンプラットフォーム環境への対応

他社ワークステーション、パソコンやORACLE, UNIFY2000, Informixなどの有力流通ソフトウェアとも連携している(図3参照)。

(5) 統合された総合データ活用支援システム

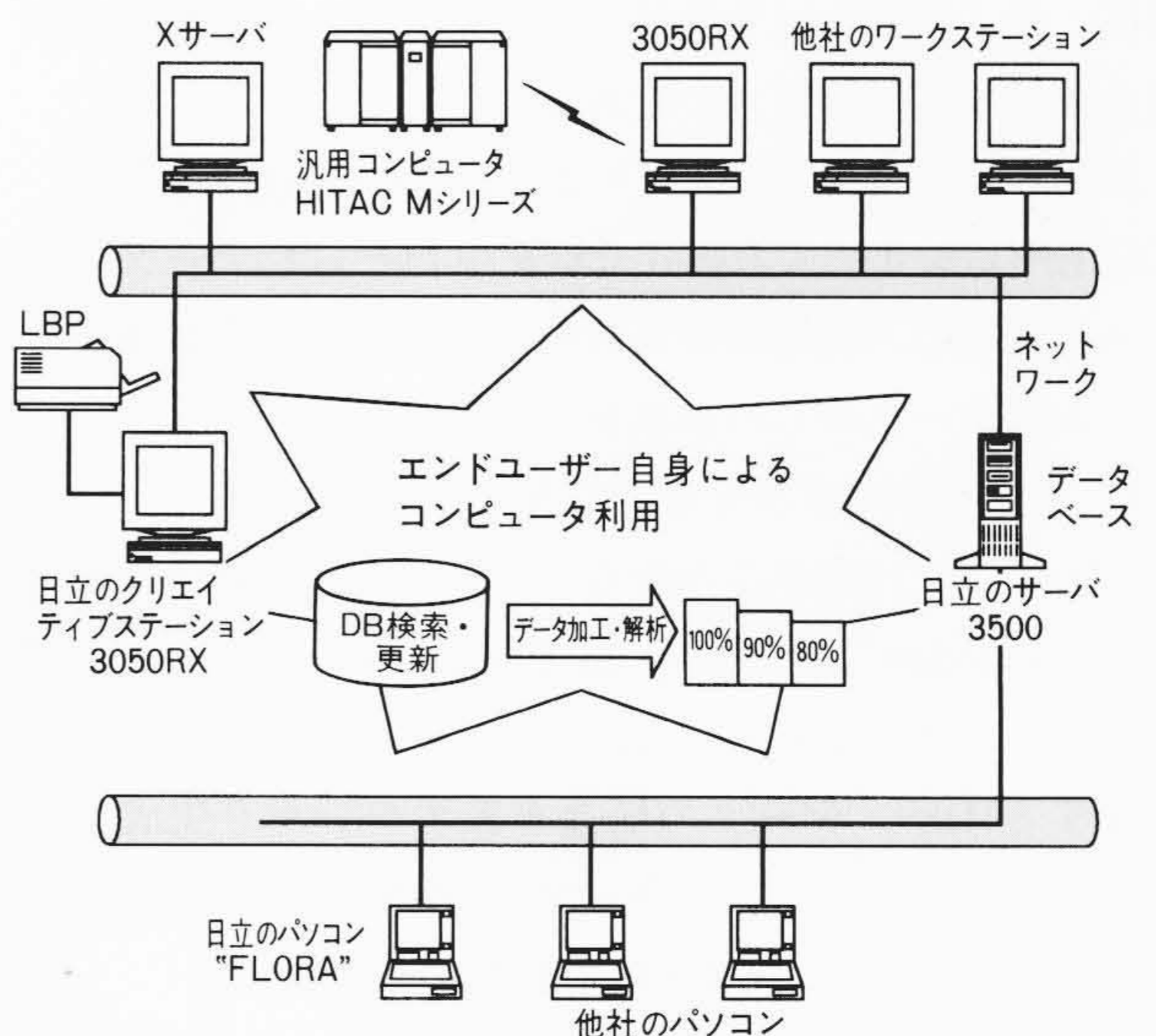
データの収集から統計解析や予測・シミュレーションまで、総合データ活用支援システムとして活用できる。

EXCEED3シリーズは、意思決定にかかわるこうした作業を総合的にとらえて支援するための機能群を豊富に用意した。

2.3 適用事例

全社情報システムの事例について以下に述べる(図4参照)。

(1) オーダー情報、生産情報、売上情報を基に、それら



注：略語説明 LBP(Laser Beam Printer)

図3 オープンプラットフォーム環境への対応

マルチベンダ環境での情報活用を実現する。

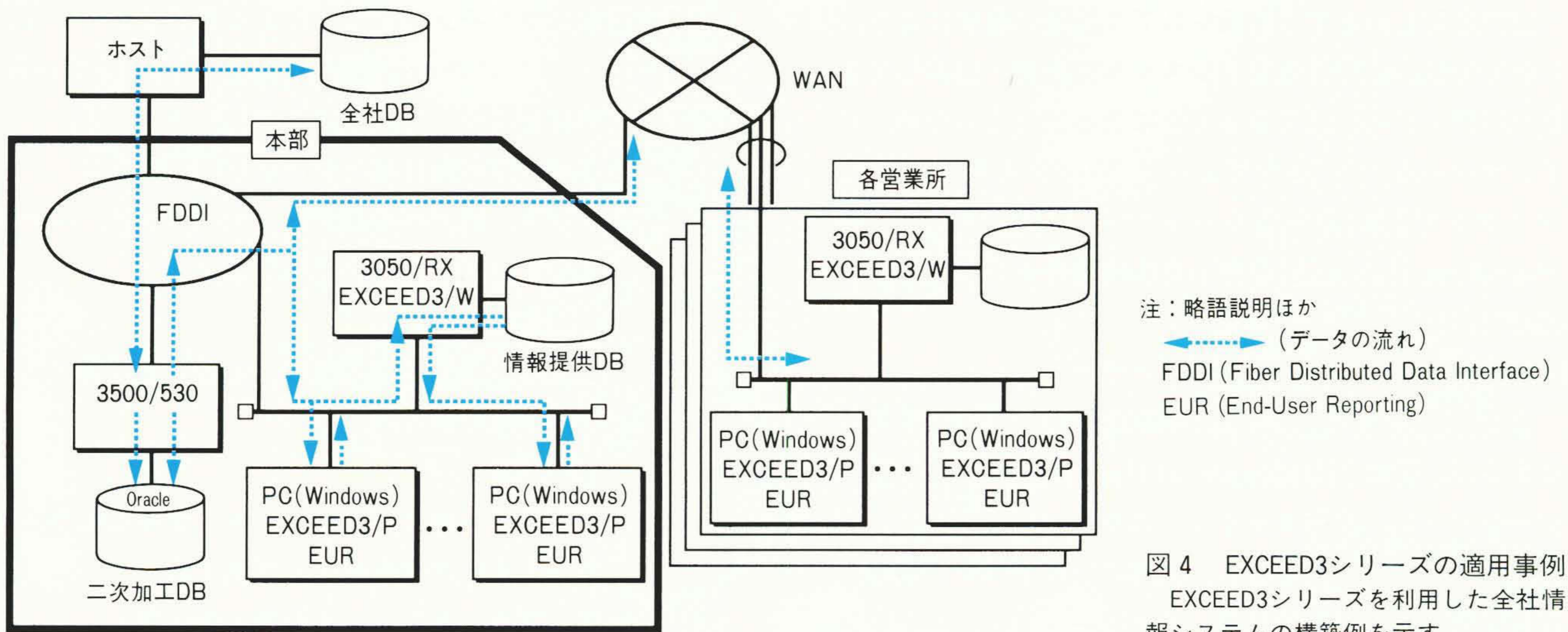


図4 EXCEED3シリーズの適用事例
EXCEED3シリーズを利用した全社情報システムの構築例を示す。

を加工・解析し、帳票を出力する。

(2) ホスト全社データベースから、二次加工データベースを作成する。EXCEED3/Wでそのデータベースにアクセスし、データを加工・解析し、帳票を出力する。

ワークステーションはデータベースサーバ、および大量データの処理に利用する。必要に応じて各担当ごとにパソコンを配布し、そこでワークステーションと同じアプリケーションを利用する。

3 データマイニングシステム“DATAFRONT”

データマイニングシステムとは、データベースに蓄えられた顧客情報や取引情報などの大量データをさまざまな角度から分析し、今まで気づかなかった「新たな因果関係」や「複雑な因果関係」を発見したり、推測した仮説を検証するためのものである。

データマイニングシステム“DATAFRONT”は、EXCEED3シリーズの統計分析を中心としたデータ解析手法を補完し、現場の担当者やマネージャの直接使用による大量データからの知識発見を可能にする。

3.1 今なぜデータマイニングか

米国では、1980年ころからデータベースからの知識発見(Knowledge Discovery in Databases)システムが開発された。著名なものでは、データを分類する判断木を生成するID3¹⁾、データの構造を表すルールを生成するIXL¹⁾、ニューロモデルを生成するDBM²⁾、自然言語形式のテキストを生成するEXPLORA³⁾などがある。

最近、大量データベース解析手法として、これらのデータマイニングシステムが注目されるようになった背景には、わが国の経済が「作れば売れる」好況期から「く

ふうしなければ売れない」低迷期に移行したことも深く関係している。このような時期に売り上げを伸ばすためには、徹底した市場調査と、ターゲットユーザーの絞り込みやユーザー嗜好(し)好の分析、販売予測や適正価格の設定などが重要になる。

市場分析の方法としては従来行われている統計手法があるが、この手法には次のような問題点がある。

- (1) 分析手法や結果の解釈など統計的な専門知識が必要
- (2) 性別や趣味などの定性データの分析が困難
- (3) 統計はデータ全体の把握が中心で、着目する特定データについての特性(または知識)の発見には不向き

これらの問題を解決する手段が、前述のルールやニューロモデル生成などのAI技術を使用して、大量のデータベースから知識を発見するデータマイニング技術である。

3.2 DATAFRONTの概要

日立製作所は、統計やAIについてまったく知識がない現場の担当者やマネージャでも簡単に使用でき、経営戦略立案に直接役立つ情報(知識)の発見を支援する目的で“DATAFRONT”を開発した。主な機能の構成を図5に示す。機能の概要について以下に述べる。

(1) 特徴ルールモデル生成・予測機能

ルールインダクション技術を使用して、データ中にある特徴的な因果関係を高速に見つけ出し、「IF 年収 = ××万円以上, THEN 契約 = 成立」などのだれもが理解しやすいルールモデルを生成して予測する。

具体的な特徴ルールの生成例を図6に示す。

(2) ニューロモデル生成・予測機能

データ中の因果関係が複雑でモデル化が困難な場合は、ニューロモデルを生成して学習・予測する。設定の

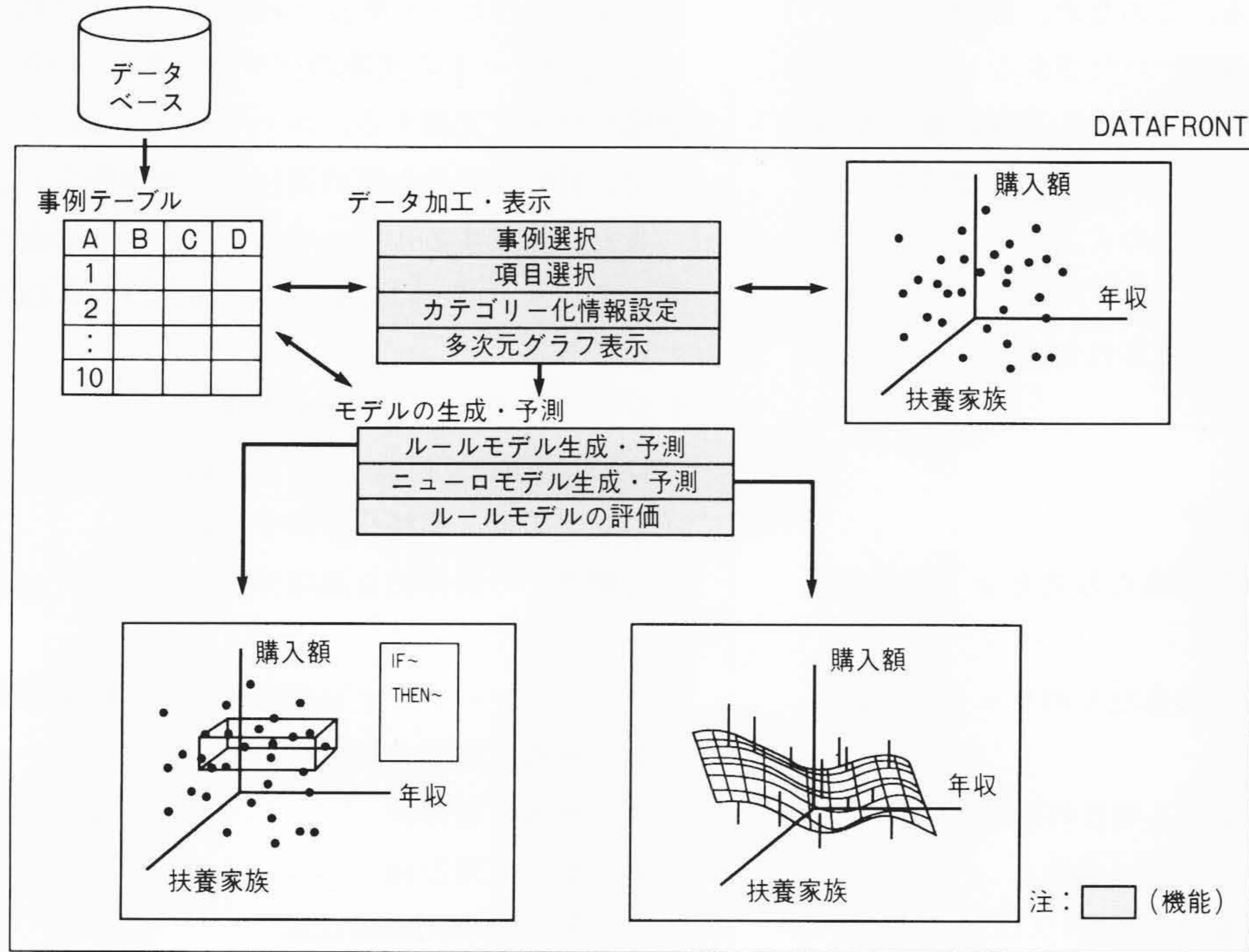


図5 DATAFRONTの機能構成
 多次元グラフ表示およびルールモデル表示の画面は、15ページの写真に掲載している。

困難な各種パラメータについての標準値が準備され、ニューロに関する知識がなくても容易に使用できる。

(3) 多次元グラフ表示機能

(a) 二～四次元の多次元散布図(四次元目は色使用)の表示

(b) 多次元散布図上の事例の詳細表示

データの分布状態を見ながら特別な値を持つ事例の詳細検討ができ、事例の絞り込みも可能である。

(c) 高速な特徴保存データ圧縮表示

密集した大量データでも分布状態の把握が容易

(d) モデルの多次元グラフ表示

モデルのふるまいを視覚的に評価できる。

(4) その他の知的データ加工機能

上記DATAFRONTの中心機能をより有効に活用するために、クラスタリングなどの統計手法やファジィなどのAI技術を使用した以下の機能がある。

(a) 事例選択機能：ランダムサンプリングや条件抽出

(b) 項目選択機能：条件となる項目の自動選択

(c) カテゴリー化情報設定機能：数値データの自動カテゴリー化(例えば「大」、「中」、「小」などに自動分類)

事例データ

事例	氏名	年齢	性別	職業	扶養家族	年収	購入商品	購入年月	購入額
事例1	佐藤××	29	男	会社員	1	512	商品A	9401	8109
事例2	田中××	30	男	会社員	0	479	商品B	9304	4627
事例3	鈴木××	31	男	会社員	1	587	商品A	9312	10725
事例4	斉藤××	31	女	会社員	0	443	商品B	9401	11236
事例5	高橋××	33	男	会社員	0	628	商品B	9207	6728
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

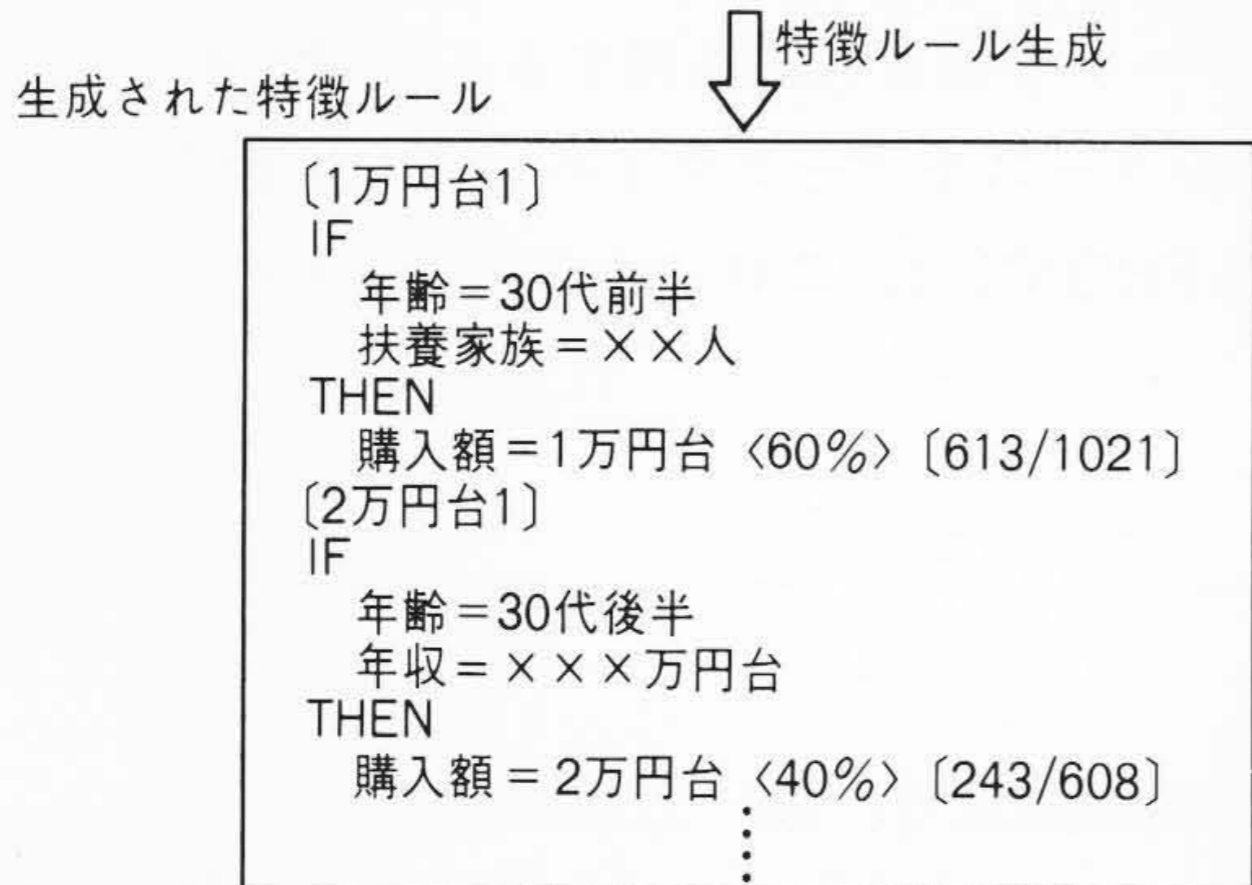


図6 事例データと生成された特徴ルールの例

購入額の区分ごとに顧客の特性を把握するルールを生成する例である。

3.3 DATAFRONTの特長

データマイニングシステムの中心機能である特徴ルール生成アルゴリズムの特長について以下に述べる。

従来のルールインダクションシステムは、「ある事例群を他の事例群から分類するためのルール(分類ルール)」を生成することを目的とし、すべての事例を分類できるようにルールを生成する。このため、例外的な事例についてもルールを生成するので雑音に弱く、生成されるルール数も多く、大量データの場合にはそのまま使用できないという欠点があった。

これに対し、DATAFRONTでは、データの持つ性質を多角的にとらえるために、「特徴的な事例群を見つけ、その性質を特徴として表すためのルール(特徴ルール)」

を生成することを目的とする。このため、数個しかないような事例はルールで表現(特徴づけ)されない。

DATAFRONTでは、大量データでも高速に質の高いルールを生成できるように、ルール生成アルゴリズムにさまざまなくふうをしている。このくふう点について簡単に述べる。

特徴ルールは以下の形式で生成される。

IF $X_1 = a_1$

$X_2 = a_1 \dots$

THEN $Y = b_1$ 〈ヒット率%〉

〔カバー事例数当たりのヒット事例数〕

$Y = b_2$ 〈ヒット率%〉

〔カバー事例数当たりのヒット事例数〕

...

ここで、 X_i : 条件の候補となる項目の名称

a_i : 項目 X_i のカテゴリ名称

Y : 結論となる項目の名称

b_i : 項目 Y のカテゴリ名称

ヒット率 : カバー事例数当たりのヒット事例数

ヒット事例 : ルールの条件と結論の両方を満たす事例

カバー事例 : ルールの条件だけを満たす事例

DATAFRONTのルール生成方式では、考えられるすべてのルールの条件節の組み合わせを、以下の尺度 $\mu(A)$ で評価し、評価尺度の高いルールを採用する。

$$\mu(A) = P(A)^\beta \sum_{i=1}^m P(b_i|A) \log(P(b_i|A)/P(b_i))$$

ここで、 $P(A)$: 事例がルールの条件を満たす確率

$P(b_i)$: 結論 Y が b_i である確率

$P(b_i|A)$: ルールの条件を満たす事例のうち結論 Y が b_i である確率

上記の評価尺度の定義式で、第1因子は「ルールのカバー率(全事例のうちでルールの条件を満たす事例の割合)」、つまり「ルールの一般性」を評価する。第2因子

は「ルールのヒット率」、つまり「ルールの精度」を評価する。 β は0~1の実数のパラメータで、一般性と精度の評価の比重を定義する。ユーザーは、これを変更することで、目的に応じた質の高いルールが得られる。

また、生成するルールの数や最大条件節数を変更できるので、ユーザーは必要なルールだけを短時間で得ることができる。

3.4 DATAFRONTの適用分野

DATAFRONTは、データが蓄積されてさえいればマイニングでき、業種や分野を問わない。

金融業での具体的な適用分野としては、次のものが想定される。

- | | |
|-----------------|--------------|
| (a) テレマーケティング分析 | (f) 相場分析 |
| (b) 融資先評価分析 | (g) アンケート分析 |
| (c) 顧客情報分析 | (h) 収益予測 |
| (d) 資金運用計画 | (i) 宣伝広告媒体選定 |
| (e) 新商品販売戦略立案 | |

4 EXCEED3シリーズおよびDATAFRONTの今後の展望

EXCEED3シリーズでは、マルチプラットフォームの対応をさらに進めている。データベースとして並列RDB(Relational Database)に対応し、パソコン版ではWindows NT*)化を推進中である。

DATAFRONTでは、現在HI-UX/WE2をサポートしている。今後はパソコンを含めたプラットフォームの拡大を検討していく。

5 おわりに

ここでは、データの戦略的活用を支援するシステムについて述べた。企業のマーケティング戦略に対する意思決定は、今まで以上に重要になってきている。そのためには、蓄積データをより有効活用することが必要である。EXCEED3シリーズやデータマイニングの技術を取り入れたDATAFRONTは、これに十分役立つものであると考える。

※) Windows NTは、米国Microsoft Corp.の商標である。

参考文献

- 1) K. Parsage, et. al. (近谷英昭訳) : 知的データベース, オーム社(1992)
- 2) 麻生川静男 : データベースから隠れた情報を発掘するツール「Data Base Mining」, 日経インテリジェントシステム, 別冊1992夏号, 102~112(1992-7)
- 3) J. Matheus, et. al. : System for Knowledge Discovery in Databases, IEEE Transactions on Knowledge and Data Engineering, Vol.5, No.6, 903~913(1993-11)