

ラフ集合理論による データマイニングシステムの開発

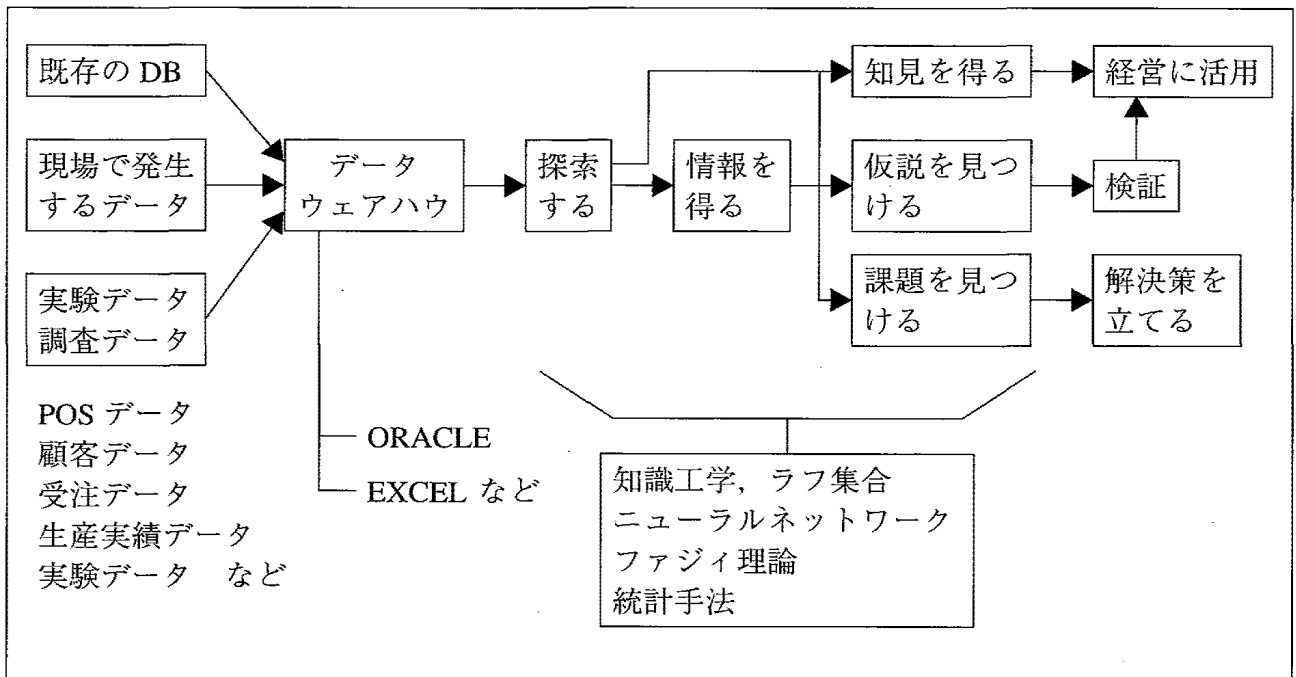
郭 沛 俊

1. はじめに

近年、大型情報システムの構築により、企業や政府機関に多量のデータが蓄積されるようになって、さらには、インターネットの普及に伴い、誰でも自由に利用できるような情報やデータが爆発的に増大した。一方、昨今のビジネス環境は著しく変化した。作れば売れるという時代は終わり、商品の類似化、ライフサイクルの短命化、ユーザーニーズの多様化に代表されるように市場は飽和した状態になってきた。新しいビジネス環境では、顔の見えない市場で勝負するのではなく、顧客一人一人のニーズに応え、よい関係を築くことが求められるようになってきている。このことから、過去に蓄積された大量なデータから如何に顧客の行動パターンを分析し、顧客ニーズの分析、新規顧客の開拓、他社との差別化、新製品の予測、競合会社の客層把握、自社の在庫管理や危険要素の把握など、あらゆる場面でデータを有効に活用することが企業の生き残るための必須条件になっている。この大量のデータの中から価値ある情報を掘り出す手法がデータマイニングである（図1）[25]。

今までデータマイニングの研究は、多変量解析、推論などの手法を中心に、様々な分析ツールを開発してきた。統計学では、平均、分散、相関関係などを用いてデータの内在的関係を数式で表すことはできたが、直感的に現象の内在的因果関係を把握することができなかった。また、これまで、通常でデータで表すことが難しい事象や、事象自体の複雑性が統計的でない場合には、統計理論を使うことができなかった。しかし、ラフ集合理論では、要因空間と結果空

図1 データマイニングの概念図



間を同値関係で分割し、二つの空間の関連性を調べることで、与えられたデータから人間の感性に近いルールを導き出し、複雑な現象からより簡潔なルールで因果関係を表し、研究対象の重要要因を取り出すことが可能になっている [1-19, 21]。

本論文では、ラフ集合理論に基づき、データマイニングソフトウェアの開発及び実用化を目的とする。応用として、穴吹興産(株)のアンケート調査データに基づき、主要マーケットの特性を分析し、顧客のマンションに対する希望価格と顧客の属性との関係を感性ルールで表し、マーケティング立案へ適用する [20]。

2. ラフ集合の基本理論

2.1 ラフ集合に関する基本概念

集合 X の二つ要素の間に規定される関係 φ を二項関係 (binary relation) という。そして、 X の二つ要素 x と y がこの二項関係を満たすとき $x\varphi y$ と書く。集合 X 上の二項関係 φ が次の三条件を満たすとき、 φ は同値関係 (equivalence relation) であるという。

- (1) 反射性 (reflexivity) : 全ての $x \in X$ に対して, $x \varphi y$
 (2) 対称性 (symmetricity) : $x, y \in X$ に対して, $x \varphi y$ ならば $y \varphi x$
 (3) 推移性 (transitivity) : $x, y, z \in X$ に対して, $x \varphi y$ 且つ $y \varphi z$ ならば $x \varphi z$

全体対象の集合を U , 対象の属性集合を A として, (U, A) を情報システム (information system) と呼ぶ。ある属性 $a \in A$ の値の集合を V_a とする。対象 $x \in U$ の属性 a の値を $a(x)$ と表して, 属性の集合 A の部分集合 B によって U 上の同値関係 $I(B)$ を次のように定義する。

定義 1.

対象 $x \in U$ と対象 $y \in U$ の $B \subseteq A$ に関して同値関係は $\forall a \in B \Rightarrow a(x) = a(y)$ によって定義され, $xI(B)y$ と表す。

全体対象の集合 U は同値関係 $I(B)$ によって分割される。この分割された集合族は U の分割と呼び, U/B と表し, その要素を基本集合と呼ぶ。対象 x が属している基本集合を $B(x)$ と表す。新たな集合 Z が与えられたとき, Z を基本集合によって近似することができる。

定義 2.

$$B^*(Z) = \bigcup \{Y \in U/B : Y \cap Z \neq \emptyset\} = \{x \in U \mid B(x) \cap Z \neq \emptyset\} \quad (1)$$

$$B_*(Z) = \bigcap \{Y \in U/B : Y \subseteq Z\} = \{x \in U \mid B(x) \subseteq Z\} \quad (2)$$

$$BN_B(Z) = B^*(Z) - B_*(Z) \quad (3)$$

ただし, \emptyset は空集合である。 $B^*(Z)$, $B_*(Z)$ 及び $BN_B(Z)$ をそれぞれ Z の上界近似 (upper approximation), 下界近似 (lower approximation) と境界領域 (boundary region) と呼ぶ。

上界近似と下界近似の性質 :

- (a) $B_*(Z) \subseteq Z \subseteq B^*(Z)$
 (b) $B_*(\emptyset) = \emptyset = B^*(\emptyset)$, $B_*(U) = U = B^*(U)$
 (c) $B_*(X \cup Y) \supseteq B_*(X) \cup B_*(Y)$
 (d) $B^*(X \cup Y) = B^*(X) \cup B^*(Y)$
 (e) $B_*(X \cap Y) = B_*(X) \cap B_*(Y)$
 (f) $B^*(X \cap Y) \subseteq B^*(X) \cap B^*(Y)$

$$(g) \quad X \subseteq Y \Rightarrow B_*(X) \subseteq B_*(Y) \text{ 且つ } B^*(X) \subseteq B^*(Y)$$

$$(h) \quad B_*(U-X) = U - B^*(X)$$

$$(i) \quad B^*(U-X) = U - B_*(X)$$

$$(j) \quad B_*(B_*(X)) = B^*(B_*(X)) = B_*(X)$$

$$(k) \quad B^*(B^*(X)) = B_*(B^*(X)) = B^*(X)$$

定義3.

(a) $B_*(Z) \neq \emptyset$, $B^*(Z) \neq U$ ならば, Z は上からも下からも定義できる。

(b) $B_*(Z) = \emptyset$, $B^*(Z) \neq U$ ならば, Z は下から定義できない。

(c) $B_*(Z) \neq \emptyset$, $B^*(Z) = U$ ならば, Z は上から定義できない。

(d) $B_*(Z) = \emptyset$, $B^*(Z) = U$ ならば, Z は上からも下からも定義できない。

$I(B)$ による Z の近似の精度は $\alpha_B(Z)$ と表し, 次のように定義する。

$$(e) \quad \alpha_B(Z) = |B_*(Z)| / |B^*(Z)|$$

ただし, $|A|$ は集合 A の要素の数である。

2.2 決定表

情報システム (U, A) において, 属性 A は条件属性と決定属性の二つの種類に分けて取り扱うこともある。このとき, 情報システムを決定表 (decision table) と呼ぶ。

2.2.1 属性の従属性 (dependency of attributes)

定義4.

D と C は属性集合の部分集合とする。 D は完全に C に従属していることを $C \Rightarrow D$ と表し, 次のように定義する。

$$C \Rightarrow D \iff I(C) \subseteq I(D) (xI(C)y \Rightarrow xI(D)y) \quad (4)$$

つまり, 任意の x と y が C に関して同値関係であるならば, D に関して同値関係が成り立つとき, D は完全に C に従属するという。このとき, C による全体集合 U の分割 U/C は D による分割 U/D より細かい。

定義5.

D と C は属性集合の部分集合であるとする。 D は k 程度で C に従属していることを $C \Rightarrow_k D$ と表し, k は次のように定義される。

$$k = |Pos_C(D)| / |U| \quad (5)$$

ただし

$$Pos_C(D) = \prod_{X \in U/D} C_*(X) \quad (6)$$

式(6)に示したように $Pos_C(D)$ は部分属性集合 D により分割された基本集合の部分属性集合 C による下界近似である。したがって, k は C による分割と D による分割との一致性を示している。

もし C が条件属性集合, D が決定属性集合であれば, $Pos_C(D)$ は決定を正しくサポートする対象の集合となる。

2.2.2 決定表における属性の簡略化 (reduction of attributes)

決定表における属性の簡略化は与えられた決定表から属性の従属性に基づいて余分な属性を取り除くことである。

定義6.

B は属性集合 A の部分集合で, a は B の要素であると仮定する。

- (1) もし $I(B) = I(B - \{a\})$ が成立すれば, a は $I(B)$ にとって“余分”(dispensable)であるという。そうでなければ, a は $I(B)$ にとって“不可欠”(indispensable)であるという。
- (2) B の全部の要素は $I(B)$ にとって不可欠であれば, B は“独立”(independent)であるという。
- (3) B の部分集合 B^* は独立であり, 且つ $I(B) = I(B^*)$ であれば, B^* は B の reduct と呼び, $Red(B)$ と表す。
- (4) 全ての $Red(B)$ の交わり集合 $I Red(B)$ は, B -コアと呼び, $Core(B)$ と表す。

定義6で示したように, $Red(B)$ は B と同様に対象集合 U を分割できる最小の B の部分集合であり, $Core(B)$ は B の中の最も重要な属性集合である。

以下では二つの属性集合 C と D の従属性に基づく属性の簡略化を考えていく。それは、 C の要素が減少することに関して、与えられた従属関係 $C \Rightarrow D$ に及ぼす影響があるかどうかを調べることである。

定義7.

D と C を属性集合の部分集合とする。 $Pos_C(D) = \prod_{X \in U/D} C_X(X)$

- (1) もし $Pos_C(D) = Pos_{C-\{a\}}(D)$ が成立すれば、 a は D にとって“余分”(D-dispensable)であるという。そうでなければ、 a は D にとって“不可欠的”(D-indispensable)であるという。
- (2) C の全部の要素は D にとって“不可欠的”ならば、 C は D にとって“独立”である (D-independent) という。
- (3) C の部分集合 C^* は独立であり、尚且つ $Pos_C(D) = Pos_{C^*}(D)$ であれば、 C^* は C の D-reduct と呼び、 $Red_D(C)$ と表す。
- (4) 全ての $Red_D(C)$ の交わり集合 $\bigcap Red_D(C)$ は、 C の D-コアと呼び、 $Core_D(C)$ と表す。

定義7で示したように、 $Red_D(C)$ は D によって分割された基本集合を C と同様に下から近似できる最小の C の部分集合であり、 $Core_D(C)$ は D に関して C の中の最も重要な属性の集合である。もし C を条件属性集合、 D を決定属性集合とすれば、 $Red_D(C)$ は決定の正しさを維持するための不可欠な条件属性の集合であり、 $Core_D(C)$ は決定に対して最も重要な条件属性の集合である。

2.2.3 対象に関する属性の簡略化

上記では、与えられた情報テーブルから余分な属性を省略する方法を紹介したが、これからはそれぞれの対象について余分な属性を省略する方法を紹介する。

定義8.

B は属性集合 A の部分集合で、 a は B の要素である。また x は対象集合 U の要素である。

- (1) $B(x) = (B - \{a\})(x)$ ならば, a は x にとって “余分” (dispensable for x) であるという。そうでなければ, a は x にとって “不可欠” (indispensable for x) であるという。
- (2) B の全ての属性は x にとって “不可欠” ならば, B は x に関して “独立” (independent for x) であるという。
- (3) B の部分集合 B^* は x にとって独立であり, 尚且つ $B(x) = B^*(x)$ が成り立つならば, B^* は x に関する B の reduct と呼び, $Red^x(B)$ と表す。
- (4) 全ての $Red^x(B)$ の交わり集合 $\bigcap Red^x(B)$ は, x に関する B -コアと呼び, $Core^x(B)$ と表す。

定義9.

D と C を属性集合 A の部分集合とする。 a は C の要素で, x は対象集合 U の要素である。

- (1) もし $C(x) \subseteq D(x) \Rightarrow (C - \{a\})(x) \subseteq D(x)$ が成立するならば, a は x に関する “ D の余分” (D-dispensable for x) であり, そうでなければ, a は x に関する “ D の不可欠” (D-indispensable for x) であるという。
- (2) C の全部の要素が x に関する D の “不可欠” ならば, C は x に関して “ D の独立” (D-independent for x) であるという。
- (3) C の部分集合 C^* は x に関する D の独立であり, 尚且つ $C(x) \subseteq D(x) \Rightarrow C^*(x) \subseteq D(x)$ ならば, C^* は x に関する C の D-reduct と呼び, $Red_D^x(C)$ と表す。
- (4) 全ての $Red_D^x(C)$ の交わり集合 $\bigcap Red_D^x(C)$ は, x に関する C の D -コアと呼び, $Core_D^x(C)$ と表す。

3. データマイニングシステムの構築

3.1 システムの分析

前章で紹介したラフ集合の基本理論に基づき, EXCEL 2003 VBA (Visual Basic for Applications) を用いて, データマイニングシステムの開発を行った。VBA とは, Excel や Access などの Microsoft Office 製品に搭載されているマク

口言語である。マクロは Excel ブックのモジュールに記述される。モジュールは「標準モジュール」, 「ユーザーフォーム」, 「クラスモジュール」, 「Excel オブジェクト」の四種類がある [22, 23, 24]。

本システムは大きく三つのサブ・システムに分けられる。即ち、データの前処理, 基本集合の形成, if-then ルールの形成である。各サブ・システムは細かく違う機能を持つモジュールから構成される (図2)。

データマイニングシステムは主に標準モジュール三つとユーザーフォーム五つにより構成されている。

標準モジュール : rough_set_main

ユーザーフォーム : データ範囲の選択 (データの取得)

ユーザーフォーム : 不要データの処理 (信頼できないデータの削除)

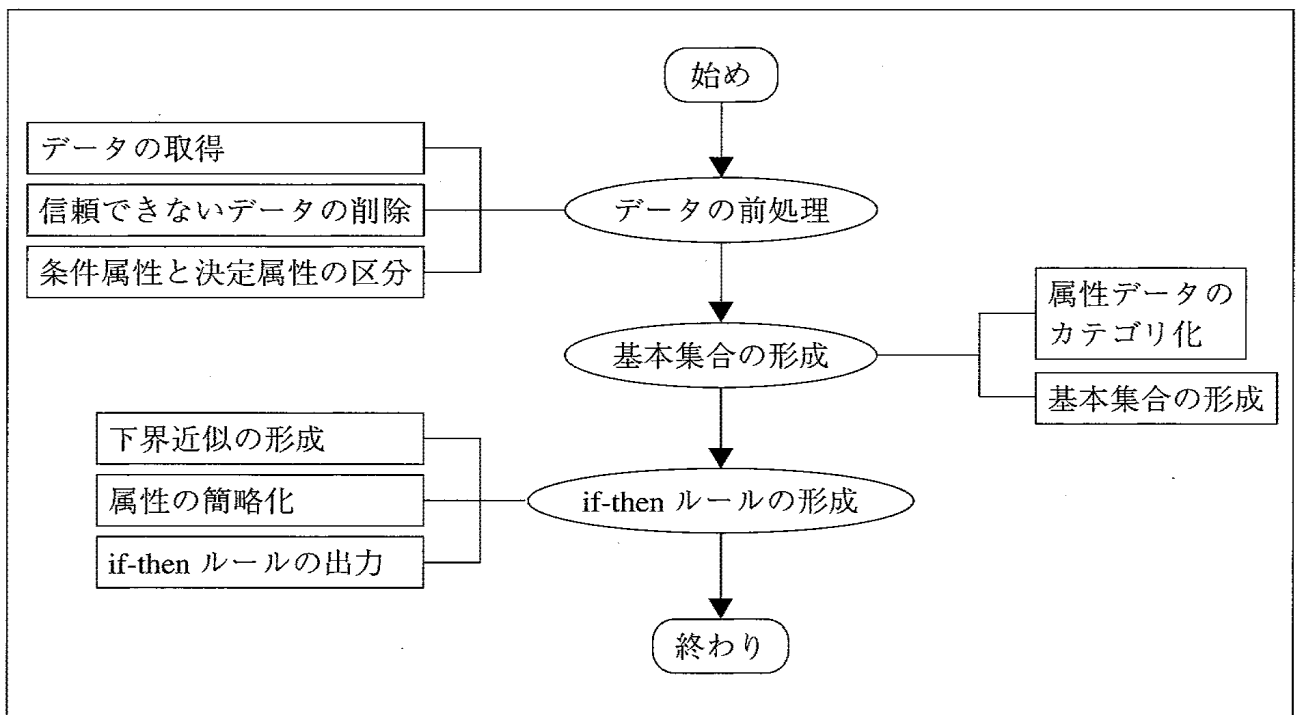
ユーザーフォーム : 条件と決定 (条件属性と決定属性の区分)

ユーザーフォーム : カテゴリの判断 (属性データのカテゴリ化)

ユーザーフォーム : 区切り値によるカテゴリ化 (基本集合の形成)

標準モジュール : low (下界近似の形成)

図2 データマイニングシステムの構造



標準モジュール：rule（属性の簡略化）

標準モジュール rough_set_main は各モジュールを順番に呼び出し，システムの全体の機能を実現する。標準モジュール rough_set_main は次のようになる。

Sub rough_set_main()

データ範囲の選択. Show

不要データの処理. Show

条件と決定. Show

区切り値によるカテゴリ化. Show

Call low

Call rule

End Sub

3.2 システムの設計

本節は図2を参照しながら，プログラムの流れの順に各モジュールを詳細に説明する。

3.2.1 データの前処理

サブ・システムーデータの前処理の構造は図3に示される。データの前処理では，データを読み込み，ユーザーの判断により絞り込み条件を入力し，信頼できないデータを削除する。そして，関連性の検定により，必要な条件属性を選択する。

3.2.1.1 データの取得

(1) コントロールの配置 (図4)

データを読み込むため，「データ範囲の選択」ユーザーフォームが用いられる。RefEdit コントロールを配置し，データのアドレスを取得する。選択されたデータにラベルが付くかどうかは「ラベル有り」チェックボックスで判断する。「実行」ボタンを押すと，データが読み込まれる。「終了」ボタンで処理を

図3 データ前処理のフローチャート

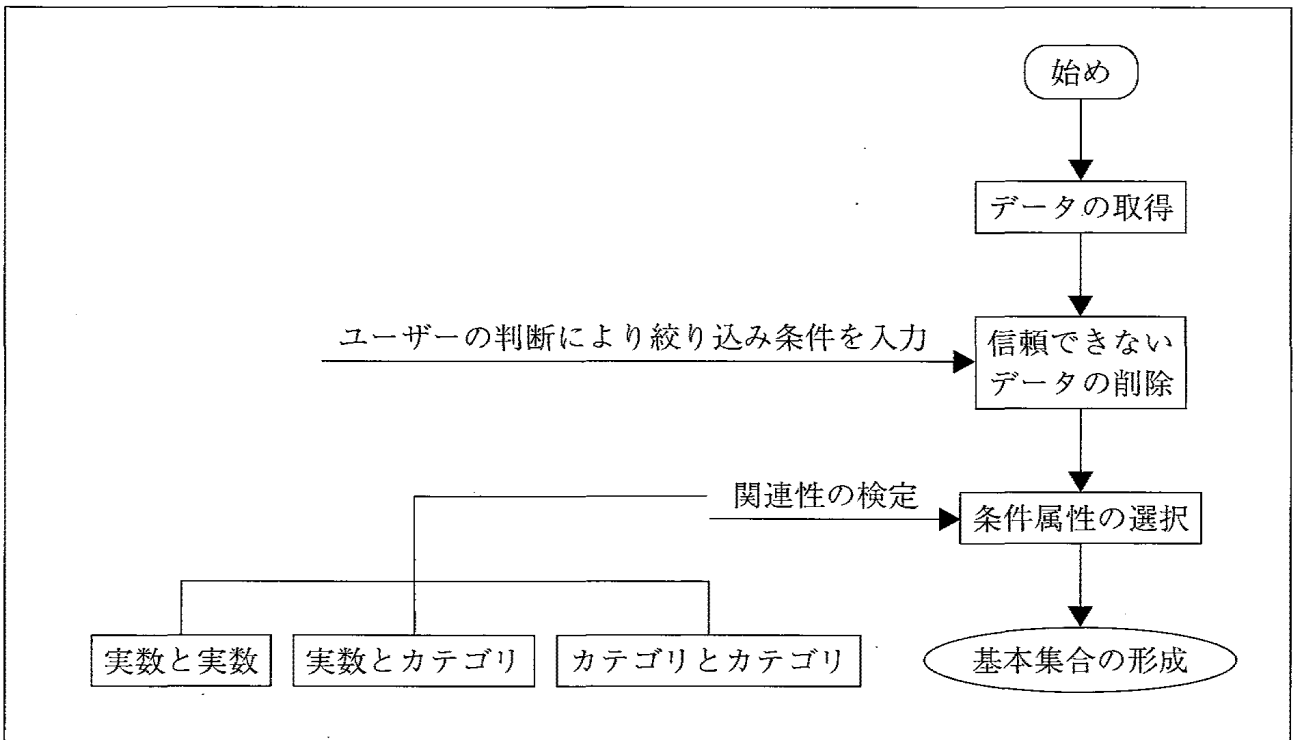


図4 「データ範囲の選択」画面

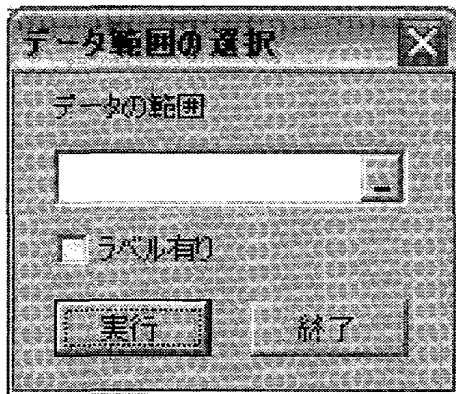
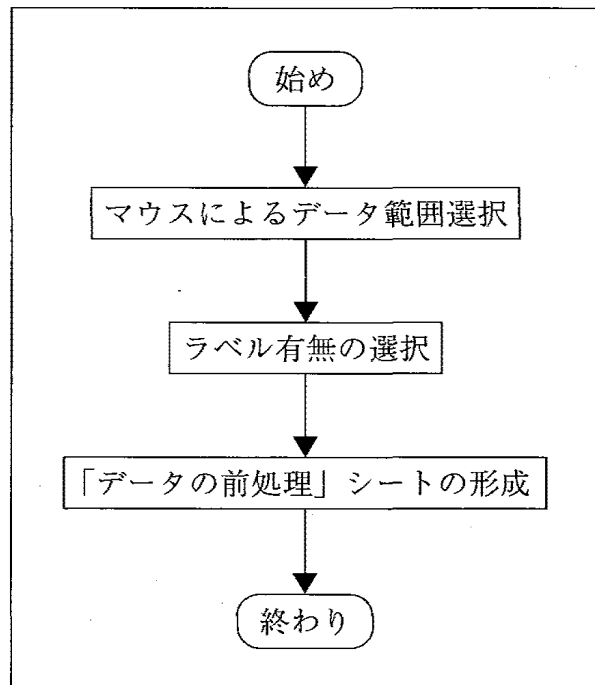


図5 データを取得するフローチャート



中止する。

(2) 「実行」ボタンで行う処理

図5に示したように Refedit からデータのアドレスを取得し、ラベルの有無を判断し、読み込んだデータを「データの前処理」シートにコピーする。

3.2.1.2 信頼できないデータの削除

(1) コントロールの配置

図6は動的コントロールで形成した「不要データの処理」ユーザーフォームである。ユーザーフォームのラベル名は選択されたデータのラベル名と自動的に対応する。選択されたデータはラベルがなければ、ユーザーフォームのラベルは属性1から属性n (nはデータの列数) と名づけられる。テキストボックスで絞り込み条件を取得する。「実行」ボタンを押すと、不要データの行が削除される。「クリア」ボタンは入力された絞り込み条件をクリアする。「終了」ボタンで処理を中止する。

(2) 「実行」ボタンで行う処理 (図7参照)

テキストボックスの内容を読み込み、「,」で分割された一つ目の文字を比較演算子、残った文字を削除データの条件値とし、セルの値と条件値を比較し、一致すれば、セルの行を削除する。

(3) 絞り込み条件の入力規則：

- ①「=」, 「<」, 「>」を含んで、条件を入力する。
- ②同じ項目に複数の絞り込み条件がある場合, 「,」で分ける。
- ③不要のデータを入力する。
- ④半角で入力する。

例えば、項目 no. 161 と 100 より小さいデータを削除するには no のテキストボックスに “=161, <100” を入力する。

図6 「不要データの処理」画面

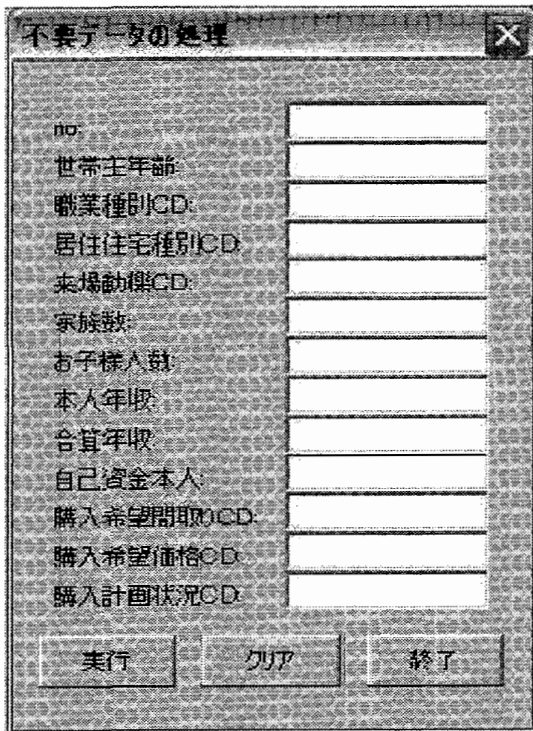
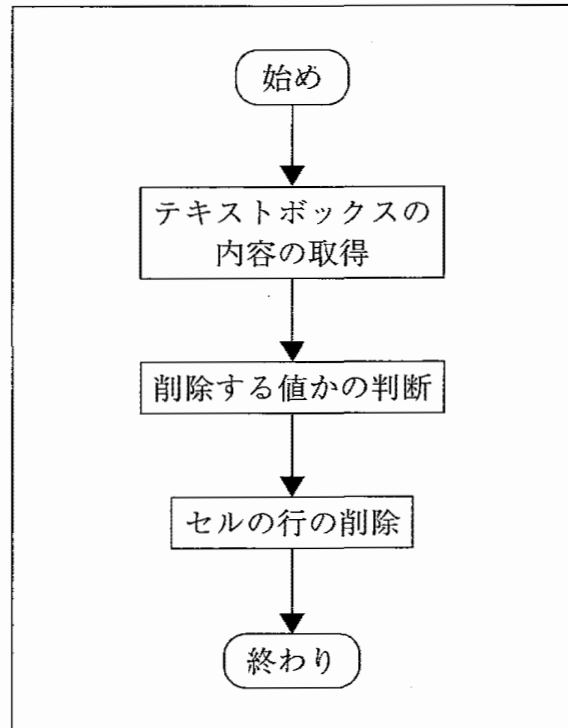


図7 不要データ処理のフローチャート



3.2.1.3 条件属性の選択

(1) コントロールの配置

図8で示したように「条件と決定」ユーザーフォームに「条件属性の範囲 RefEdit」と「決定属性の範囲 RefEdit」を配置し、条件属性データと決定属性データのアドレスを分けて取得する。「実行」ボタンを押すと、条件属性と決定属性が選択される。属性を選択する際に、条件属性と決定属性のデータ形態は「カテゴリの判断」ユーザーフォーム（図9）より、取得する。

(2) 「実行」ボタンで行う処理

RefEdit から条件と決定属性データのアドレスを取得し、「条件と決定」シートを形成する。その後、条件と決定属性のデータ形態を判断する。条件と決定属性のデータともにカテゴリデータの場合にはカイ二乗検定、条件と決定属性のデータともに実数データの場合にはT検定、条件と決定属性のデータの 하나가実数、もう一つがカテゴリの場合にはF検定を行い、相関でない属性を削除する（表1）。

表1 検定の種類

2項目のデータ形態	相関係数	検定手法
実数データ 実数データ	単相関係数	T検定
実数データ カテゴリデータ	相関比	F検定
カテゴリデータ カテゴリデータ	独立係数	カイ二乗検定

図8 「条件と決定」画面

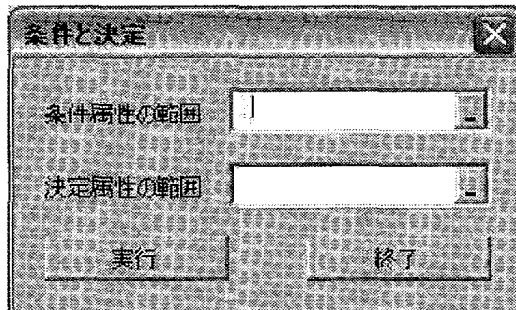
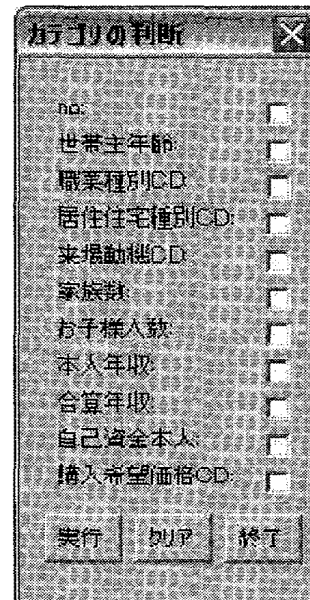


図9 「カテゴリの判断」画面



3.2.2 基本集合の形成

基本集合の形成モジュールでは、処理されたデータに基づき、ユーザーの入力により、属性データをカテゴリ化し、第1章で説明したラフ集合の同値関係により基本集合を形成した上で、下界近似を求め、if-thenルールを出力する(図11参照)。

3.2.2.1 属性データのカテゴリ化

(1) コントロールの配置 (図12)

「カテゴリ化」ユーザーフォームも動的コントロールで形成したものである。

図10 条件と決定のフローチャート

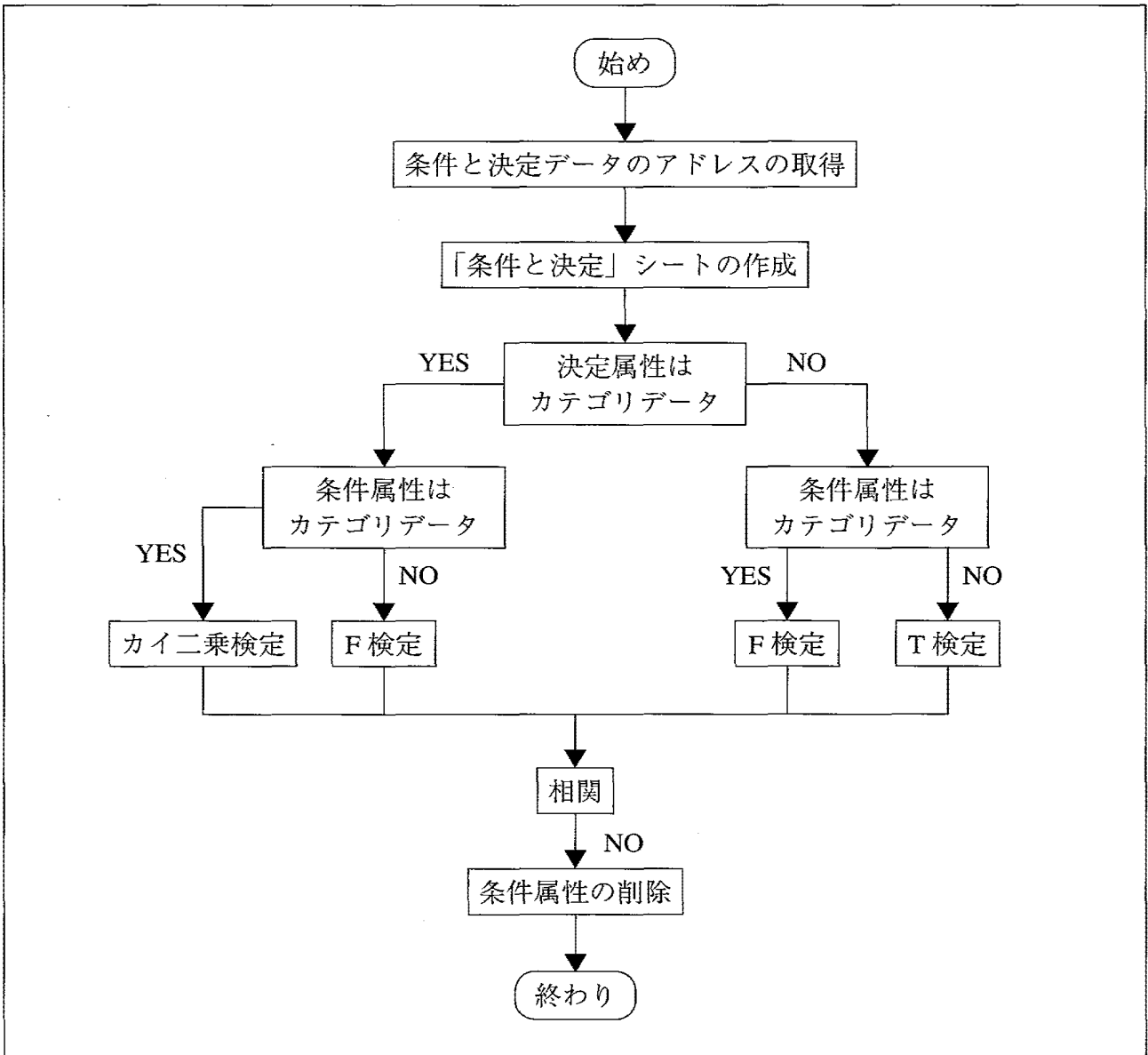


図11 基本集合形成のフローチャート

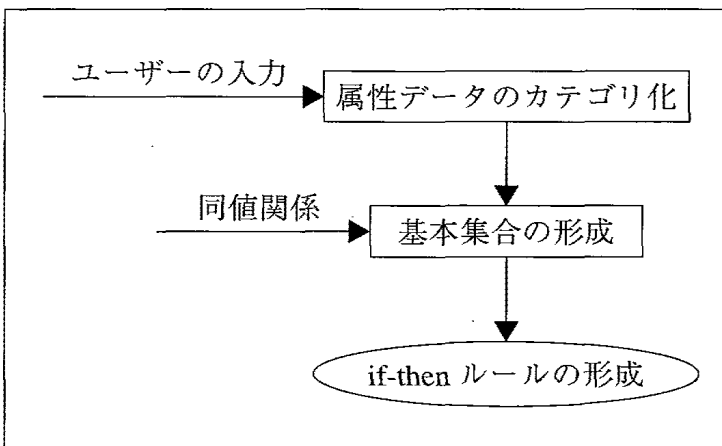
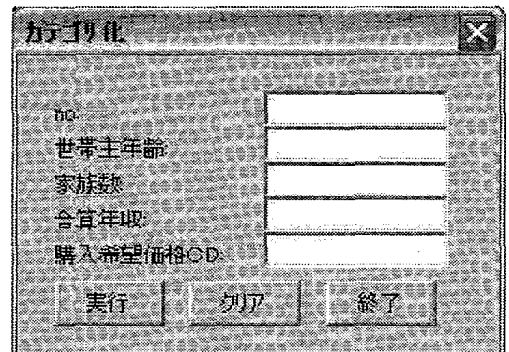


図12 「カテゴリ化」画面



テキストボックスでカテゴリ化条件を取得する。「実行」ボタンを押せば、カテゴリ化を実行する。

(2) 「実行」ボタンで行う処理 (図 13)

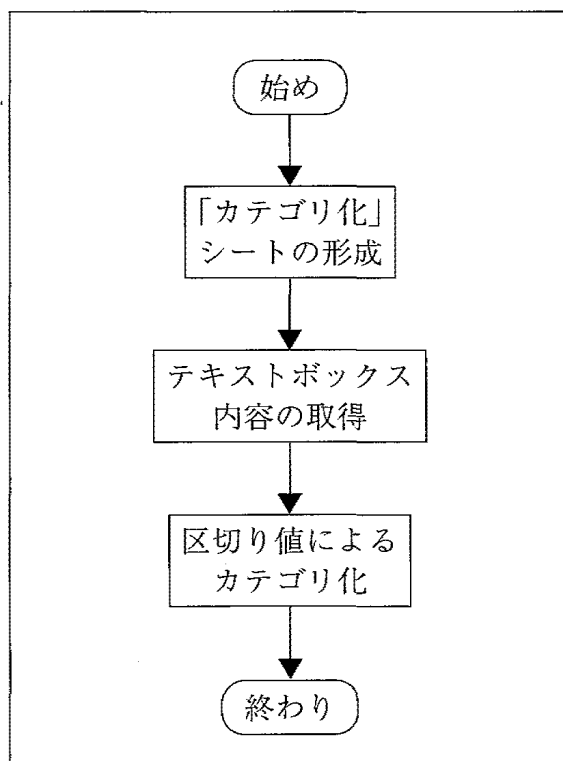
まず、「条件と決定」シートを「カテゴリ化」シートにコピーする。このテキストボックスは「カテゴリ化」シートの列と対応する。テキストボックスの内容 (カテゴリ化入力規則に基づき、入力されたカテゴリ化条件) を読み込み、「,」で分割し、文字を境界値とし、セルの値と境界値を比較し、カテゴリ化する。

(3) カテゴリ化入力規則：

- ①境界値を入力する。
- ②同じ項目に複数の絞り込み条件がある場合、「,」で分ける。
- ③半角で入力する。

例えば、1を25歳以下、2を26～50歳、3を51～75歳、4を76歳以上にする場合には、「25, 50, 75」を入力すればいい。

図13 カテゴリ化のフローチャート



3.2.2.2 基本集合の形成と if-then ルールの形成

基本集合の形成と if-then ルールの形成では、カテゴリデータに基づき、同値関係を用いて、基本集合を求める。求めた基本集合を用いて、決定属性による分割された集合の下界近似を形成し、属性の簡略化及び対象に関する属性を簡略化し、求められたルールを出力する（図 14）。下界近似と属性の簡略化の詳細を図 15 と図 16 によって示している。

図14 基本集合と if-then ルールの形成のフローチャート

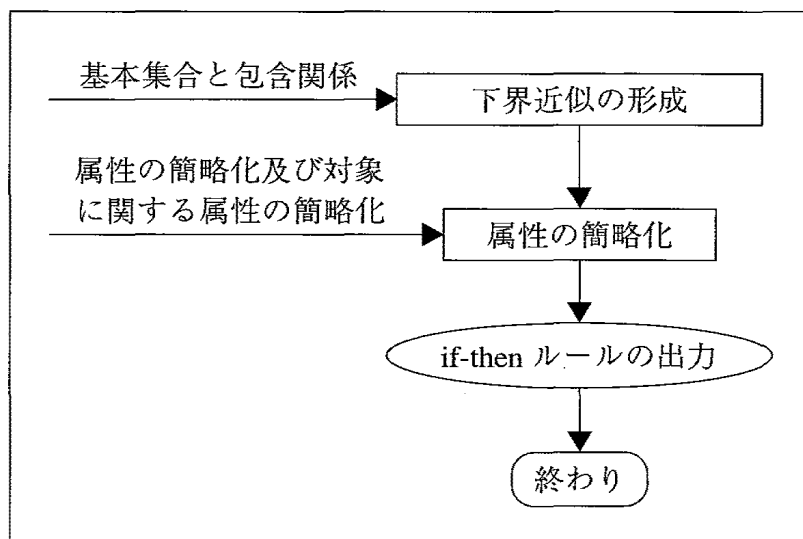


図15 下界近似のフローチャート

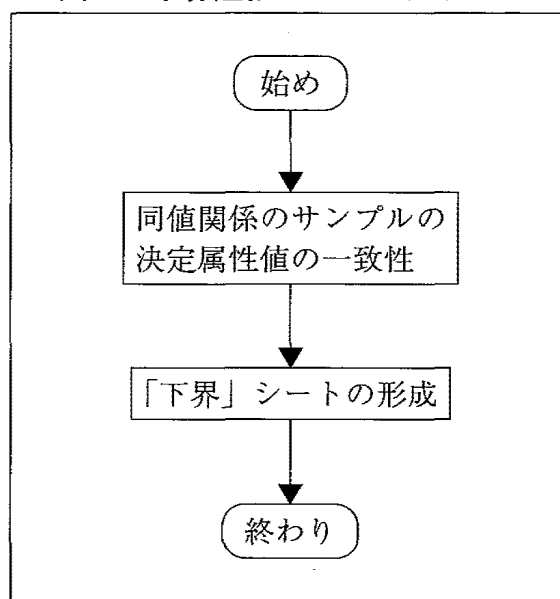
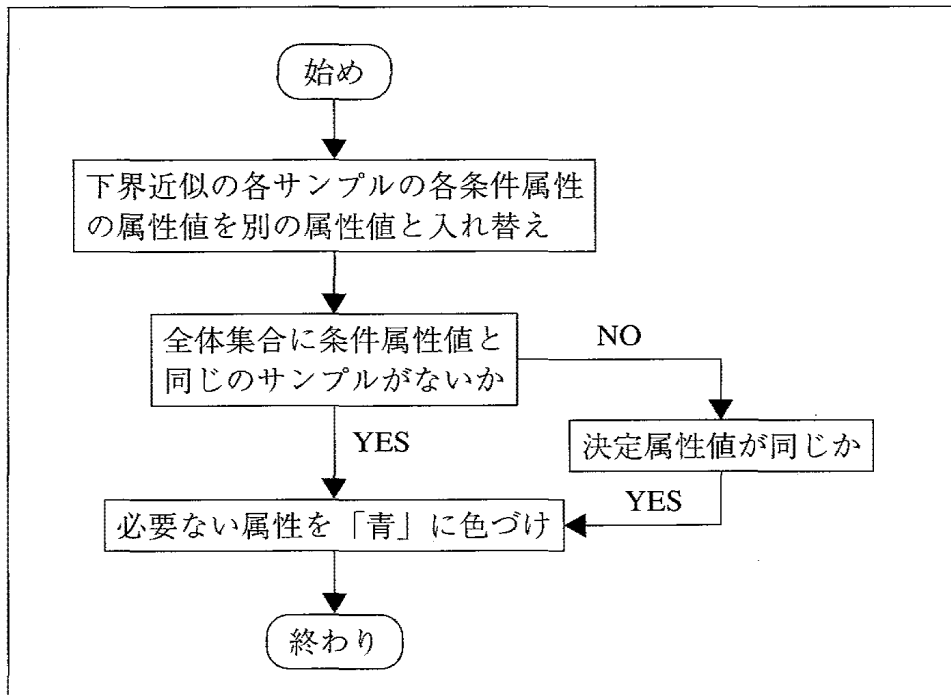


図16 属性の簡略化のフローチャート



4. マーケティングへの応用

4.1 データ概要

1. アンケート実施者：穴吹興産(株)
2. 時期：2003年10月1日～12月31日
3. 場所：四国地区（高松市，徳島市，高知市，松山市）における穴吹モデルルーム
4. サンプル数：857
5. 関連する分析用項目：
 - (1) 世帯主年齢：0 = 不明
 - (2) 職業種別 CD：1 = 会社員，2 = 会社役員，3 = 公務員，4 = 商工自営業，5 = 勤務医，6 = 開業医，7 = 自由業，8 = 団体職員，9 = 農林漁業，10 = 無職，11 = その他，0 = 不明
 - (3) 居住住宅種別 CD：1 = 本人持家分譲マンション，2 = 本人持家分譲一戸建て，3 = 本人持家その他，4 = 借家マンション，5 = 借家コーポ，6 = 借家公営住宅，7 = 社宅，8 = 借家一戸建，9 = 借家官舎，10 = 借家その

他, 11=家族持家分譲マンション, 12=家族持家一戸建, 13=家族持家その他, 0=不明

(4) 来場動機 CD: 1=現在住まいの老朽化, 2=現在住まいが狭い, 3=家賃無駄, 4=子供の為, 5=日常生活が不便, 6=結婚, 7=転勤・仕事の都合, 8=投資および資産形成, 9=セカンドハウス, 10=世帯分け, 11=親・子供と同居, 12=その他, 13=現状の間取・設備に不満, 14=現状の周辺環境に不満, 15=現状の通勤・通学が不便, 16=資金面の充実, 0=不明

(5) 家族数: 0=不明

(6) 合算年収: 0=不明, 99=欠損

(7) 購入希望価格 CD: 1=~1,000万円, 2=~1,500万円, 3=~1,800万円, 4=~2,000万円, 5=~2,100万円, 6=~2,200万円, 7=~2,300万円, 8=~2,400万円, 9=~2,500万円, 10=~2,600万円, 11=~2,700万円, 12=~2,800万円, 13=~2,900万円, 14=~3,000万円, 15=~3,100万円, 16=~3,200万円, 17=~3,300万円, 18=~3,400万円, 19=~3,500万円, 20=~3,600万円, 21=~3,700万円, 22=~3,800万円, 23=~3,900万円, 24=~4,000万円, 25=~4,500万円, 26=~5,000万円, 27=~6,000万円, 28=~7,000万円, 29=~8,000万円, 30=~9,000万円, 31=~1億円, 32=1億円以上, 0=不明

4.2 分析結果

アンケート調査データに基づき、開発したデータマイニングシステムを用いて、顧客の購買特性を表す if-then ルールを求めることが出来た。表2は不動産市場における顧客の人口動態変数（年齢、家族構成と所得）と顧客の購買決定変数（希望価格）の関係を表している。これらのルールに基づき、顧客の購買特性を把握し、市場をセグメント化し、標的市場を選択することで、効果的且つ効率的なマーケティング戦略の策定が可能になる。業者との合議により、分析結果は長い間累積した経験と合うルールもあったし、前に気付いていなかった新しい発見もあった。

表2 マーケティングシステムの出力

ルール	if			then
	世帯主年齢	家族数	合算年収	購入希望価格
1	~29	4人以上		2,000万以下
2		4人以上	250~499万	2,000万以下
3	40~49		250~499万	2,000万以下
4	~29	3, 4人	250~499万	2,000万以下
5	40~49	1, 2人	750~999万	2,000万以下
6	30~39	4人以上		2,000万以下
7		4人以上	250~499万	2,000万以下
8	50~	1, 2人	500~749万	~2,500万
9	~29	1, 2人	250~499万	~2,500万
10	40~49	1, 2人	1,000万以上	~2,900万
11	40~49	4人以上		3,000万以上
12		4人以上	1,000万以上	3,000万以上
13	~29		1,000万以上	3,000万以上
14	30~39		1,000万以上	3,000万以上
15	40~49	1, 2人	500~749万	3,000万以上
16	50~	4人以上		3,000万以上
17		4人以上	500~749万	3,000万以上
18	40~49	4人以上		3,000万以上
19		4人以上	500~749万	3,000万以上

5. 結 論

現在 POS データやカードデータ、Web サイトにより得られるデータから、顧客管理情報や売上データ、気象や地震のデータに至るまで、多種多様なデータが溢れている。この大量のデータから利用者のニーズに対し、必要な情報を集め、分析することで知識を発見する手法として、データマイニングがよく用いられている。今まで主に統計学、機械学習などの手法が使われているが、最近では、ラフ集合の研究が進められており、今後応用分野において大きな成果が期待されている [18, 19, 20, 21]。

本研究では、EXCEL VBA を用いて、ラフ集合理論に基づくデータマイニングソフトウェアを開発し、開発されたソフトウェアを用いて、不動産市場のアンケート調査データを分析することにより、大量のデータから顧客の購買特性

を表す 19 個の簡潔な if-then ルールを導き出すことができた。これらのルールにより，市場をセグメント化し，標的市場を選択することで，効果的且つ効率的なマーケティング戦略の策定が可能になると期待される。なお，開発されたシステムはマーケティングへの応用に限らず，他の分野でも適用できる。

謝 辞

本論文は平成 17 年度香川大学萌芽研究「データマイニングシステムの開発及びマーケティングへの応用」として助成を受けて実施しているものである。研究にあたっては（元）香川大学大学院経済学研究科の王波さんの協力，穴吹興産(株)からのアンケート調査データを受けた。ここに謝意を表す。

参 考 文 献

- [1] V. Ananthanarayana, M. Murty and D. Subramanian, Tree structure for efficient data mining using rough sets, *Pattern recognition letters* 24 (2003) 851-862.
- [2] S. Asharaf, S. Shevade and M. Murty, Rough support vector clustering, *Pattern recognition* 38 (2005) 1779-1783.
- [3] T. Beaubouef, R. Ladner and F. Petry, Rough set spatial data modeling for data mining, *International journal of intelligent systems* 19 (2004) 567-584.
- [4] S. Greco, B. Matarazzo and R. Slowinski, Rough sets theory for multicriteria decision analysis, *European journal of operational research* 129 (2001) 1-47.
- [5] P. Lingras and Mofreh Hogo and Miroslav Snorek and Chad West, Temporal analysis of clusters of supermarket customers: conventional versus interval set approach, *Information Sciences* 172 (2005) 215-240.
- [6] Z. Pawlak, Rough set, *International Journal of Computer and Information Science* 11 (1982) 341-356.
- [7] Z. Pawlak, Rough set approach to knowledge-based decision support, *European journal of operational research* 99 (1997) 48-57.
- [8] Z. Pawlak, Rough set theory and its applications to data analysis, *Cybernetics and systems* 29 (1998) 661-688.
- [9] Z. Pawlak, Rough sets and intelligent data analysis, *Information sciences* 147 (2002) 1-12.
- [10] Z. Pawlak, A rough set view on Bayes' theorem, *International journal of intelligent systems* 18 (2003) 487-498.

- [11] L. Polkowski and A. Skowron, (eds.), *Rough Sets in Knowledge Discovery* (Physica-Verlag, Heidelberg, 1998).
- [12] L. Polkowski, S. Tsumoto and T. Y. Lin (eds) : *Rough Set Methods and Applications : New Developments in Knowledge Discovery in Information Systems* (Physica-Verlag, New York, 2000).
- [13] M. Rebolledo, Integrating rough sets and situation-based qualitative models for processes monitoring considering vagueness and uncertainty, *Engineering Applications of Artificial Intelligence*, Volume 18, Issue 5, 2005, Pages 617-632.
- [14] D. Slezak and W. Ziarko, The investigation of the Bayesian rough set model, *International journal of approximate reasoning* 40 (2005) 81-91.
- [15] H. Tanaka and P. Guo, Possibility data analysis and its similarity to rough set concepts, in the book of "Data mining, Rough Sets and Granular Computing", T. Lin, Y. Yao and L. A. Zadeh (Eds) (Physica-Verlag, 518-536, 2002).
- [16] H. Tanaka and P. Guo, Upper and Lower Possibility Distributions with Rough Set concepts, in the book of "Rough Sets Theory and Granular Computing", M. Inuiguchi, S. Hirano and S. Tsumoto (Eds) (Physica-Verlag, 243-250, 2003).
- [17] S. Tsumoto, Automated extraction of medical expert system rules from clinical databases based on rough set theory, *Information Sciences* 112 (1998) 67-84.
- [18] 郭沛俊, 藤村和宏, 「ラフ集合理論による不動産市場データ分析」, 日本経営システム学会第 35 回 (2005 年秋期) 全国研究発表大会論文集, pp. 46-49.
- [19] 郭沛俊, 李曉偉, 「ラフ集合理論による業種特性分析」, 日本経営システム学会第 35 回 (2005 年秋期) 全国研究発表大会論文集, pp. 50-53.
- [20] 郭沛俊, 「ラフ集合理論によるデータマイニングシステムの開発及びマーケティングへの応用」, 日本経営システム学会第 37 回 (2006 年秋期) 全国研究発表大会論文集, pp. 224-227.
- [21] 郭沛俊, 李曉偉, 本田道夫, 「ラフ集合理論による高齢者生活満足度分析について」, 日本経営システム学会第 37 回 (2006 年秋期) 全国研究発表大会論文集, pp. 228-231.
- [22] 田中亨『Excel 2003 完全制覇 関数・数式』翔泳社, 2004 年。
- [23] 田中亨『Excel VBA 完全制覇パーフェクト』翔泳社, 2003 年。
- [24] 河西朝雄『Visual Basic Ver. 6.0 初級プログラミング入門〈上〉』技術評論社, 2000 年。
- [25] 上田太一郎『データマイニング事例集』共立出版株式会社, 1998 年。