多次元正規近似に基づく例外知識の発見

鈴木 英之進†

本論文では，データ集合から信頼性が高い例外知識を発見するアルゴリズムを提案する．例外知識は，よく知られている事実とは異なるパターンと定義され，発見が困難であるが意外性を示し非常に有用となる場合がある．この種の知識を発見する従来の手法では，データ集合から発見された知識の信頼性が評価されていなかった．しかし，このような評価は，信頼性が高い知識と偶然に成立したパターンを，発見システムのユーザに労力をかけることなく区別するためには必要不可欠である．この問題に対処するために，統計的推定手法に基づいて，例外知識と対応する常識知識を，ルールペアの形式で得る手法を提案する．この手法は，多項分布の多次元正規近似に基づいて，一般性，意外性およびデータへの適合度などのすべての条件を確実に満たすルールペアを発見する．発見過程に要する計算時間は，新しく提案する停止条件によって短縮されている．この手法を知識発見システムPEDREとして実装し，実験によってその有効性を確認している．

Discovery of Exception Knowledge Based on Multi-dimensional Normal Approximations

EINOSHIN SUZUKI†

This paper presents an algorithm for discovering exception knowledge from data sets. A piece of exception knowledge, which is defined as a deviational pattern to a well-known fact, exhibits unexpectedness and is sometimes extremely useful in spite of its obscurity. Previous discovery approaches for this type of knowledge have neglected the problem of evaluating the reliability of the knowledge extracted from a data set. It is clear, however, that this question is mandatory in distinguishing reliable knowledge from unreliable patterns without annoying the users. In order to circumvent these difficulties we propose a probabilistic estimation approach in which we obtain a piece of exception knowledge associated with a piece of common sense knowledge in the form of a rule pair. Our approach discovers, based on the normal approximations of the multinomial distributions, rule pairs which satisfy, with high confidence, all the conditions such as the simplicity, the unexpectedness and the goodness-of-fit to the data. The time efficiency of the discovery process is improved by the newly-derived stopping criteria. PEDRE, which is a knowledge discovery system based on our approach, has been validated using the benchmark data sets in the machine learning community.

1. はじめに

データベースからの知識発見（Knowledge Discovery in Databases: KDD）は，データベースとその中に格納されているデータの急速な増加を背景に誕生した研究分野であり，今日ますます世間の関心を集めている[7),10),15)]．KDDにおいて，確率的プロダクションルール[18)]は，高い確率で成立する規則性を表し，KDDが発見対象とする情報のクラスとして，単純だがその一般性のために重要とされている．確率的プロダクションルールは，多数の例について当てはまる常識知識と，比較的少数の例が常識知識とは異なる法則に従うことを表す例外知識に分類することができる[19)~21)]．データベースからの常識知識の発見は，意味情報を用いた問合せ最適化[9)]や，知識ベースの自動構築[18)]などの応用があり，重要である．一方，例外知識は；よく知られている常識知識とは異なることを表すため，意外性を示し，有用となることが多い．たとえば，「助手席に座った身長の低い子供が，シートベルトを装着するのは，事故の際に危険である」という知識は，「助手席に座った人がシートベルトを装着するのは，事故の際に安全である」という常識知識に対する例外知識を表し，数年前に交通事故データから発見されたときには多くの人にとって意外であり，現在でも有用である．また，例外知識は，多数の人間の行動

† 横浜国立大学工学部電子情報工学科
Division of Electrical and Computer Engineering, Faculty of Engineering, Yokohama National University

規範となっている常識知識とは異なることを表すために，有用であることが多い．たとえば，ある種のキノコが，そのほとんどが毒を持つが，食べられる例外も存在するものとする．ここで，食べられる例外となる条件を知ることにより，多数の人間が食べないキノコを独占する利益を得ることができる．

例外知識は比較的少数の例について成立するので，データベースから例外知識を発見する際に最も重要な点の1つは，発見したパターンが本当に信頼できる知識なのか，それともたまたま成立したノイズなのかを区別することである．しかし，従来の知識発見システムにおいては，発見された例外知識についての信頼性評価の問題は考慮されておらず，発見された例外知識をノイズと区別することは，ユーザに任されていた．このことは，EXPLORA[11]やINLEN 2[12]のように領域知識を用いるシステム，KEFIR[16]のように領域固有の評価規準を用いるシステム，MEPRO[19),21)]やMEPROUX[20]のように常識知識と例外知識を一体として発見するシステムすべてに当てはまる．例外知識の信頼性をユーザが評価する方式は，人間の主観的な信頼性判定に依存するために不確実であり，データベースから発見される例外知識が多数となる場合には，ユーザに大きな負担をかけてしまう．さらに，Chernoff bound[2]や二項分布の正規近似[8]などの従来KDDで用いられてきた信頼性評価手法[1),4),13),17)]は，単一確率の信頼区間を推定するためのものであり，条件つき確率の信頼区間推定が必要となる例外知識の信頼性評価には不適当である．

本論文では，この問題に対処するために，多項分布の多次元正規近似に基づき，信頼性を考慮して例外知識を発見する新しい手法を提案する．この手法は，領域知識，領域固有の規準およびユーザに依存しないで例外知識と対応する常識知識を発見し，それらの信頼性を評価する．機械学習の標準問題として用いられているデータ集合を用いた実験の結果，本手法は，信頼性が高い有用な例外知識を効率的に発見できることが確認された．

2. 対象問題

データ集合中に蓄えられているオブジェクトを例 e_i と呼ぶとき，データ集合中には n 個の例 $e_1, e_2, \cdots, e_n$ が保存されているとする．この例 e_i は，m 個の属性についての属性値 $a_{i1}, a_{i2}, \cdots, a_{im}$ から構成されるタプル $\langle a_{i1}, a_{i2}, \cdots, a_{im} \rangle$ で表されている．ただし，実数の属性値をとる連続値属性はすべて，種々の離散化手法[5),6)]によって，離散値の属性値をとる離散値属性に変換されているものとする．また，ある属性が1つの属性値をとる事象を，アトムと呼ぶことにする．

アトムあるいはアトムの連言として表される前提部が成立するときに，アトムとして表される結論部がある確率で成立するプロダクションルールを，確率的プロダクションルール[18]と呼ぶことにする．本論文では，それぞれ確率的プロダクションルールとして表される常識知識と例外知識を一体として発見する問題を考える．常識知識を，アトムあるいはアトムの連言

$$A_\mu \equiv a_1 \wedge a_2 \wedge \cdots \wedge a_\mu \tag{1}$$

とアトム c を用いて，「A_μ ならば c」と表すとする．このとき，例外知識は，常識知識を表すルールの前提部に付加条件を表すアトムあるいはアトムの連言

$$B_\nu \equiv b_1 \wedge b_2 \wedge \cdots \wedge b_\nu \tag{2}$$

が付き，結論部のアトム c' はアトム c と属性は同じだが属性値が異なる「A_μ かつ B_ν ならば c'」で表されることになる．したがって，発見対象を，両者を組にしたルールペア $r(\mu, \nu)$ で表すことにする．

$$r(\mu,\nu) \equiv \begin{cases} A_\mu & \rightarrow \quad c \\ A_\mu \wedge B_\nu & \rightarrow \quad c'. \end{cases} \tag{3}$$

なお，$A_\mu \rightarrow c$ を常識ルール，$A_\mu \wedge B_\nu \rightarrow c'$ を例外ルールと呼ぶことにする．

例からの学習において，学習仮説の評価規準としては，仮説の単純さすなわち一般性と，データへの適合度すなわち適合性が，最も一般的である[18]．確率的プロダクションルール $A_\mu \rightarrow c$ の場合，一般性と適合性は，それぞれ前提部が成立する確率 $p(A_\mu)$ と，前提部が成立するときに結論部が成立する条件つき確率 $p(c|A_\mu)$ に相当する[18]．確率的プロダクションルールの評価は，一般性と適合性を組み合わせた数式で定義される単一の規準を仮定する種々の方法[11),18)]と，一般性と適合性それぞれについてユーザが指定した閾値を上回るルールを求める方法[13)]に分類することができる．ここでは手法の一般性を考えて，後者の方法を用いることにし，常識ルールの一般性と適合性についてそれぞれ閾値 θ_1^{S}, θ_1^{F} を指定し，例外ルールの一般性と適合性についてそれぞれ閾値 θ_2^{S}, θ_2^{F} を指定する．ただし，本研究では信頼性を考えるので，データ集合は実世界をサンプリングした結果得られたものであると考える．したがって，データ集合から点推定によって得られる確率 $\hat{p}(A_\mu)$, $\hat{p}(c|A_\mu)$, $\hat{p}(A_\mu, B_\nu)$, $\hat{p}(c'|A_\mu, B_\nu)$ ではなく，それぞれの真の確率 $p(A_\mu)$, $p(c|A_\mu)$, $p(A_\mu, B_\nu)$, $p(c'|\ A_\mu, B_\nu)$ が，各閾値を信頼度 $1-\delta$ 以上で上回るルールを求めることにする．ただし，確率 $p(c'|B_\nu)$ が大きい場合，例外ルール $A_\mu \wedge B_\nu \rightarrow c'$ は，関連す

るルール $B_\nu \to c'$ から容易に推測できるため，意外性が高いとはいえない．したがって，真に意外な例外ルールを得るために，確率 $p(c'|B_\nu)$ が，閾値を信頼度 $1-\delta$ 以上で下回る条件も指定することにする．

$$\Pr\{p(A_\mu) \geq \theta_1^{\mathrm{S}},\ p(c|A_\mu) \geq \theta_1^{\mathrm{F}},\ p(A_\mu, B_\nu) \geq \theta_2^{\mathrm{S}},\ p(c'|A_\mu, B_\nu) \geq \theta_2^{\mathrm{F}},\ p(c'|B_\nu) \leq \theta_2^{\mathrm{I}}\} \geq 1-\delta. \quad (4)$$

以上より，本論文が対象とする問題は，データ集合から，制約 (4) を満たすルールペア $r(\mu,\nu)$ を求めることとして表せる．

3. 信頼性評価手法

3.1 従来の手法

KDD において，Chernoff bound[2]は，発見したルールの信頼性評価に頻繁に用いられている[1),13),17)]．Chernoff bound によれば，あるアトムが成立する真の確率 p についての $1-\delta$ 信頼区間は，データ集合が n 個の例から構成される場合，次のようになる[17)]．ただし，$\ln(x)$ は x の自然対数であり，$\hat{p}$ は点推定によって得られる確率である．

$$\hat{p} - \sqrt{\frac{1}{n}\ln\left(\frac{2}{\delta}\right)} \leq p \leq \hat{p} + \sqrt{\frac{1}{n}\ln\left(\frac{2}{\delta}\right)}. \quad (5)$$

一方，この問題は，二項分布の正規近似[8)]によっても解くことができ，その結果は KDD における信頼性評価にも用いられている[4)]．この手法によれば，確率 p の $1-\delta$ 信頼区間は，次のようになる．

$$\hat{p} - \alpha_\delta\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + \alpha_\delta\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (6)$$

ただし，α_δ は，

$$\frac{1}{\sqrt{2\pi}}\int_{-\alpha_\delta}^{\alpha_\delta}\exp\left(-\frac{x^2}{2}\right)dx = 1-\delta \quad (7)$$

を満たす値であり，正規分布表などから求めることができる．

ここで，本節で紹介した 2 つの近似手法は，単一確率の信頼区間を推定するためのものであり，条件つき確率は，2 つの確率の商として定義されるので，その推定には適用できないことに注意されたい．言いかえれば，式 (4) は条件つき確率を含むため，Chernoff bound も二項分布の正規近似も，本研究が対象とする例外知識発見問題に用いることはできない．

3.2 多項分布の多次元正規近似

3.1 節より，式 (4) に関連する確率すべてについての信頼区間を同時推定し，信頼範囲を求めればよいことが分かる．本論文では，そのような推定方法として多項分布の多次元正規分布による近似を用いることにする．まず，アトム $D_1, D_2, \cdots, D_8$ を次のように定義する．

$$D_1 \equiv c \wedge A_\mu \wedge B_\nu, \quad (8)$$
$$D_2 \equiv c' \wedge A_\mu \wedge B_\nu, \quad (9)$$
$$D_3 \equiv (\overline{c \vee c'}) \wedge A_\mu \wedge B_\nu, \quad (10)$$
$$D_4 \equiv c \wedge A_\mu \wedge \overline{B_\nu}, \quad (11)$$
$$D_5 \equiv \overline{c} \wedge A_\mu \wedge \overline{B_\nu}, \quad (12)$$
$$D_6 \equiv \overline{c'} \wedge \overline{A_\mu} \wedge B_\nu, \quad (13)$$
$$D_7 \equiv c' \wedge \overline{A_\mu} \wedge B_\nu, \quad (14)$$
$$D_8 \equiv \overline{A_\mu} \wedge \overline{B_\nu}. \quad (15)$$

$\hat{p}(D_i) \neq 0$ となるアトム D_i を，順に $E_1, E_2, \cdots, E_{k+1}$ とおき，それぞれの事象が起きる回数を，$x_1, x_2, \cdots, x_{k+1}$ とおく．事象 $E_1, E_2, \cdots, E_{k+1}$ は互いに排反であるために，それぞれの事象が $(x_1, x_2, \cdots, x_{k+1})$ 回起きる確率は，多項分布に従う．データ集合において各事象を満たす例の数を $u_1, u_2, \cdots, u_{k+1}$ と定義し，$\boldsymbol{G}$ の転置行列を ${}^t\boldsymbol{G}$ で表すことにする．また，

$$\vec{u} \equiv (u_1, u_2, \cdots, u_k) \quad (16)$$
$$\vec{x} \equiv (x_1, x_2, \cdots, x_k) \quad (17)$$

と定義する．以下，n が十分大きいと仮定して，この多項分布を，$\vec{x}$ の同時確率密度関数が次のように表される k 次元正規分布で近似する．

$$f(\vec{x}) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{H}|^{1/2}} \cdot \exp\left\{-\frac{{}^t(\vec{x}-\vec{u})\boldsymbol{H}^{-1}(\vec{x}-\vec{u})}{2}\right\}. \quad (18)$$

ただし，共分散行列 $\boldsymbol{H}$ は，

$$\boldsymbol{H} = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1k} \\ h_{21} & h_{22} & \cdots & h_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ h_{k1} & h_{k2} & \cdots & h_{kk} \end{pmatrix}, \quad (19)$$

$$h_{ij} = \begin{cases} u_i(n-u_i)/n & (i=j \text{ のとき}) \\ -u_iu_j/n & (i \neq j \text{ のとき}) \end{cases} \quad (20)$$

となる．

ここで，k 次元楕円面によって囲まれる範囲

$$V_\delta : {}^t(\vec{x}-\vec{u})\boldsymbol{H}^{-1}(\vec{x}-\vec{u}) \leq \beta(\delta,k)^2 \quad (21)$$

を考える．ただし，

$$\Pr(\vec{x} \in V_\delta) = 1-\delta \quad (22)$$

とする．この楕円面は，$\vec{x}$ の $1-\delta$ 信頼範囲に相当する．なお，正数 $\beta(\delta,k)$ は，付録 A.1 に示す方法で計算できる．また，以下において $\beta(\delta,k)$ を β と略記する．

アトム D を満たす例についての真の数を $x(D)$ と

表すと，式(8)〜(14)より，

$$p(A_\mu)=\frac{\sum_{i=1}^{5}x(D_i)}{n}, \tag{23}$$

$$p(c|A_\mu)=\frac{x(D_1)+x(D_4)}{\sum_{i=1}^{5}x(D_i)}, \tag{24}$$

$$p(A_\mu,B_\nu)=\frac{\sum_{i=1}^{3}x(D_i)}{n}, \tag{25}$$

$$p(c'|A_\mu,B_\nu)=\frac{x(D_2)}{\sum_{i=1}^{3}x(D_i)}, \tag{26}$$

$$p(c'|B_\nu)=\frac{x(D_2)+x(D_7)}{\sum_{i=1}^{3}x(D_i)+\sum_{i=6}^{7}x(D_i)} \tag{27}$$

であるから，式(4)の判定問題は，

$$\frac{\sum_{i=1}^{5}x(D_i)}{n}\geq\theta_1^{\mathrm{S}}, \tag{28}$$

$$\frac{x(D_1)+x(D_4)}{\sum_{i=1}^{5}x(D_i)}\geq\theta_1^{\mathrm{F}}, \tag{29}$$

$$\frac{\sum_{i=1}^{3}x(D_i)}{n}\geq\theta_2^{\mathrm{S}}, \tag{30}$$

$$\frac{x(D_2)}{\sum_{i=1}^{3}x(D_i)}\geq\theta_2^{\mathrm{F}}, \tag{31}$$

$$\frac{x(D_2)+x(D_7)}{\sum_{i=1}^{3}x(D_i)+\sum_{i=6}^{7}x(D_i)}\leq\theta_2^{\mathrm{I}} \tag{32}$$

が，式(21)で表される楕円面内で成立することを判定する問題に帰着される．この問題に対して，付録A.2より，以下の結果を得ている．

$$\left(1-\beta\sqrt{\frac{1-\hat{p}(A_\mu)}{n\hat{p}(A_\mu)}}\right)\hat{p}(A_\mu)\geq\theta_1^{\mathrm{S}}, \tag{33}$$

$$\left(1-\beta\sqrt{\frac{\hat{p}(\overline{c},A_\mu)}{\hat{p}(c,A_\mu)\{(n+\beta^2)\hat{p}(A_\mu)-\beta^2\}}}\right)\cdot\hat{p}(c|A_\mu)\geq\theta_1^{\mathrm{F}}, \tag{34}$$

$$\left(1-\beta\sqrt{\frac{1-\hat{p}(A_\mu,B_\nu)}{n\hat{p}(A_\mu,B_\nu)}}\right)\hat{p}(A_\mu,B_\nu)\geq\theta_2^{\mathrm{S}}, \tag{35}$$

$$\left(1-\beta\sqrt{\frac{\hat{p}(\overline{c'},A_\mu,B_\nu)}{\hat{p}(c',A_\mu,B_\nu)\{(n+\beta^2)\hat{p}(A_\mu,B_\nu)-\beta^2\}}}\right)\cdot\hat{p}(c'|A_\mu,B_\nu)\geq\theta_2^{\mathrm{F}}, \tag{36}$$

$$\left(1+\beta\sqrt{\frac{\hat{p}(\overline{c'},B_\nu)}{\hat{p}(c',B_\nu)\{(n+\beta^2)\hat{p}(B_\nu)-\beta^2\}}}\right)\cdot\hat{p}(c'|B_\nu)\leq\theta_2^{\mathrm{I}}. \tag{37}$$

したがって，本研究が対象とする問題は，式(33)〜(37)をすべて満たすルールペア $r(\mu,\nu)$ を求めることに帰着された．

4. 発見アルゴリズム

発見アルゴリズムでは，データベースからの知識発見を，式(3)で与えられるルールペア $r(\mu,\nu)$ を表すノードから構成される探索木についての探索問題として考える．ここで，$\mu=0$, $\nu=0$ はそれぞれ，前提部が a_i あるいは b_i を含まない場合である．$\mu=\nu=0$ の場合を深さ1のノードとし，以降探索木において深さが1増すごとに，常識ルールあるいは例外ルールの前提部にアトムを1つ加えることになる．深さ2のノードでは $\mu=1$, $\nu=0$ であり，深さ l (≥ 3) では，$\mu+\nu=l-1$ $(\mu,\nu\geq 1)$ である．

探索法としては深さ優先探索法を用い，μ, ν の最大値 M はユーザが指定するものとする．ただし，単純な深さ優先探索法では多数のノードを調べる必要があるので，以下の定理における式(38)〜(42)を探索の停止条件として用いている．

定理1 現在探索中のノードが表すルールペア $r(\mu',\nu')$ が，式(38)〜(42)のいずれかを満たす場合，子孫ノードが表すルールペア $r(\mu,\nu)$ は，式(33)〜(37)をすべて満たすことはない．

$$\left(1-\beta\sqrt{\frac{1-\hat{p}(A_{\mu'})}{n\hat{p}(A_{\mu'})}}\right)\hat{p}(A_{\mu'})<\theta_1^{\mathrm{S}}, \tag{38}$$

$$\left(1-\beta\sqrt{\frac{1-\hat{p}(A_{\mu'})}{n\hat{p}(A_{\mu'})}}\right)\hat{p}(c,A_{\mu'})<\theta_1^{\mathrm{S}}\theta_1^{\mathrm{F}}, \tag{39}$$

$$\left(1-\beta\sqrt{\frac{1-\hat{p}(A_{\mu'},B_{\nu'})}{n\hat{p}(A_{\mu'},B_{\nu'})}}\right)\cdot\hat{p}(A_{\mu'},B_{\nu'})<\theta_2^{\mathrm{S}}, \tag{40}$$

$$\left(1-\beta\sqrt{\frac{1-\hat{p}(A_{\mu'},B_{\nu'})}{n\hat{p}(A_{\mu'},B_{\nu'})}}\right)\cdot\hat{p}(c',A_{\mu'},B_{\nu'})<\theta_2^{\mathrm{S}}\theta_2^{\mathrm{F}}, \tag{41}$$

$$\left(1-\beta\sqrt{\frac{1-\hat{p}(A_{\mu'},B_{\nu'})}{n\hat{p}(A_{\mu'},B_{\nu'})}}\right)\hat{p}(B_{\nu'})<\frac{\theta_2^{\mathrm{S}}\theta_2^{\mathrm{F}}}{\theta_2^{\mathrm{I}}}. \tag{42}$$

証明 ルールペア $r(\mu,\nu)$ が，式(33)〜(37)をすべて満たすと仮定する．まず，$n,\beta>0$ より，関数 $(1-\beta\sqrt{(1-x)/n/x})$ は，正数 x に関して単調増加する．式(33)と $\hat{p}(A_{\mu'})\geq\hat{p}(A_\mu)$ より，(式(38)の左辺) $\geq(1-\beta\sqrt{(1-\hat{p}(A_\mu))/n/\hat{p}(A_\mu)})\hat{p}(A_\mu)\geq\theta_1^{\mathrm{S}}$ となるが，これは式(38)と矛盾する．同様に，式(35)は，式(40)と矛盾する．次に，式(33)，(34)と $\hat{p}(c,A_{\mu'})\geq\hat{p}(c,A_\mu)$ より，(式(39)の左辺) $\geq(1-\beta\sqrt{(1-\hat{p}(A_\mu))/n/\hat{p}(A_\mu)})\hat{p}(c,A_\mu)\geq\theta_1^{\mathrm{S}}\hat{p}(c|A_\mu)\geq\theta_1^{\mathrm{S}}\theta_1^{\mathrm{F}}$ となるが，これは式(39)と矛盾する．同様に，式(41)は式(35)，(36)と，式(42)は式(35)〜(37)

表1 PEDREによって国勢調査データ集合から発見されたルールペアと，関連するルール．結論部の属性は収入クラスに限定されている．

Table 1 The rule pair with its associated reference rule discovered by PEDRE from the census data set, where the salary class is the only attribute allowed in the conclusions.

No.	常識ルール 例外ルール 関連ルール	(min) A (min)	$c \wedge A$	(min) $\hat{p}(c\|A)$ (min) (max)
	relationship=Husband → class $\leq 50K$	19,716 (19,444)	10,870	0.551 (0.542)
1	$\mathcal{C}$, workclass=Private, occupation=Exec-managerial → class $> 50K$	1,933 (1,792)	1,417	0.733 (0.700)
	workclass=Private, occupation=Exec-managerial → class $> 50K$	3,995	1,889	0.473 (0.499)

と矛盾する． □

この定理に基づき，式(38)〜(42)のいずれかを満たすノードを展開しないことにより，発見されるルールペアの集合を変えることなく探索の効率化を実現している．

5. データ集合への適用

本論文で提案した手法に基づく知識発見システム PEDRE（Probabilistic Estimation-based Data mining system for Reliable Exceptions）を構築し，種々の実験を行った．ここではPEDREを，機械学習における標準問題[14]である，国勢調査データ集合とマッシュルームデータ集合に適用した結果について述べる．

国勢調査データ集合とは，米国国勢調査局が，合衆国に住む48,842人について，年収が5万ドルを超えるかを14個の属性について記述したデータ集合である．各属性は，連続値属性を最小記述長原理に基づく離散化手法[6]で離散化した結果，2個から41個の属性値をとる．ここでは結論部の属性として年収を表すクラスを考え，$M = 2$, $\delta = 0.1$, $\theta_1^{\mathrm{S}} = 0.3$, $\theta_2^{\mathrm{S}} = 0.03$, $\theta_1^{\mathrm{F}} = 0.5$, $\theta_2^{\mathrm{F}} = 0.7$, $\theta_2^{\mathrm{I}} = 0.5$ の場合について実験を行った．その結果を**表1**に示す．ただし，表中の $\mathcal{C}$, A, $c \wedge A$ は，それぞれ常識ルールの前提部，前提部 (A) を満たす例の数，前提部と結論部 (c) をともに満たす例の数を表す．また，min と max は，それぞれ左にある値の，信頼度 $1-\delta$ での下限値と上限値を表す．

国勢調査データ集合より得られたルールペアは，比較的興味深い例外性を示すことが表1より分かる．表中のルールペアは，属性 “relationship” が “Husband” である者 19,716 人の 55.1%は収入が5万ドル以下であるが，この中で属性 “workclass” が “Private”, “occupation” が “Exec-managerial” である 1,933 人の 73.3%は収入が5万ドルより多いことを示している．それぞれのルールは，信頼度90%で，少なくとも 19,444 人あるいは 1,792 人について条件つき確率 54.2%あるいは 70.0%以上で成立する．ここで，“workclass” が “Private”, “occupation” が “Exec-managerial” であっても収入が5万ドルより多い人は 3,995 人中の 47.3%にすぎず，これは信頼度90%で 49.9%より少ない．したがって，発見された例外知識は意外性を示すことが分かる．なお，$\theta_1^{\mathrm{F}} = 0.6$, $\theta_2^{\mathrm{F}} = 0.8$ と定め，$\theta_2^{\mathrm{S}} \geq 0.01$ となる種々の条件下において実験を行ったが，知識は発見されなかった．その理由は，収入に関して例外性があれば，それを平均化する強い圧力が働くという，国勢調査データ集合が対象とする問題の性質であると考えられる．

2番目の標準問題であるマッシュルームデータ集合とは，北アメリカに自生するキノコ 8,124 本について，毒の有無を22個の属性で記述したデータ集合である．各属性は2個から12個の属性値をとる．ここでは結論部の属性として毒の有無を表すクラスを考え，$M = 3$, $\delta = 0.1$, $\theta_1^{\mathrm{S}} = 0.2$, $\theta_2^{\mathrm{S}} = 0.05$, $\theta_1^{\mathrm{F}} = 0.7$, $\theta_2^{\mathrm{F}} = 1.0$, $\theta_2^{\mathrm{I}} = 0.5$ の場合について実験を行った．その結果を**表2**に示す．

発見されたルールペアは，非常に興味深い例外性を示すことが表2より分かる．たとえば1番目のルールペアは，属性 “bruises” が “f”, “g-size” が “b”, “stalk-shape” が “e” であるキノコの 72.9%は毒性があるが，この中で属性 “stalk-root” が “?” であるものは 100%毒性がないことを示している．それぞれのルールは，信頼度90%で，少なくとも 1,734 本あるいは 415 本のキノコについて条件つき確率 70.3%あるいは 100%以上で成立する．なお，“stalk-root” が “?” であっても毒性がないキノコは 29.0%にすぎず，これは信頼度90%で 31.8%より少ない．したがって，発見された例外知識は真に意外であることが分かる．

μ, ν の最大値 M は，前提部が多数のアトムから構成されるルールペアを調べられるように，十分大きくなければならない．しかし，ルールペアの数は，深さ優先探索においては深さに対して指数関数的に増加する．以下，このような非効率性を改善するうえで，停止条件が有効であることを実験によって示す．

表2 PEDREによってマッシュルームデータ集合から発見されたルールペアと，関連するルール．結論部の属性は毒性クラスに限定されている．

Table 2 The rule pairs with their associated reference rules discovered by PEDRE from the mushroom data set, where the edibility class is the only attribute allowed in the conclusions.

No.	常識ルール 例外ルール 関連ルール	(min) A (min)	$c \wedge A$	(min) $\hat{p}(c\|A)$ (min) (max)
1	bruises=f, g-size=b, stalk-shape=e → class=p	1,828 (1,734)	1,332	0.729 (0.703)
	$\mathcal{C}$, stalk-root=? → class=e	480 (415)	480	1.000 (1.000)
	stalk-root=? → class=e	2,480	720	0.290 (0.318)
2	g-attachment=f, stalk-root=? → class=p	2,288 (2,187)	1,760	0.769 (0.747)
	$\mathcal{C}$, g-size=b, stalk-shape=e, veil-color=w → class=e	480 (415)	480	1.000 (1.000)
	g-size=b, stalk-shape=e, veil-color=w → class=e	2,636	1,232	0.467 (0.497)
3	stalk-root=?, sp-color=w → class=p	2,240 (2,139)	1,760	0.786 (0.764)
	$\mathcal{C}$, g-attachment=f, g-size=b, stalk-shape=e → class=e	480 (421)	480	1.000 (1.000)
	g-attachment=f, g-size=b, stalk-shape=e → class=e	2,618	1,232	0.471 (0.498)
4	stalk-root=?, sp-color=w → class=p	2,240 (2,139)	1,760	0.786 (0.764)
	$\mathcal{C}$, g-size=b, stalk-shape=e, veil-color=w → class=e	480 (421)	480	1.000 (1.000)
	g-size=b, stalk-shape=e, veil-color=w → class=e	2,636	1,232	0.467 (0.495)

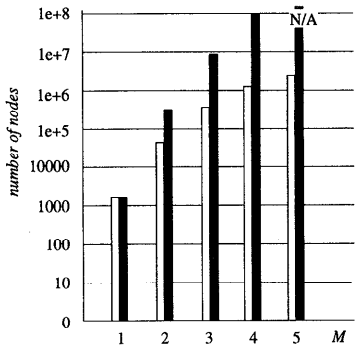

図1 停止条件を用いた場合（白）と用いない場合（黒）の比較

Fig. 1 Performance of PEDRE with/without the stopping criteria (white/black bars).

図1は，マッシュルームデータ集合において，停止条件を用いた場合と用いない場合に探索するノード数を，表2を求めた実験と M 以外は同じ条件のもとで，底が10の対数表示で M に関して示したものである．停止条件を用いた場合のノード数は白い棒グラフで，用いない場合のノード数は黒い棒グラフで表されている．実験にはPentium Pro 200 MHz搭載のパーソナルコンピュータを用い，計算時間はノード数が 10^8 個程度で約9日間であった．$M = 5$ で停止条件を用いない場合には，9日間経過しても計算が終了しなかったためN/Aと表記している．図より，停止条件は，探索の時間的効率を改善するうえで有効であることが分かる．たとえば，$M = 5$ で停止条件を用いた場合は，$M = 3$ で停止条件を用いない場合よりも約3.9倍効率的であり，$M = 4$ では停止条件により約70倍の高速化が実現されている．この実験より，停止条件は例外知識の効率的な発見に必要不可欠であると考えられる．

6. おわりに

本論文では，統計的推定手法に基づき，信頼性が高い例外知識を発見する手法を提案した．従来の手法は，ユーザによる信頼性評価に依存するか，単一確率の信頼区間しか考慮しないために，例外知識のように比較的少数の例について成立する知識を発見する際には，偶然成立したパターンを知識と間違って見なしてしまう問題があった．本論文で提案した手法では，多項分布の多次元正規近似に基づき，複数の変数についての信頼範囲を同時推定することでこの問題を解決している．すなわち，本手法は，主観的な信頼性判定に依存せず，ユーザが指定した条件を確実に満たす例外ルールだけを発見する．もちろん，本研究は主観的な信頼性判定や領域知識の利用を完全に否定するものではなく，ユーザによる発見結果の再評価や，制約による探索空間の限定などによって，容易にこれらを実現することができる．また，本手法は，同時推定を計算が容易な解析解に基づいて行うことと，発見結果を変えないで探索木を枝刈りする新しい停止条件を用いることにより，高速な発見過程を実現することにも成功している．

提案した手法は，例外知識発見システムPEDREとして実装され，機械学習におけるいくつかの標準問題に適用されている．実験の結果，PEDREは，信頼性が高く興味深い例外知識を効率的に発見できることが確認されている．PEDREは，このようにすぐれた性能を有しており，特に主観的な信頼性判定や領域知識の利用が困難であるデータベースにおける知識発見に有効である．また，PEDREは，データ集合中に強い例外性が存在しないことの確認や，既知の例外知識の再評価にも用いることができ，主観的な信頼性判定や

領域知識の利用が原因で見落とした有用な例外知識の発見にも有効であると考えられる．

参考文献

1) Agrawal, R., Mannila, H., Srikant, R., Toivonen, H. and Verkamo, A.I.: Fast Discovery of Association Rules, *Advances in Knowledge Discovery and Data Mining*, Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (Eds), pp.307–328, AAAI Press/MIT Press (1996).
2) Alon, N., Spencer, J.H. and Erdös, P.: *The Probabilistic Method*, John Wiley & Sons (1992).
3) Cramér, H.: *Mathematical Methods of Statistics*, Princeton Univ. Press (1966).
4) Chan, K.C.C. and Wong, A.K.C.: A Statistical Technique for Extracting Classificatory Knowledge from Databases, *Knowledge Discovery in Databases*, Piatetsky-Shapiro, G. and Frawley, W.J. (Eds), pp.107–123, AAAI Press/MIT Press (1991).
5) Dougherty, J., Kohavi, R. and Sahami, M.: Supervised and Unsupervised Discretization of Continuous Features, *Proc. 12th International Conference on Machine Learning*, pp.194–202 (1995).
6) Fayyad, U.M. and Irani, K.B.: Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning, *Proc. 13th International Joint Conference on Artificial Intelligence*, pp.1022–1027 (1993).
7) Fayyad, U.M., Piatetsky-Shapiro, G. and Smyth, P.: From Data Mining to Knowledge Discovery: An Overview, *Advances in Knowledge Discovery and Data Mining*, Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (Eds), pp.1–34, AAAI Press/MIT Press (1996).
8) Feller, W.: *An Introduction to Probability Theory and Its Applications Volume 1*, John Wiley & Sons (1957).
9) Hsu, C. and Knoblock, C.A.: Using Inductive Learning to Generate Rules for Semantic Query Optimization, *Advances in Knowledge Discovery and Data Mining*, Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (Eds), pp.425–445, AAAI Press/MIT Press (1996).
10) 河野浩之，西尾章治郎，Han, J.：データベースからの知識獲得技術，人工知能学会誌，Vol.10, No.1, pp.38–44 (1995).
11) Klösgen, W.: Explora: A Multipattern and Multistrategy Discovery Approach, *Advances in Knowledge Discovery and Data Mining*, Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (Eds), pp.249–271, AAAI Press/MIT Press (1996).
12) Michalski, R.S.: Multistategy Data Mining, *Proc. 4th International Workshop on Rough Sets, Fuzzy Sets, and Machine Discovery*, p.9 (1996).
13) Mannila, H., Toivonen, H. and Verkamo, A.I.: Efficient Algorithms for Discovering Association Rules, *AAAI-94 Workshop on Knowledge Discovery in Databases*, Fayyad, U.M. and Uthurusamy, R. (Eds), pp.181–192 (1994).
14) Murphy, P.M. and Aha, D.W.: UCI Repository of Machine Learning Databases, http://www.ics.uci.edu/~mlearn/MLRepository.html, Dept. of Information and Computer Sci., Univ. of California (1996).
15) 西尾章治郎：大規模データベースにおける知識獲得，情報処理，Vol.34, No.3, pp.343–350 (1995).
16) Piatetsky-Shapiro, G. and Matheus, C.J.: The Interestingness of Deviations, *AAAI-94 Workshop on Knowledge Discovery in Databases*, Fayyad, U.M. and Uthurusamy, R. (Eds), pp.25–36 (1994).
17) Siebes, A.: Homogeneous Discoveries Contain No Surprises: Inferring Risk-profiles from Large Databases, *AAAI-94 Workshop on Knowledge Discovery in Databases*, Fayyad, U.M. and Uthurusamy, R. (Eds), pp.97–107 (1994).
18) Smyth, P. and Goodman, R.M.: An Information Theoretic Approach to Rule Induction from Databases, *IEEE Trans. Knowledge and Data Eng.*, Vol.4, No.4, pp.301–316 (1992).
19) Suzuki, E. and Shimura, M.: Exceptional Knowledge Discovery in Databases Based on Information Theory, *Proc. 2nd International Conference on Knowledge Discovery and Data Mining*, pp.275–278 (1996).
20) Suzuki, E.: Discovering Unexpected Exceptions: A Stochastic Approach, *Proc. 4th International Workshop on Rough Sets, Fuzzy Sets, and Machine Discovery*, pp.225–232 (1996).
21) 鈴木英之進，志村正道：情報理論的手法を用いたデータベースからの例外的知識の発見，人工知能学会誌，Vol.12, No.2, pp.305–312 (1997).

付　録

A.1 β の計算

3.2 節の β を計算するアルゴリズムについて説明する．まず，k 次元楕円面

$$^t(\vec{x}-\vec{u})\boldsymbol{H}(\vec{x}-\vec{u}) \leq \beta^2 \qquad (43)$$

の体積 $V(k,\boldsymbol{H},\beta)$ は，次のようになる[3]．ただし，$\boldsymbol{H}$ は，$k\times k$ 正則行列とする．

$$V(k,\boldsymbol{H},\beta)=\frac{\pi^{k/2}}{\Gamma(k/2+1)}\frac{\beta^k}{|\boldsymbol{H}|^{1/2}}. \tag{44}$$

ここで，$\Gamma()$ はガンマ関数を表し，次の関係を用いて計算することができる．

$$\Gamma(1)=1, \tag{45}$$

$$\Gamma(1/2)=\sqrt{\pi}, \tag{46}$$

$$\Gamma(n+1)=n\Gamma(n). \tag{47}$$

$|\boldsymbol{H}|^{1/2}|\boldsymbol{H}^{-1}|^{1/2}=1$ となることを用いると，式 (18)，(44) より，次の関係が得られる．

$$V(k,\boldsymbol{H}^{-1},\beta)f(\vec{x})=\frac{\beta^k}{2^{k/2}\Gamma(k/2+1)}\cdot\exp\left(-\frac{{}^t(\vec{x}-\vec{u})\boldsymbol{H}^{-1}(\vec{x}-\vec{u})}{2}\right). \tag{48}$$

式 (21)，(43)，(48) より，β は，次のように数値積分によって求めることができる．

手続き: obtain_$\beta(\delta,k,\beta)$

入力: δ, k.

出力: β.

パラメータ: 十分小さい正数 $\Delta\beta$, 体積 V.

begin

V:=0, β:=0.

while $(V<1-\delta)$

begin

$V:=V+\frac{(\beta+\Delta\beta)^k-\beta^k}{2^{k/2}\Gamma(k/2+1)}\exp(-\frac{\beta^2}{2})$.

$\beta:=\beta+\Delta\beta$.

end

$\beta:=\beta-\Delta\beta$.

end

A.2 式 (33)〜(37) の導出

3.2 節より，式 (33)〜(37) が成立するときに，式 (28)〜(32) が式 (21) で表される楕円面内 V_δ でつねに成立すればよい．以下，式 (28)〜(31) の左辺が楕円面内でとりうる最小値 S_1,F_1,S_2,F_2 と，式 (32) の左辺が楕円面内でとりうる最大値 I_2 を求め，これらがそれぞれ式 (33)〜(37) の左辺に一致することを示す．ただし，$\forall\,\hat{p}(D_i)\neq 0\ (i=1,2,\cdots,8)$ すなわち $k=7$，$\forall u_i\neq 0\ (i=1,2,\cdots,8)$ の場合だけを扱い，同様にして証明できる $\exists\hat{p}(D_i)=0\ (i=1,2,\cdots,8)$ の場合は省略する．

まず，楕円面の式を整理する．式 (19)，(20) より，共分散行列の逆行列 $\boldsymbol{H}^{-1}$ は，

$$\boldsymbol{H}^{-1}=\frac{1}{n-\sum_{i=1}^{7}u_i}\cdot\begin{pmatrix}h'_{11}&h'_{12}&\cdots&h'_{17}\\h'_{21}&h'_{22}&\cdots&h'_{27}\\\vdots&\vdots&\ddots&\vdots\\h'_{71}&h'_{72}&\cdots&h'_{77}\end{pmatrix}, \tag{49}$$

$$h'_{ij}=\begin{cases}\left(n-\sum_{l=1(l\neq i)}^{7}u_l\right)\Big/u_i & (i=j\ \text{のとき})\\ 1 & (i\neq j\ \text{のとき})\end{cases} \tag{50}$$

となる．式 (21)，(49)，(50) より V_δ は，次のように表される．

$$\frac{1}{n-\sum_{i=1}^{7}u_i}\left\{\sum_{i=1}^{7}\frac{n-\sum_{j=1(j\neq i)}^{7}u_j}{u_i}(x_i-u_i)^2+\sum_{i=1}^{7}\sum_{j=1(j\neq i)}^{7}(x_i-u_i)(x_j-u_j)\right\}\leq\beta^2. \tag{51}$$

ここで，式 (28)〜(32) の各左辺の表現を一定とおいた式はすべて平面を表しており，楕円面内での各表現の最大・最小値は，楕円面上での極値点で起こることが容易に分かる．ラグランジュの未定乗数法によれば，関数 $g=0$ のもとで，関数 f の極値点は，

$$\lambda\left(\frac{\partial g}{\partial x_1}\frac{\partial g}{\partial x_2}\cdots\frac{\partial g}{\partial x_7}\right)=\left(\frac{\partial f}{\partial x_1}\frac{\partial f}{\partial x_2}\cdots\frac{\partial f}{\partial x_7}\right) \tag{52}$$

を満たす．f を式 (51) の楕円面，g を式 (28) の左辺とおいて，ラグランジュの未定乗数法を適用して整理すると，次の式を得る．

$$\begin{cases}\frac{\lambda}{2n} &= \frac{n-\sum_{j=1(j\neq i)}^{7}u_j}{u_i}(x_i-u_i)+\sum_{j=1(j\neq i)}^{7}(x_j-u_j)\\ &\quad (i=1,2,\cdots,5)\\ 0 &= \frac{n-\sum_{j=1(j\neq i)}^{7}u_j}{u_i}(x_i-u_i)+\sum_{j=1(j\neq i)}^{7}(x_j-u_j)\\ &\quad (i=6,7).\end{cases} \tag{53}$$

式 (53) から λ を消去し，各辺を (x_2-u_2) について解くと，

$$\begin{cases}(x_i-u_i) &= \frac{u_i}{u_2}(x_2-u_2)\\ &\quad (i=1,3,4,5)\\ (x_i-u_i) &= -\frac{u_i}{u_2}\frac{\sum_{j=1}^{5}u_j}{n-\sum_{j=1}^{5}u_j}(x_2-u_2)\\ &\quad (i=6,7).\end{cases} \tag{54}$$

式 (51) の楕円面に代入して x_2 について解き，式 (54) を用いて x_1, x_3, x_4, x_5 も求めると，

$$x_i = u_i \left(1 \pm \beta \sqrt{\frac{n - \sum_{j=1}^{5} u_j}{n \sum_{j=1}^{5} u_j}}\right)$$
$$(i = 1, 2, \cdots, 5) \text{ (複号同順)}. \qquad (55)$$

したがって，

$$S_1 = \frac{\sum_{i=1}^{5} u_i}{n} \left(1 - \beta \sqrt{\frac{n - \sum_{i=1}^{5} u_i}{n \sum_{i=1}^{5} u_i}}\right)$$
$$= \left(1 - \beta \sqrt{\frac{1 - \hat{p}(A_\mu)}{n \hat{p}(A_\mu)}}\right) \hat{p}(A_\mu) \qquad (56)$$

となり，これは式 (33) の左辺に一致する．

同様に，f を式 (51) の楕円面，g を式 (30) の左辺とおいて，ラグランジュの未定乗数法を適用し，x_1, x_2, x_3 を求めると，

$$x_i = u_i \left(1 \pm \beta \sqrt{\frac{n - \sum_{j=1}^{3} u_j}{n \sum_{j=1}^{3} u_j}}\right)$$
$$(i = 1, 2, 3) \text{ (複号同順)}. \qquad (57)$$

したがって，

$$S_2 = \frac{\sum_{i=1}^{3} u_i}{n} \left(1 - \beta \sqrt{\frac{n - \sum_{i=1}^{3} u_i}{n \sum_{i=1}^{3} u_i}}\right)$$
$$= \left(1 - \beta \sqrt{\frac{1 - \hat{p}(A_\mu, B_\nu)}{n \hat{p}(A_\mu, B_\nu)}}\right) \hat{p}(A_\mu, B_\nu) \qquad (58)$$

となり，これは式 (35) の左辺に一致する．

次に，f を式 (51) の楕円面，g を式 (29) の左辺とおいて，ラグランジュの未定乗数法を適用して整理すると，次の式を得る．

$$\begin{cases} \frac{\lambda(x_2+x_3+x_5)}{2\left(\sum_{j=1}^{5} x_j\right)^2} = \frac{n - \sum_{j=1(j\neq i)}^{7} u_j}{u_i}(x_i - u_i) + \sum_{j=1(j\neq i)}^{7}(x_j - u_j) \\ \qquad (i = 1, 4) \\ \frac{-\lambda(x_1+x_4)}{2\left(\sum_{j=1}^{5} x_j\right)^2} = \frac{n - \sum_{j=1(j\neq i)}^{7} u_j}{u_i}(x_i - u_i) + \sum_{j=1(j\neq i)}^{7}(x_j - u_j) \\ \qquad (i = 2, 3, 5) \\ 0 = \frac{n - \sum_{j=1(j\neq i)}^{7} u_j}{u_i}(x_i - u_i) + \sum_{j=1(j\neq i)}^{7}(x_j - u_j). \\ \qquad (i = 6, 7). \end{cases} \qquad (59)$$

式 (59) から λ を消去し，各辺を x_2, x_4 について解くと，

$$\begin{cases} x_1 = \frac{u_1}{u_4} x_4 \\ x_i = \frac{u_i}{u_2} x_2 \\ \qquad (i = 3, 5) \\ -\frac{x_2+x_3+x_5}{x_1+x_4} = \left\{ \frac{n - \sum_{i=1(i\neq 4)}^{7} u_i}{u_4}(x_4 - u_4) + \sum_{i=1(i\neq 4)}^{7}(x_i - u_i) \right\} \\ \qquad \Big/ \left\{ \frac{n - \sum_{i=1(i\neq 2)}^{7} u_i}{u_2}(x_2 - u_2) + \sum_{i=1(i\neq 2)}^{7}(x_i - u_i) \right\} \\ x_i = u_i - \frac{u_j}{n - \sum_{j=1}^{5} u_j} \cdot \left(\frac{u_2+u_3+u_5}{u_2} x_2 + \frac{u_1+u_4}{u_4} x_4 - \sum_{j=1}^{5} u_j \right) \\ \qquad (i = 6, 7). \end{cases} \qquad (60)$$

式 (51) の楕円面と，式 (60) より $x_1, x_2, x_3, x_5, x_6, x_7$ を消去して x_4 について解き，式 (60) を用いて x_1, x_2, x_3, x_5 も求めると，

$$\begin{cases} x_i = \frac{u_i \left\{(n+\beta^2)\sum_{j=1}^{5} u_j - n\beta^2\right\}}{n \sum_{j=1}^{5} u_j} \left(1 \pm \beta \cdot \sqrt{\frac{n(u_2+u_3+u_5)}{(u_1+u_4)\left\{(n+\beta^2)\sum_{j=1}^{5} u_j - n\beta^2\right\}}}\right) \\ \qquad (i = 1, 4). \\ x_i = \frac{u_i \left\{(n+\beta^2)\sum_{j=1}^{5} u_j - n\beta^2\right\}}{n \sum_{j=1}^{5} u_j} \left(1 \mp \beta \cdot \sqrt{\frac{n(u_1+u_4)}{(u_2+u_3+u_5)\left\{(n+\beta^2)\sum_{j=1}^{5} u_j - n\beta^2\right\}}}\right) \\ \qquad (i = 2, 3, 5) \text{ (複号同順)}. \end{cases} \qquad (61)$$

したがって，

$$F_1 = \frac{u_1 + u_4}{\sum_{i=1}^{5} u_i} \left(1 - \beta \cdot \sqrt{\frac{n(u_2 + u_3 + u_5)}{(u_1 + u_4)\left\{(n + \beta^2)\sum_{i=1}^{5} u_i - n\beta^2\right\}}}\right)$$
$$= \left(1 - \beta \sqrt{\frac{\hat{p}(\bar{c}, A_\mu)}{\hat{p}(c, A_\mu)\{(n + \beta^2)\hat{p}(A_\mu) - \beta^2\}}}\right) \cdot \hat{p}(c|A_\mu) \qquad (62)$$

となり，これは式 (34) の左辺に一致する．

同様に，f を式 (51) の楕円面，g を式 (31) の左辺と

おいて，ラグランジュの未定乗数法を適用し，x_1, x_2, x_3 を求めると，

$$\begin{cases} x_i = \dfrac{u_i\left\{(n+\beta^2)\sum_{j=1}^{3} u_j - n\beta^2\right\}}{n\sum_{j=1}^{3} u_j}\left(1 \pm \beta \cdot\sqrt{\dfrac{nu_2}{(u_1+u_3)\left\{(n+\beta^2)\sum_{j=1}^{3} u_j - n\beta^2\right\}}}\right) \\ \quad (i=1,3). \\ x_2 = \dfrac{u_2\left\{(n+\beta^2)\sum_{i=1}^{3} u_i - n\beta^2\right\}}{n\sum_{i=1}^{3} u_i}\left(1 \mp \beta \cdot\sqrt{\dfrac{n(u_1+u_3)}{u_2\left\{(n+\beta^2)\sum_{i=1}^{3} u_i - n\beta^2\right\}}}\right) \\ \quad (\text{複号同順}). \end{cases} \tag{63}$$

したがって，

$$\begin{aligned} F_2 &= \frac{u_2}{\sum_{i=1}^{3} u_i}\left(1-\beta \cdot\sqrt{\frac{n(u_1+u_3)}{u_2\left\{(n+\beta^2)\sum_{i=1}^{3} u_i - n\beta^2\right\}}}\right) \\ &= \left(1-\beta \cdot\sqrt{\frac{\hat{p}(\overline{c'}, A_\mu, B_\nu)}{\hat{p}(c', A_\mu, B_\nu)\{(n+\beta^2)\hat{p}(A_\mu, B_\nu) - \beta^2\}}}\right) \cdot \hat{p}(c'|A_\mu, B_\nu) \end{aligned} \tag{64}$$

となり，これは式 (36) の左辺に一致する．

また同様に，f を式 (51) の楕円面，g を式 (32) の左辺とおいて，ラグランジュの未定乗数法を適用し，x_1, x_2, x_3, x_6, x_7 を求めると，

$$\begin{cases} x_i = \dfrac{u_i\left\{(n+\beta^2)(u_1+u_3+u_6+u_2+u_7) - n\beta^2\right\}}{n(u_1+u_3+u_6+u_2+u_7)}\left(1 \pm \beta \cdot\sqrt{\dfrac{n(u_1+u_3+u_6)}{(u_2+u_7)\left\{(n+\beta^2)(u_1+u_3+u_6+u_2+u_7) - n\beta^2\right\}}}\right) \\ \quad (i=2,7). \\ x_i = \dfrac{u_i\left\{(n+\beta^2)(u_1+u_3+u_6+u_2+u_7) - n\beta^2\right\}}{n(u_1+u_3+u_6+u_2+u_7)}\left(1 \mp \beta \cdot\sqrt{\dfrac{n(u_2+u_7)}{(u_1+u_3+u_6)\left\{(n+\beta^2)(u_1+u_3+u_6+u_2+u_7) - n\beta^2\right\}}}\right) \\ \quad (i=1,3,6)\ (\text{複号同順}). \end{cases} \tag{65}$$

したがって，

$$\begin{aligned} I_2 &= \frac{u_2+u_7}{u_1+u_3+u_6+u_2+u_7}\left(1+\beta \cdot\sqrt{\frac{n(u_1+u_3+u_6)}{(u_2+u_7)\left\{(n+\beta^2)(u_1+u_3+u_6+u_2+u_7) - n\beta^2\right\}}}\right) \\ &= \left(1+\beta\sqrt{\frac{\hat{p}(\overline{c'}, B_\nu)}{\hat{p}(c', B_\nu)\{(n+\beta^2)\hat{p}(B_\nu) - \beta^2\}}}\right) \cdot \hat{p}(c'|B_\nu) \end{aligned} \tag{66}$$

となり，これは式 (37) の左辺に一致する．

(平成 9 年 4 月 11 日受付)
(平成 10 年 6 月 5 日採録)

鈴木英之進（正会員）

昭和 40 年生．昭和 63 年東京大学工学部卒業．平成 5 年同大学大学院工学系研究科博士課程修了．博士（工学）．同年東京工業大学工学部情報工学科助手．平成 8 年横浜国立大学工学部電子情報工学科講師を経て，平成 9 年より同助教授．平成 9 年度人工知能学会論文賞，平成 9 年度人工知能学会研究奨励賞，第 11 回人工知能学会全国大会優秀論文賞受賞．データマイニング，機械学習など人工知能に関する研究に従事．人工知能学会，AAAI，IEEE-CS 各会員．