不動点探索形構造受精変換多段階認識の、確率過程論的取り扱い

鈴木昇一、佐久間拓也、釈氏孝浩,
前田英明、下平丕作士

A Stochastic-Process Treatment of Multi-Stage Recognition concerning a Structural-Fertilization Transformation of a Fixed-Point Searching Type

Shoichi Suzuki, Takuya Sakuma, Takahiro Shakushi,
Hideaki Maeda and Hisashi Shimodaira

あらまし

パターン認識過程は、入力パターンの帰属するであろうカテゴリ候補の集合を単一の元から成るカテゴリ候補の集合に変換するもの（カテゴリ候補を絞っていき、該当しないカテゴリ候補（非カテゴリ候補）を除去して行く過程）であると考えてみよう。

非カテゴリ候補の除去過程にソフトウェア信頼性の理論を適用するため、非カテゴリ候補の個数を数えていく計数過程 $N(t)$（時刻 t 迄に発見された非カテゴリ候補の個数）が非定常ポアソン確率過程とみなされる。$N(t)$ の期待値である平均値関数（強度関数）$M(t)$ の満たす微分方程式を仮定し、時刻 t における残存非カテゴリ候補1個当りの非カテゴリ候補発見率などにつき、研究する。

問題とする処理の対象パターン φ が最終的に認識されるためには、1つのカテゴリ候補を除き、残りのすべてのカテゴリ候補が非カテゴリ候補として除去されなければならないが、S.Suzukiの提案したカテゴリ帰属知識に関する不動点方程式が成立し、認識の働きが終了するという過程に関し、その途中の認識過程に残存するカテゴリ候補の個数を推定するための分析などがなされている。特に、単位時間当りに発見される非カテゴリ候補の平均個数はその時刻の残存非カテゴリ候補の数に比例するとしない遅れS字型ソフトウェア信頼度成長モデルに対応するパターン認識過程が、非カテゴリ候補の存在を観測･確認する過程と、構造受精変換を行って非カテゴリ候補の抽出に至る過程との2つの過程からなるとした認識過程が研究される。

キーワード

知能情報メディア　指数型分布　遅れS字型ソフトウェア信頼度成長モデル
信頼度　構造受精変換　不動点方程式

Abstract

Let us assume that a process of recognizing a pattern may be that of narrowing down a set of categories to which a pattern may possibly belong. A counting process N(t) denotes the number of categories (non-category candidates) which a recognizer detected till time t and to which a pattern may possibly belong.

Let us consider that N(t) is a nonhomogenuous stochastic Poisson process according to the theory of software reliability growth model. Assuming a differential equation of mean value function which is the expectation of N(t), we shall investigate a rate of non-category candidates detected per residual non-category candidates.

Two conditions for the recognizer to be able to recognize the pattern in question is that in a final stage of a multi-step recognition only category candidate remains eventually and all the other category candidates must be removed.When a fixed-point equation about categorical membership knowledges suggested by S.Suzuki holds,the muti-step recognition of the pattern comes to an end.An analysis for a fixed-point induction of such a recognition is made to estimate a number of category candidates which remains at each stage.The corresponding recognition process of the delayed S-shaped SRGM (Software Reliability Growth Model) DSSRGM ,where in this DSSRGM the expected number of non-category-candidates in existence detected per unit time (the instantaneous detection rate) is not proportional to the number of the current residual non-category candidates is specially in detail studied.This recognition process consists of two successive phases,namely a former phase observing an existence of non-catagory candidates, and a latter phase extracting non-catagory candidates by the application of selected structure-fertilization transformations to the current categorical membership knowledge.

Key words : intelligent information media exponential distribution delayed S-shaped reliability growth modeling reliability structural-fertilization transformation fixed-point equation

1．はじめに

1.1 知能現象としての、パターン認識過程

本論文では、入力パターンφを認識［6］,［9］,［41］する過程が、

φの帰属するカテゴリ候補を1つ2つと削除し、絞り込んでゆく計数過程N(t)≧0である

と見做される。このように、入力パターンφを認識する過程とは、カテゴリ帰属知識［55］で表されるカテゴリ候補の**カテゴリ候補の絞り込みの働き**［54］**が発現する知能情報メディアにおける知能現象**と見て、この働きを非定常ポアソン確率過程［23］,［63］として記述する。

1.2 ソフトウェア信頼度成長モデル

大規模ソフトウェアにおいては、その機能が失われてることにより被る被害が大きいし、設計段階でソフトウェアシステム状態の遷移やシステムのエラーを完全に把握することは難しい[66]。

ソフトウェア開発のテスト段階において、ソフトウェアに内在するエラーはソフトウェア故障として発見される。エラーは原因であり、故障はその結果である。

プログラムの部分にエラーがあると、その部分を実行経路に含むか含まないかによって、テストケースが失敗するか成功するかが決まる［27］。エラーは分離・修正され、テスト時間の経過と

ともに、エラーは減少.し、ソフトウェアの信頼度は成長する。このような過程を数式で表現したものを**ソフトウェア信頼度成長モデル**（software reliability growth model;SRGM）と呼ぶ［23］。

ソフトウェア内に潜在するエラー、フォールト（fault ; 欠陥）により、ソフトウェアが期待通りに動作しない現象に関する管理技術が多方面から、研究されている［14］，［16］，［20］～［22］，［25］～［35］，［37］。

適切にソフトウェア信頼度成長モデルを設定し、この成長モデルから、検出フォールトデータから残存しているフォールト数と、その検出に必要な時間を推定するのが、**ソフトウェア工学**における目的である。

1.3 ソフトウェア信頼度の手法を適用した“カテゴリ候補の並列的絞り込み方法”

処理の対象とする問題のパターンφについて、その帰属するカテゴリ候補を絞り込む方法としては、

①カテゴリ候補の逐次的絞り込み方法

…注目した1つのカテゴリが正しいカテゴリであるかどうかを、各認識段階で確かめ、正しいカテゴリでなかったら、このカテゴリを捨て、次の認識段階へ進む方式

②カテゴリ候補の並列的絞り込み方法

…可能性のあるカテゴリ候補の集合を想定し、各認識段階で複数のカテゴリ候補を捨て、残りのカテゴリ候補を次の認識段階で処理の対象とする方式

があるが、本研究では、

$$\text{発見された1つのソフトウェア故障 (software failures)} = \text{発見された1つの非カテゴリ候補} \tag{1.1}$$

と見て、ソフトウェア信頼度の手法を適用し、後者の②を論じている。

Let $\{N(t), t \geqq 0\}$ be a counting process that has independent increments so that the numbers of errors detected during disjoint time-intervals are independent ［14］.

入力パターンφに関するパターン認識過程を、φの帰属するカテゴリ候補を1つ2つと削除し、絞り込んでいく**計数過程**（counting process）

$$N(t) \in \{0,1,2,\cdots\}, t \geqq 0 \tag{1.2}$$

と考えるものである。ここに、$N(t)$は、入力パターンφに対し、時刻tでk_t個のカテゴリ候補が発見・残存されているとき、カテゴリ総数をm_0として、

$$N(t) = m_0 - k_t \tag{1.3}$$

と定義される。このように、

$$N(t) \equiv N(t;\varphi) \tag{1.4}$$

は、入力パターンφに関し、時刻tで発見された非カテゴリ候補の個数を表している。

1.4 本研究内容の新規性,有効性

パターン認識過程をカテゴリ候補の除去過程と想定し、然も、カテゴリ候補を除去し、唯一のカテゴリ候補を残すというパターン認識過程をソフトウェア信頼度成長過程と想定し、パターン認識の働きを研究した論文は他にない（研究内容の**新規性**）。

パターンφの整形化後、φから想起される唯1つのカテゴリに対応する“代表パターン”を決定する方、つまり、想起の方が、φの帰属する唯1つのカテゴリを推定する方、つまり、識別の方

より、難しい手続きといわれるが、このようなパターン認識の働きを確率過程論的に多段階的に分析出来たが、人間による認識の働きと結びつけられるとき、認知科学への貢献が期待されよう(研究内容の**有効性**)。

1.5 遅れS字型ソフトウェア信頼度成長モデルを利用した残存するカテゴリ候補の個数推定

パターン認識過程に残存するカテゴリ候補の個数を推定するためのモデルが研究される。そのために、ソフトウェアに残存するエラーの個数を推定するときに、経験的に中小規模のソフトウェアの信頼性モデルによく適合すると言われている

「**遅れS字型ソフトウェア信頼度成長モデル**(delayed S-shaped reliability growth modeling for software error detection)[16]」

を適用する。

フォールト除去過程を、ソフトウェア故障の発生現象を観測する過程と、その原因となったフォールトを発見する過程とに区別して、説明したモデルであるが、大部分のソフトウェア信頼度成長モデルはそれら2つの過程を同一視している [21]。

具体的には、遅れS字型ソフトウェア信頼度成長モデル(delayed S-shaped SRGM)は次のように説明される [20]:

(イ)「障害発生数は残存フォールト数に比例するが、障害検出からフォールト認定迄時間を要する」と解釈し、時刻t迄に検出された期待フォールト数H(t)を、

$$H(t)=a\ [1-(1+bt)\exp(-bt)],\ a, b>0 \tag{1.5}$$

で表したモデルである。ここで、a, bは、

a:最終的に検出される期待フォールト数(the expected number of faults to be eventually detected) (1.6)

b:フォールト出現率(the fault-occurrence rate) (1.7)

であり、H(t)はS字型成長曲線を形成する。 □

その他の、ソフトウェア信頼度成長モデルとして、3種類(ロ),(ハ),(ニ)が付録Cで説明されている。

1.6 研究内容の信頼性

さて、任意のパターン認識手法より、認識率が下回らない認識手法が不動点探索形(構造受精多段階帰納推論)認識手法として、存在することがで証明されている(文献 [55] の定理6. 1(**万能認識定理**))。認識システム内の3要素 T, SM, BSCが十分、問題としている処理対象パターン集合 Φに関し適切に選ばれていなければ、入力パターンφのカテゴリ帰属知識のあいまい性を正しく解消できないことには、注意しておかなければならない。任意の通常の認識法を1段の認識過程で模擬できるが、このときの通常の認識法での正認識率を高めるには、多段決定過程を導入することがその1つの解決法であることが、**万能認識定理**の証明などから、容易に理解できよう。

Assuming that the expected number of errors detected per unit testing time (the instantaneous error-detection rate) is proportional to the current residual error content, many SRGM,s are formulated as

$$dH(t)/dt=b(t)\cdot[a-H(t)]\ (b(t)>0, t\geqq 0) \tag{1.8}$$

where b(t) is the error-detection rate per error at testing time t. Solving the differential equation (1.8) in terms of $H(t)$ under the condituion $H(0)=0$ yields

$$H(t)=a\cdot[1-\exp(-\int_0^t dx\, b(x))] \quad [14]. \tag{1.9}$$

□

計算機ソフトウェアの信頼性を定量的に評価する手法を適用し、ソフトウェアに残存するエラーの個数を推定する時、経験的に中小規模のソフトウェアの信頼性のモデルによく適合するといわれている山田･尾関によって提案された日本独自の［23］“遅れS字型ソフトウェア信頼度成長モデル”に対応した“不動点探索形構造受精多段階帰納推論を行うパターン認識の分析”を行っているので、その結果は信頼出来るものであろう（研究内容の**信頼性**)。

式（1.4）の$N(t)$の期待値である平均値関数(mean value function) $H(t)$の満たす微分方程式（1.8）を仮定し、時刻 t における残存非カテゴリ候補1個当りの非カテゴリ候補発見率$b(t)$ (the error-detection rate per error at testing time t）などにつき、研究する。

1.7 遅れS字型SRGMに対応するパターン認識の働きにおける2つの素過程

問題とする処理の対象パターンφが最終的に認識されるためには、1つのカテゴリ候補を除き、残りのすべてのカテゴリ候補が非カテゴリ候補として除去されなければならない。S.Suzukiの提案したカテゴリ帰属知識に関する不動点方程式（fixed-point equation）が成立し、認識の働きが終了するという過程［55］に関し、その途中の認識過程に残存するカテゴリ候補の個数を推定するための分析などが成されている。特に、単位時間当りに発見される非カテゴリ候補の個数はその時刻の残存非カテゴリ候補の数に比例するという“指数型SRGM”でなくて、**遅れS字型SRGMに対応する“不動点探索形構造受精多段階帰納推論を行うパターン認識過程”**、即ち、

その認識の働きが、非カテゴリ候補の存在を観測・ 確認する素過程と、構造受精変換を行って非カテゴリ候補の抽出に至る素過程との2つの過程からなるとした認識過程

が研究される。

1.8 パターン認識における自由パラメータの推定手法

また、n個のデータ $\{\langle \underline{x_i}, y_i\rangle \mid i=1\sim n\}$ にパラメータθを持つ関数 $s(x;\theta)$ を当てはめる場合、モデル

$$\text{Model}: y_i=s(\underline{x_i};\theta)+\varepsilon \tag{1,10}$$

におけるεが$0, \sigma^2$ を各々平均値、分散とする正規分布［15]，［56］$N(0,\sigma^2)$に従う独立な確率変数と想定すると、真の分布に近いモデルを得るためには、対数尤度

$$\ell\,(y_1, y_2,\cdots, y_n/\theta, \sigma\, 2)$$
$$\equiv \log_e \prod_{i=1}^{n}\cdot(1/\sqrt{2\pi\sigma^2})\cdot\exp[-\{y_i-s(\underline{x_i};\theta)\}^2/(2\sigma^2)] \tag{1.11}$$

を最大にすればよくて、残差分散の最尤推定量

$$\sigma^{2\wedge}\equiv(1/n)\cdot\sum_{i=1}^{n}[y_i-s(\underline{x_i};\theta)]^2 \tag{1.12}$$

を用意すると、最大対数尤度$\ell^\wedge$は、

$$\ell^\wedge\,(y_1, y_2,\cdots, y_n/\theta,\ \sigma^2)$$
$$=-(n/2)\cdot\log_e(2\pi)-(n/2)\cdot\log e\,\sigma^{2\wedge}-n/2 \tag{1.13}$$

と計算されるから、結局、残差分散の最尤推定量［17]，［56］ $\sigma^{2\wedge}$を最少とする $\theta=\theta^\wedge$が最良モデルであることが判明する。つまり、最小二乗法を使ってのモデル決定法に一致する［64]。

このように、自由パラメータの数が同一の、確率分布で表される数理モデルについては、最小

二乗法で自由パラメータの値を決めればよいが、自由パラメータの数が異なるモデルの集合から、最良モデルを決定するには、K-L情報量(amount of Kullback-Leibler information)[32], [57] の観点からは、赤池情報量基準(Akaike Information Criterion) [19]

$$\mathrm{AIC} \equiv (-2) \times [(\text{モデルの最大対数尤度}) - (\text{モデル内の自由パラータの数})] \tag{1.14}$$

を最少とするモデルを選べばよい(付録D)。

ソフトウェア信頼度成長モデル内の自由パラメータの推定 (parameter estimation) には、様々な具体的手法が研究されているが [14], [21], [23], [26], [28]、本研究でも、同様な手法を研究する周知の基礎を付録Dで解説しておいた。

2. カテゴリ候補を絞って行く過程としてのパターン認識過程

本章では、ソフトウェアエラーと非カテゴリ候補との違い、つまり、

「ソフトウェア修理では、ソフトウェアエラーは終局段階では完全に除去されなければならないが、パターン認識では、終局段階で少なくとも1つのカテゴリ候補が残らなければならない」

という制約

を意識しながら、カテゴリ帰属知識空間 [54], [55] と称される$\langle \Phi, 2^J \rangle$での不動点探索形認識手法

「The pattern φ can be recognized as the corresponding model $T\omega_j$ if there is some t such that a fixed-point equation

$$TA(\mu_t)\, T\langle \psi_t, \lambda_t \rangle =_{\triangle} \langle \psi_t, \lambda_t \rangle \quad \wedge \lambda_t = [j] \quad \text{holds」} \tag{2.1}$$

が説明される。ここに、$[j_1, j_2, \cdots, j_k]$ は k 個の要素 $j_1, j_2, \cdots, j_k$ からなるリストを表し、集合 $\{j_1, j_2, \cdots, j_k\}$ と同一視することがある。特に、要素を1つも持たないリストを空集合 ϕ で表す。

2.1 A category-candidates elimination modeling in stochastic pattern-recognition process

We deals with the problems of removing redundant information from a given knowledges $\langle \varphi, \gamma \rangle \in \langle \Phi, 2^J \rangle$.

「パターン $\varphi \in \Phi$ が、カテゴリ集合

$$\underline{\mathfrak{C}} = \{\mathfrak{C}_j \mid j \in J\} \tag{2.2}$$

の何れか1つのカテゴリ $\mathfrak{C}_j$ に帰属する可能性がある」

という **"カテゴリ帰属知識"** を、認識システムRECOGNITRONが持っているとする。この知識を、

$$\langle \varphi, J \rangle \in \langle \Phi, 2^J \rangle \tag{2.3}$$

と表す。一般に、式(2.3)において、J の代りに、$\gamma \in 2^J$ と置き換えたものを、

$$\langle \varphi, \gamma \rangle \in \langle \Phi, 2^J \rangle \tag{2.4}$$

と表記し、

「パターン $\varphi \in \Phi$ が、カテゴリ集合 $\underline{\mathfrak{C}}$ の部分集合

$$\{\mathfrak{C}_j \mid j \in \gamma\} \subseteq \underline{\mathfrak{C}} \tag{2.5}$$

の何れか1つのカテゴリ $\mathfrak{C}_j$ に帰属する可能性がある」

という "認識システム RECOGNITRON がパターン $\varphi \in \Phi$ に対し持っているカテゴリ帰属知識" を意味している。ここに、2^J は、カテゴリ番号 $j \in J$ から成るカテゴリ番号集合 J のすべての部分集合の成す集合、ベキ集合 (power set) である。また、$\langle \Phi, 2^J \rangle$ は、**カテゴリ帰属知識空間**

(categorical membership-knowledge space) と呼ばれ、すべてのパターン $\varphi \in \Phi$ と、すべてのカテゴリ番号集合 $\gamma \in 2^J$ との成す対であり、**対リストの集合**と呼ばれることがある。

The condensing algorithm (a category-candidates elimination modeling in stochastic pattern-recognition process) uses as a subroutine a function procedure SUBS which after receiving $\langle\varphi, \gamma\rangle$ returns a fertilization-transformation $TA(\mu)T$ such that selecting some $\mu \in 2^J$ specially

$TA(\mu)T\langle\varphi, \gamma\rangle =_\triangle \langle\psi, \lambda\rangle$ holds. □

一般に、2式 (2,3), (2.4) の $\langle\varphi, J\rangle$, $\langle\varphi, \gamma\rangle$ は a non-condensed knowledge であり、このカテゴリ帰属知識などを、ある1つのカテゴリ知識

$$\langle\omega_j, [j]\rangle \tag{2.6}$$

と絞ってゆく過程（**凝縮過程**; condensing process)、或いは、**精密化過程**（refining process) がパターン認識過程である。ここに、ω_j は第 $j \in J$ 番目のカテゴリ $\mathfrak{C}_j$ の諸性質を典型的に代表しているパターン（代表パターン）である。問題とする入力パターン φ に対し、式 (2.6) が得られたとき、

処理対象とした入力パターン φ から、$T\omega_j$ が連想（想起）されたといい、

$$\text{この}\varphi\text{は、第}\, j \in J\,\text{番目のカテゴリ}\, \mathfrak{C}_j\,\text{に認識された} \tag{2.7}$$

という。

$\langle\varphi, J\rangle$, 或いは, $\langle\varphi, \gamma\rangle$ が式 (2.6) の $\langle\omega_j, [j]\rangle$ に凝縮される過程は、一般に、

$$\langle\psi_0, \lambda_0\rangle \in \langle\Phi, 2^J\rangle, \text{ where } \psi_0 \equiv T\varphi,\ \lambda_0 \equiv \gamma \tag{2.8}$$

$$\to TA(\mu_0)T\langle\psi_0, \lambda_0\rangle =_\triangle \langle\psi_1, \lambda_1\rangle \in \langle\Phi, 2^J\rangle \tag{2.9}$$

$$\to TA(\mu_1)T\langle\psi_1, \lambda_1\rangle =_\triangle \langle\psi_2, \lambda_2\rangle \in \langle\Phi, 2^J\rangle \tag{2.10}$$

$$\to TA(\mu_2)T\langle\psi_2, \lambda_2\rangle =_\triangle \langle\psi_3, \lambda_3\rangle \in \langle\Phi, 2^J\rangle \tag{2.11}$$

$$\to$$

$$\cdots$$

$$\to TA(\mu_{t-2})T\langle\psi_{t-2}, \lambda_{t-2}\rangle \tag{2.12}$$

$$=_\triangle \langle\psi_{t-1}, \lambda_{t-1}\rangle \in \langle\Phi, 2^J\rangle, \tag{2.13}$$

$$\to TA(\mu_{t-1})T\langle\psi_{t-1}, \lambda_{t-1}\rangle =_\triangle \langle\psi_t, \lambda_t\rangle \in \langle\Phi, 2^J\rangle \tag{2.14}$$

$$\to TA(\mu_t)T\langle\psi_t, \lambda_t\rangle (=_\triangle \langle\psi_{t+1}, \lambda_{t+1}\rangle)$$

$$= \triangle\langle\psi_t, \lambda_t\rangle \in \langle\Phi, 2^J\rangle$$

$$\text{(fixed-point equation)} \tag{2.15}$$

と表される。ここに、

$$\langle\psi_{s+1}, \lambda_{s+1}\rangle =_\triangle TA(\mu_s)T\langle\psi_s, \lambda_s\rangle,$$

$$\text{for any } s \in \{0, 1, 2, \cdots, t\} \tag{2.16}$$

where

$$\psi_{s+1} = TA(\mu_s \cap \lambda_s)T\psi_s \tag{2.17}$$

and

$$\lambda_{s+1} = CSF(\psi_s, \mu_s \cap \lambda_s) \tag{2.18}$$

である。

カテゴリ候補を絞っていく性質が、

$$J \supseteq \mu_0 \supseteq \mu_1 \supseteq \mu_2 \supseteq \cdots \supseteq \mu_{t-2} \supseteq \mu_{t-1} \supseteq \mu_t \tag{2.19}$$

$$\wedge$$

$$J \supseteq \mu_0 \cap \lambda_0 \supseteq \mu_1 \cap \lambda_1 \supseteq \mu_2 \cap \lambda_2 \supseteq \cdots$$

$\supseteq \mu_{t-2} \cap \lambda_{t-2} \supseteq \mu_{t-1} \cap \lambda_{t-1} \supseteq \mu_t \cap \lambda_t$ (2.20)

或いは、

$J \supseteq \lambda_0 \supseteq \lambda_1 \supseteq \lambda_2 \supseteq \cdots \supseteq \lambda_{t-2} \supseteq \lambda_{t-1} \supseteq \lambda_t$ (2.21)

$\wedge$

$J \supseteq \lambda_0 \cap \mu_0 \supseteq \lambda_1 \supseteq \lambda_1 \cap \mu_1 \supseteq \lambda_2 \supseteq \cdots$

$\supseteq \lambda_{t-2} \cap \mu_{t-2} \supseteq \lambda_{t-1} \supseteq \lambda_{t-1} \cap \mu_{t-1} \supseteq \lambda_t$

$\supseteq \lambda_t \cap \mu_t$ (2.22)

と成立していなければならない。

2.2 不動点探索形認識手法に登場する諸記号

前節で登場した諸記号、説明、意味との重複を恐れず、更に、不動点探索形認識手法に関連する事実が説明される。

本論文は、4つの記号 $\Phi, J, 2^J, \langle \varphi, \gamma \rangle$ を各々、

Φ：処理の対象とする問題のパターン φ の集合 (2.23)

J：各パターン φ の帰属する概念的カテゴリ集合

(a set of conceptual categories) の集合

$\underline{\mathfrak{C}} = \{\mathfrak{C}_j \mid j \in J\}$ (2.24)

での添字(カテゴリ番号) j の集合 (2.25)

2^J：カテゴリ番号の集合 J の、すべての部分集合のなす集合 (2.26)

$\langle \varphi, \gamma \rangle (\in \langle \Phi, 2^J \rangle)$：パターン φ $(\in \Phi)$ とその帰属するカテゴリ候補 $\mathfrak{C}_j$ のすべてから成る集合 $\{\mathfrak{C}_j \mid j \in \gamma\} \subseteq \underline{\mathfrak{C}}$ でのカテゴリ番号 j の集合（リスト）γ との対（認識システム RECOGNITRON がパターン φ に関し持つカテゴリ帰属知識) (2.27)

とし、処理の対象とする問題のパターン（入力パターン）φ に関する不動点探索形認識手法でのパターン認識過程

$\langle \psi_0, \lambda_0 \rangle \in \langle \Phi, 2^J \rangle$

$,\text{where } \psi_0 \equiv T\varphi, \lambda_0 \equiv J$ (2.28)

$\rightarrow TA(\mu_0)T\langle \psi_0, \lambda_0 \rangle =_{\triangle} \langle \psi_1, \lambda_1 \rangle \in \langle \Phi, 2^J \rangle$ (2.29)

$,\text{where } \mu_0 \in \langle \Phi, 2^J \rangle$ (2.30)

$\rightarrow TA(\mu_1)T\langle \psi_1, \lambda_1 \rangle =_{\triangle} \langle \psi_2, \lambda_2 \rangle \in \langle \Phi, 2^J \rangle$ (2.31)

$,\text{where } \mu_1 \in \langle \Phi, 2^J \rangle$ (2.32)

$\rightarrow TA(\mu_2)T\langle \psi_2, \lambda_2 \rangle =_{\triangle} \langle \psi_3, \lambda_3 \rangle \in \langle \Phi, 2^J \rangle$ (2.33)

$,\text{where } \mu_2 \in \langle \Phi, 2^J \rangle$ (2.34)

$\rightarrow$

$\cdots$

$\rightarrow TA(\mu_{t-2})T\langle \psi_{t-2}, \lambda_{t-2} \rangle$ (2.35)

$=_{\triangle} \langle \psi_{t-1}, \lambda_{t-1} \rangle \in \langle \Phi, 2^J \rangle$ (2.36)

$,\text{where } \mu_{t-2} \in \langle \Phi, 2^J \rangle$ (2.37)

$\rightarrow TA(\mu_{t-1})T\langle \psi_{t-1}, \lambda_{t-1} \rangle$

$=_{\triangle} \langle \psi_t, \lambda_t \rangle \in \langle \Phi, 2^J \rangle$ (2.38)

,where

$\mu_{t-1} \in \langle \Phi, 2^J \rangle$ (2.39)

$\rightarrow TA(\mu_t)T\langle \psi_t,\ \lambda_t \rangle$

$(=_{\triangle} \langle \psi_{t+1},\ \lambda_{t+1} \rangle) =_{\triangle} \langle \psi_t,\ \lambda_t \rangle \in \langle \Phi, 2^J \rangle$

(**fixed-point equation**) (2.40)

,where $\mu_t \in \langle \Phi, 2^J \rangle$ (2.41)

を、パターンφの帰属するカテゴリ候補に関する**絞り込み性質**

$J \supset \lambda_0 \supset \lambda_1 \supset \lambda_2 \supset \cdots \supset \lambda_{t-2} \supset \lambda_{t-1} \supset \lambda_t$ (2.42)

を備えた

非定常ポアッソン過程としてのパターン認識過程

(Pattern-recognizing process as a nonhomogeneous poisson process)

と見做したものであり、この認識手法で登場するとして、本章で登場する5つの記号$1^{\#}$～$5^{\#}$の各定義の意味、説明の詳細については、文献［55］にある：

$1^{\#}$（**モデル構成作用素**；model-construction operator）

$T : \Phi \rightarrow \Phi$ (2.43)

$2^{\#}$（**類似度関数**；similarity-measure function）

$SM : \Phi \times \Omega \rightarrow \{s \mid 0 \leqq s \leqq 1\}$ (2.44)

$3^{\#}$（**大分類関数**；binary-state classifier）

$BSC : \Phi \times J \rightarrow \{0, 1\}$ (2.45)

$4^{\#}$（**構造受精作用素**；structure-fertilization operator）

$A(\mu) : \Phi \rightarrow \Phi$ (2.46)

$5^{\#}$（**カテゴリ選択関数**；category-selection function）

$CSF : \Phi \times 2^J \rightarrow 2^J$ (2.47)

また、カテゴリ帰属知識$\langle \varphi, \gamma \rangle$から$\langle \psi, \lambda \rangle$への変換（**構造受精変換**；structure-fertilization transformation）

$TA(\mu)T : \langle \Phi, 2^J \rangle \rightarrow \langle \Phi, 2^J \rangle$ (2.48)

の定義は、

$6^{\#}\ TA(\mu)T\langle \varphi,\ \gamma \rangle =_{\triangle} \langle \psi,\ \lambda \rangle$ (2.49)

,where

$\psi \equiv TA(\mu \cap \gamma)T\varphi$ (2.50)

$\lambda \equiv CSF(\varphi,\ \mu \cap \gamma)$ (2.51)

と与えられる。 □

上の$6^{\#}$に登場する“構造受精作用素”$A(\gamma)$と、$5^{\#}$でのカテゴリ選択関数CSFの両定義から得られる表現について、説明しておこう。

実変数uの、関数

$psn(u) \equiv 0 \text{ if } u < 0, \equiv 1 \text{ if } u \geqq 0$ (2.52)

と、不等式

$0 < b \leqq 1$ (2.53)

を満たす閾値bを用意する。その定義から、

①$\varphi = 0 \vee \gamma = \phi$の場合、

$A(\gamma)\varphi = 0$ (2.54)

②$\varphi \neq 0 \wedge \gamma \neq \phi$の場合、

$$A(\gamma)\varphi =$$
$$\Sigma_{k\in\gamma}[1+BSC(\varphi, k)$$
$$-psn(\Sigma_{k\in\gamma} BSC(\varphi, k)-b)]\cdot SM(\varphi, \omega_k)\cdot T\omega_k. \qquad (2.55)$$

であることが導かれている（文献［55］の、式（6.12）、並びに、付録Eの定理E2）。

また、式（2.14）のカテゴリ選択関数CSFは、その定義から、

①$\varphi = 0 \vee \gamma = \phi$の場合、

$$CSF(\varphi, \gamma) = \phi \qquad (2.56)$$

②$\varphi \neq 0 \wedge \gamma \neq \phi$の場合、

$$CSF(\varphi, \gamma) =$$
$$\{k\in\gamma \mid [1+BSC(\varphi, k)$$
$$-psn(\Sigma_{k\in\gamma} BSC(\varphi, k)-b)]\cdot SM(\varphi, \omega_k) > 0\}. \qquad (2.57)$$

と表現されることが導かれている(文献[55]の、付録Eのaxiom 4の（i）、並びに、付録Eの定理E2)。

尚、2式（2.49），（2.16）で登場しているカテゴリ帰属知識間の、2元関係（a binary relation on $\langle \Phi, 2^J \rangle$）としての、等形式関係（equi-form relation）$=_\triangle$は、次の定義2.1で与えられる。

［定義2.1］**（カテゴリ帰属知識間の等形式関係）**

$\langle \varphi, \gamma \rangle =_\triangle \langle \psi, \lambda \rangle$（恒等的に等しい）

$\Leftrightarrow \varphi = \psi \wedge \gamma = \lambda$. □

上の定義2.1の等形式関係 $=\triangle$ より弱い等構造関係（a equi-structure relation）$=$ を、次の定義2.2のごとく定義する。形式が同じであれば、構造も同じであるが、構造が同じだからいって、形式が同じとは限らないカテゴリ帰属知識 $\langle \varphi, \gamma \rangle$ が存在する事実に注意しておこう。

［定義2.2］**（カテゴリ帰属知識間の等構造関係）**

$\langle \varphi, \gamma \rangle = \langle \psi, \lambda \rangle$

$\Leftrightarrow CSF(\varphi, \gamma) = CSF(\psi, \lambda) \wedge$

$[\forall\ j \in CSF(\varphi, \gamma) \cap CSF(\psi, \lambda),$

$SM(\varphi, \omega_j) = SM(\psi, \omega_j)]$. □

このとき、2カテゴリ帰属知識$\langle \varphi, \gamma \rangle, \langle \psi, \lambda \rangle$間の内積 $(\langle \varphi, \gamma \rangle, \langle \psi, \lambda \rangle)$を、

$$(\langle \varphi, \gamma \rangle, \langle \psi, \lambda \rangle)$$
$$\equiv \Sigma_{j\in CSF(\varphi,\gamma)\cap CSF(\psi,\lambda)} \sqrt{SM(\varphi, \omega_j)}\cdot\sqrt{SM(\psi, \omega_j)} \qquad (2.58)$$

と定義し、そのノルム $\|\langle \varphi, \gamma \rangle\|$ を

$$\|\langle \varphi, \gamma \rangle\| \equiv \sqrt{(\langle \varphi, \gamma \rangle, \langle \varphi, \gamma \rangle)} \qquad (2.59)$$

と定義すれば、カテゴリ帰属知識間の等構造関係 $=$ について、**カテゴリ帰属知識** $\langle \varphi, \gamma \rangle$ の所要の直交直和分解を与える次の定理2.1が成り立つ。パターン$\varphi \in \Phi$のカテゴリ依存構造を明らかにするこの分解は$\langle \varphi, \gamma \rangle \in \langle \Phi, 2^J \rangle$の標準分解（canonical decomposition）とも呼ばれる。ここに、$s \overline{\in} S$は、要素sが集合Sの元でないの意である。

［定理2.1］**（カテゴリ帰属知識の直交直和分解（標準分解）定理）**［55］

3写像SM,BSC,γについて、全射性が成立するという仮定［55］の下で、

$\forall \langle \varphi, \gamma \rangle \in \langle \Phi, 2^J \rangle$,

（i）(直交展開式；SS展開)

$\langle \varphi, \gamma \rangle$

$= \Sigma_{j \in J} (\langle \varphi, \gamma \rangle, \langle \omega_j, [\,j\,] \rangle) \cdot \langle \omega_j, [\,j\,] \rangle$ (2.60)

$= \Sigma_{j \in CSF(\varphi, \gamma)} \sqrt{SM(\varphi, \omega_j)} \cdot \langle \omega_j, [\,j\,] \rangle$ (2.61)

(ii) (ノルム $\| \langle \varphi, \gamma \rangle \|$の表現)

$\| \langle \varphi, \gamma \rangle \|$

$= [\Sigma_{j \in J} | (\langle \varphi, \gamma \rangle, \langle \omega_j, [\,j\,] \rangle) |^2]^{1\ \ 2}$ (2.62)

$= [\Sigma_{j \in CSF(\varphi, \gamma)} \ SM(\varphi, \omega_j)]^{1\ \ 2}$ (2.63)

が成り立つ。

ここに、次の(iii), (iv)が成立している：

(iii) ($\langle \Phi, 2^J \rangle$の基底$\langle \Omega, J \rangle$の存在)

$(\langle \omega_i, [\,i\,] \rangle, \langle \omega_j, [\,j\,] \rangle) =$

$$\begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \qquad (2.64)$$

が成り立ち、

$\langle \Omega, J \rangle \equiv \{ \langle \omega_j, [\,j\,] \rangle \mid j \in J \}$ (2.65)

は、カテゴリ帰属知識空間$\langle \Phi, 2^J \rangle$の完全正規直交系 (complete orthonormal system) である。

(iv) (直交展開係数$(\langle \varphi, \gamma \rangle, \langle \omega_j, [\,j\,] \rangle)$の決定)

$(\langle \varphi, \gamma \rangle, \langle \omega_j, [\,j\,] \rangle) =$

$$\begin{cases} \sqrt{SM(\varphi, \omega_j)} & \text{if } j \in CSF(\varphi, \gamma) \\ 0 & \text{if } j \in CSF(\varphi, \gamma) \end{cases} \qquad (2.66)$$

□

2.3 不動点探索形認識手法の、荒っぽい説明

2.3.1 パターン集合 Φ の決定

先ず、式 (2.20) の写像 (モデル形式) $T: \Phi \rightarrow \Phi$ について、3性質

① $0 \in \Phi$

② $\forall \varphi \in \Phi$, $a \cdot \varphi \in \Phi$ for any positive real number a

③ $\forall \varphi \in \Phi, T\varphi \in \Phi$

を満たすようなパターン集合 Φ を想定しておく。このようなパターン集合 Φ は、初期条件

$\Phi(t) |_{t=0} = \Phi_B \ni 0$ (2.67)

の下で、パターン集合 Φ に関する逐次的再帰領域方程式 (reflective domain equation)

$\Phi(t+1) = \Phi_B \cup T \cdot \Phi(t) \cup R^{++} \cdot \Phi(t)$,

where R^{++} is a set of positive real numbers (2.68)

を解いて、

$\Phi = \lim_{t \rightarrow \infty} \Phi(t)$ (2.69)

$= R^{++} \cdot \Phi_B \cup R^{++} \cdot T \cdot \Phi_B$ (2.70)

と与えられる (文献 [55] の定理2.1)。

2.3.2 最大類似度法に関する精密化としてのパターン認識手法

入力パターン$\varphi \in \Phi$の帰属すべきカテゴリ (類概念) を決定するという**認識の働き**としては、パターンφの代りとしてのパターンモデル$T\varphi \in \Phi$を使い、次の不動点探索形構造受精認識法が考えられる．

次項で説明される本認識法（**不動点探索形構造受精多段階変換に基づくパターン認識手法**）は、入力パターンφと、第 $k \in J$ 番目のカテゴリ$\mathfrak{C}_k$の代表パターンω_kとの類似度 $SM(\varphi, \omega_k)$ が最大となる最も若いカテゴリ番号

$$\mathrm{argmax}_{k \in J}\ SM(\varphi, \omega_k) = j \in J \tag{2.71}$$

を見つけ、

$$\varphi \text{ belongs to the j-th category } \mathfrak{C}_j \tag{2.72}$$

と、認識する

と説明される最大類似度法［50］に関する精密化であり、14式（2.28）～（2.41）で説明される認識法がカテゴリ帰属知識を用いないで説明されたものである。

2.3.3 不動点探索形（構造受精）認識手法

各カテゴリ番号 $j \in J$ の集合 J のすべての部分集合のなす集合を 2^J と表そう。$\mu \in 2^J$ を、パターン$\varphi \in \Phi$の帰属する可能性のあるカテゴリ番号のリストとして採用しながら、このμを助変数に持つパターン変換作用素（構造受精作用素）

$$A(\mu): \Phi \to \Phi \tag{2.73}$$

を用意し、パターンモデル$T\varphi$を、

$$TA(\mu)T\varphi = \eta \in \Phi \tag{2.74}$$

というように、パターンモデル$T\eta$へと変換することを考えよう。写像

$$TA(\mu)T: \Phi \to \Phi \tag{2.75}$$

は、**構造受精変換**（structure-fertilization transformation）と呼ばれているものである。このとき、写像 T のベキ等性（文献［55］の付録Aでのaxiom 1, (iii)）

$$\forall \varphi \in \Phi, T(T\varphi) = T\varphi \tag{2.76}$$

より、式（2.33）のパターンηの不動点性

$$T\eta = \eta \tag{2.77}$$

が成立しており、この構造受精変換段階で得られた式（2.77）のパターンηは写像 T の不動点となっている。

このような$\mu \in 2^J$を、多段階認識過程における各多段階でその都度、適切に選び、式（2.75）の構造受精変換を多段階的に何回か繰り返して行き、最終的にパターンモデル$T\varphi$の帰属する可能性のあるカテゴリの番号が唯一に、式（2.40）のλ_tが、例えば、

$$\lambda_t = [j] \tag{2.78}$$

となり、第 $j \in J$ 番目のカテゴリ$\mathfrak{C}_j$に絞られることによって、認識システム RECOGNITRON は、入力パターン$\varphi \in \Phi$を、

$$\text{An input pattern } \varphi \text{ can be assigned to a category } \mathfrak{C}_j\text{, and can be reproduced as } T\omega_j \tag{2.79}$$

のごとく、$\mathfrak{C}_j$に認識し（カテゴリ番号の付値を行い、識別し）、$T\omega_j$として再現する（$T\omega_j$を想起する）。

2.4 fixed-point searching recognition by selecting some category-candidates for an unknown pattern φ

2.4.1 Recognition and potential energy of categorical membership-knowledge

The recognition of a pattern φ stands for the decision about a categorical membership on the basis of the corresponding model $T\varphi$ of the observed pattern φ. A convergence to the subset of set

$\{\langle\psi, \lambda\rangle \in \langle\Phi, 2^J\rangle \mid E(\psi, \lambda) = 0\}$ (2.80)

of critical points depending on the pattern φ may be regarded as a recognition process of φ, where $E(\psi, \lambda)$ is called a corresponding potential energy of the knowledge $\langle\psi, \lambda\rangle$, and is defined as follows:

①$\varphi = 0 \vee \gamma = \phi$ の場合、

$E(\psi, \lambda) \equiv 0$ (2.81)

②$\varphi \neq 0 \wedge \gamma \neq \phi$ の場合、

$E(\psi, \lambda) \equiv |\lambda| - \Sigma_{j\in\lambda}\, SM(\psi, \omega_j)$ (2.82)

where $|\lambda|$ is the cardinality of set λ (i.e., the total number of elements in the set λ) [55].

2.4.2 An implementation method for category-candidate selection

(i) A deformed knowledge $\langle\varphi, \gamma\rangle$ of $\langle\psi, \lambda\rangle$

A knowledge $\langle\varphi, \gamma\rangle$ is called a variant of a knowledge $\langle\psi, \lambda\rangle$ if and only if there is a fertilization-transform $TA(\mu)\ T(\mu \in 2^J)$ such that

$TA(\mu)\ T\langle\varphi, \gamma\rangle =_\triangle \langle\psi, \lambda\rangle$, (2.83)

where

$\psi = TA(\mu \cap \gamma)\, T\varphi$ (2.84)

and

$\lambda = CSF(\varphi, \mu \cap \gamma)$. (2.85)

(ii) A knowledge $\langle\varphi, \gamma\rangle \in \langle\Phi, 2^J\rangle$ subsumes a knowledge $\langle\psi, \lambda\rangle$, denoted by

$\langle\varphi, \gamma\rangle \bigcirc\!\!\rightarrow \langle\psi, \lambda\rangle$ (2.86)

, if and only if there is a fertilization-transformation $TA(\mu)T(\mu \in 2^J)$ such that

$TA(\mu)T\langle\varphi, \gamma\rangle =_\triangle \langle\psi, \lambda\rangle$. (2.87)

We say that a knowledge

$\langle\psi, \lambda\rangle$ implies $\langle\varphi, \gamma\rangle$ and write

$\langle\psi, \lambda\rangle \Rightarrow \langle\varphi, \gamma\rangle$ (2.88)

if and only if $\langle\psi, \lambda\rangle$ logically implies $\langle\varphi, \gamma\rangle$. Note that from equation (2.83) it follows that equation (2.88); but the converse is not true in general.

2.4.3 fixed-point equation and recognition

At the t-th stage in such a way that a fixed-point equation

$\langle\psi_t, \lambda_t\rangle = TA(\mu_t)T\langle\psi_t, \lambda_t\rangle$ for a pattern φ to be recognized, starting from the initial state $\langle\psi_0, \lambda_0\rangle \in \langle\Phi, 2^J\rangle$, where $\psi_0 \equiv T\varphi$ and $\lambda_0 \equiv J$

holds, the potential energy $E(\psi_t, \lambda_t)$ become a minimum. Then the recognizer RECOGNITRON shall determine that φ belongs to one of the narrowed category set $\{\mathfrak{C}_j \mid j \in \lambda_t\}$ as if to tell us that the pattern φ was the fixed-point model φ_t.

2.5 カテゴリ分布の変換と認識結果の分類

2.5.1 カテゴリ分布の変換

想定している式（2.2）の全カテゴリ集合 $\underline{\mathfrak{C}}$ に注目すると、任意のパターン $\varphi \in \Phi$ は事前に、共通して、カテゴリ分布

$$p(\mathfrak{C}_j), j \in J \tag{2.89}$$
を持っている。ここに、$p(\mathfrak{C}_j)$ は、確率条件

$$[\forall j \in J, 0 < p(\mathfrak{C}_j) < 1] \wedge \Sigma_{j \in J}\, p(\mathfrak{C}_j) = 1 \tag{2.90}$$
を満たしており、第 $j \in J$ 番目のカテゴリ $\mathfrak{C}_j$ の生起確率である。

処理の対象とする問題のパターン φ の集合 Φ に対し適切に選んだモデル構成作用素 $T:\Phi \rightarrow \Phi$ を固定し、式（2.44）の類似度関数 SM を選ぶと、式（2.90）のカテゴリ分布より精密な φ の事前確率分布

$$SM(\varphi, \omega_j), j \in J \tag{2.91}$$
が得られる。ここに、$\omega_j \in \Phi$ は第 $j \in J$ 番目のカテゴリ $\mathfrak{C}_j$ の持つ諸性質を典型的に所有している代表パターンであり、その適応的決定法は文献 [55] の付録 I で説明されている。代表パターン集合

$$\Omega \equiv \{\omega_j \mid j \in J\} \tag{2.92}$$
が導入されたことに注意しておく。確率条件

$$[\forall j \in J, 0 \leqq SM(\varphi, \omega_j) \leqq 1 \wedge \sum_{j \in J} SM(\varphi, \omega_j) = 1 \tag{2.93}$$
が成立しており、

SM (φ, ω_j) は、パターン φ が第 $j \in J$ 番目のカテゴリ $\mathfrak{C}_j$ に帰属している確率（φ が ω_j に似ている確率）である (2.94)

と解釈される。

処理の対象とする問題のパターン（入力パターン）φ について、式（2.91）の事前確率分布を以下の式（2.97）の事後確率分布に変換するのが、不動点探索形構造受精多段階変換によるパターン認識の働きを備えている認識システム RECOGNITRON の情報処理機能である。

不動点探索形構造受精多段階変換によるパターン認識の働きを備えている認識システム RECOGNITRON は、モデル構成作用素 T, 類似度関数 SM, 大分類関数 BSC などで代表されるその持っている知識の範囲で、処理の対象とする問題のパターン（入力パターン）φ について、各カテゴリ番号リスト $\mu \subseteq J$ を各認識段階でその都度適切に選んで得られる各構造受精変換 $TA(\mu)T$ による変換を、認識の初期段階のカテゴリ帰属知識 $\langle T\varphi, J\rangle$ に対し、不動点方程式

$$TA(\mu)T\langle\psi, \lambda\rangle =_{\triangle} \langle\psi, \lambda\rangle \tag{2.95}$$
が成立するまで繰り返し、φ に対応する終局的なカテゴリ極限分布

$$S_j(\psi; \underline{\mathfrak{C}}), \ j \in J \tag{2.96}$$
を、推定しようとする。各 $S_j(\psi; \underline{\mathfrak{C}})$ は、

$$[\forall j \in J, 0 \leqq S_j(\psi; \underline{\mathfrak{C}}) \leqq 1 \wedge \sum_{j \in J} S_j(\psi; \underline{\mathfrak{C}}) = 1 \tag{2.97}$$
を満たし、不動点パターン ψ が第 $j \in J$ 番目のカテゴリ $\mathfrak{C}_j$ に帰属している程度（帰属度 ; grade of membership）である。実は、$S_j(\psi; \underline{\mathfrak{C}})$ は、

$$S_j(\psi; \underline{\mathfrak{C}}) = \begin{cases} SM(\psi, \omega_j) \ / \Sigma_{j \in \lambda}\, SM(\psi, \omega_j) & \cdots j \in J \text{のとき} \\ 0 & \cdots j \in J - \lambda \text{のとき} \end{cases} \tag{2.98}$$
と求められる。

2.5.2　認識結果の分類

不動点方程式（2.95）が成立しているから、

$$\mathrm{TA}(\mu\cap\lambda)\mathrm{T}\psi=\psi \tag{2.99}$$

$$\wedge\mathrm{CSF}(\psi,\mu\cap\lambda)=\lambda \tag{2.100}$$

が成立している。登場した式（2.100）の CSF（$\psi,\mu\cap\lambda$）$\subseteq$J はカテゴリ帰属知識$\langle\psi,\mu\cap\lambda\rangle$の有効なカテゴリ候補の番号リストである。

入力パターンφから想起されるパターンは、不動点方程式（2.95）を満たすという意味での不動点パターンψであり、2.5.1項で説明された不動点探索形構造受精多段階認識の、入力パターンφに関する結果は、

(i)（認識不定）　$\lambda=[\mathrm{j}_1\,\mathrm{j}_2\cdots\mathrm{j}_k]$ならば、

φ belongs to one of the category subset $\{\mathfrak{C}_j \mid j\in\lambda\}$

(ii)（認識確定）　$\lambda=[\,\mathrm{j}\,]$ならば、

φ belongs to the j-th category $\mathfrak{C}_j$　(2.101)

(iii)（認識不能）　$\lambda=\phi$ ならば、

φ belongs by no means to one of the category set $\{\mathfrak{C}_j \mid j\in\lambda\}$

と分類される（文献［55］，付録Gの定理G14（連想形認識不動点解の分類定理）を参照）が、(ii)の場合を除いて、(i)、(iii）の場合は、強制的に、カテゴリ番号

$$\mathrm{j}=\mathrm{argmax}_{k\in\lambda}\ s_k(\varphi;\underline{\mathfrak{C}})\in\lambda\in\mathrm{J} \tag{2.102}$$

を求め、式（2.101）のごとく、認識確定させることが出来よう。ここに、argmax$_{k\in\lambda}\cdots$ は、量 …の最大値を与える最も若いカテゴリ番号 $k\in\lambda$ の意である。

認識の3結果（i),（ii),（iii）において、入力パターンφに対応し想起される不動点パターンψと、φに対応する終局的な式（2.96）のカテゴリ極限分布とについては、文献［55］を参照されたい。

3．非定常ポアッソン過程としてのパターン認識過程

本章では、第1章での ②カテゴリ候補の並列的絞り込み方法 に限って、式（1.1）の対応を設定し、ソフトウェア信頼モデル（software reliability model）の理論を適用し、ソフトウェア故障として発見され、分離･修正されるエラーはテスト時間の経過と共に減少し、ソフトウェアの信頼度は成長するような過程を確率過程として表したソフトウェア信頼度成長モデルを、

“入力パターンφの不動点探索形構造受精多段階変換パターン認識過程”

＝“入力パターンφの帰属しない非カテゴリ候補を除去しながらの、カテゴリ候補の絞り込み過程”　(3.1)

の想定の下で、取り扱ってみよう。この取り扱いによって、第2章で説明された不動点探索形多段階構造受精多段階変換に基づくパターン認識の働きの、確率過程としての諸性質を明らかにするため、式（1.4）のN(t)の期待値である平均値関数H(t)の満たす微分方程式（1.8）を仮定し、時刻tにおける残存非カテゴリ候補1個当りの非カテゴリ候補発見率b(t)などにつき、研究する。

3.1　構造受精の多段階変換を用いたパターン認識

第2章の内容を要約すると、次のようになる。

初期条件

$$\langle \psi_t, \lambda_t \rangle \mid_{t=0} =_{\triangle} \text{式}(2.3)\text{の}\langle \varphi, J \rangle \tag{3.2}$$

或いは、更に一般化した初期条件

$$\langle \psi_t, \lambda_t \rangle \mid_{t=0} =_{\triangle} \text{式}(2.27)\text{の}\langle \varphi, \gamma \rangle \tag{3.3}$$

の下で、各カテゴリ番号リスト μ_{s-1} を適切に選定しながら、構造受精の多段階変換

$$TA(\mu_{s-1})T\langle \psi_{s-1}, \lambda_{s-1} \rangle =_{\triangle} \langle \psi_s, \lambda_s \rangle$$
$$s = 1, 2, \cdots, \tag{3.4}$$

を繰り返し、その結果、有限の認識段階、つまり、式 (3.4) の第 t $(\in \{0,1.2,\cdots\})$ 認識段階で、2式 (2.15), (2.40) の不動点方程式

$$TA(\mu_t)T\langle \psi_t, \lambda_t \rangle =_{\triangle} \langle \psi_t, \lambda_t \rangle \tag{3.5}$$

が終了条件 (termination criterion) として成立したとき、認識システム RECOGNITRON が入力パターン φ に対し、カテゴリ帰属知識 $\langle \psi_t, \lambda_t \rangle$ を最終的に持つことになり、

The recognizer RECOGNITRON shall determine that a pattern φ to be recognized belongs to one of the narrowed category set $\{\mathfrak{C}_j \mid j \in \lambda_t\}$ as if to tell us that the pattern φ in question was the fixed-point model ψ_t (3.6)

ということになる。

2式 (3.2) , (3.3) の何れかを初期条件として採用し、不動点方程式 (3.5) の成立を終了条件とした不動点探索過程で得られるカテゴリ帰属知識の列

$$\langle \psi_0, \lambda_0 \rangle, \langle \psi_1, \lambda_1 \rangle, \cdots, \langle \psi_s, \lambda_s \rangle, \cdots, \langle \psi_t, \lambda_t \rangle \tag{3.7}$$

は、陽に表されていないけれども、一般的には式 (2.75) で表される各構造受精変換

$$TA(\mu_{s-1})T : \langle \Phi, 2^J \rangle \rightarrow \langle \Phi, 2^J \rangle \tag{3.8}$$

の適用様相の意味する a more efficient condensing algorithm によって得られていると見做すことができ、

a sequence of categorical membership-knowledges obtained by replacing a noncondensed knowledge $\langle \psi_{s-1}, \lambda_{s-1} \rangle$ by a condensed $\langle \psi_s, \lambda_s \rangle$ which is subsumed by a transformation $TA(\mu_{s-1})T$ (3.9)

と考えられる。

3.2　ソフトウェア信頼度成長モデルになぞられた構造受精多段階変換パターン認識過程

ソフトウェア信頼度成長モデルでは、テスト工程や運用段階をソフトウェア信頼度過程とみなし、発生したソフトウェア故障の数、或いは、発見されたフォールト数に関する計数過程は非定常ポアッソン過程と見做されて良い [21]。

非定常ポアッソン過程としてのパターン認識過程

(Pattern-recognizing process as a nonhomogeneous Poisson process) (3.10)

を導入するには、このパターン認識過程を、

ソフトウェア故障の現象を観測する過程（ソフトウェア故障発見過程 ; failure detection process）を、非カテゴリ候補を除去する過程 (3.11)

と見立てるのがよい。

card(K) : the cardinality of set K(i.e., the number of elements in the set K) (3.12)

として、入力パターンφに関するパターン認識過程を、φの帰属するカテゴリ候補を1つ2つと削除し、絞り込んでいく式（1.2）の計数過程

$$\{N(t)\in\{0,1,2,\cdots\} \mid t\in\{0,1,2,\cdots\}\} \tag{3.13}$$

と考えるものである。ここに、N(t)は、入力パターンφに対し、時刻 t で card（γ_t）個のカテゴリ候補が発見されているとき、カテゴリ総数を m_0 として、

$$N\ (t;\varphi)\ \equiv N(t)=m_0-\mathrm{card}(\gamma_t)$$

：the number of non-category candidates detected up to time t (3.14)

と定義される。ここに、

m_0：カテゴリ総数 (3.15)

card(γ_t)：パターンφについて、時刻 t 迄に残存しているカテゴリ候補の数（the number of residual category-candidates detected up to time t） (3.16)

である。ここに、初期条件（3.2），(3.3）の下では、各々、

$$m_0=\mathrm{card}(J) \tag{3.17}$$

$$m_0=\mathrm{card}(\gamma) \tag{3.18}$$

である。

3.3 非カテゴリ候補の計数に関する非定常ポアッソン確率過程

処理の対象としている問題の認識されるべき入力パターンφに対し、認識システム RECOGNITRON が構造受精多段階変換パターン認識過程の第 t 段階でカテゴリ帰属知識$\langle\psi_t,\lambda_t\rangle$を持っているとき、式(3.14)の N(t)を考えているが、特に、不動点方程式（3.5）が成立する第 t 段階で、2.5.2項での“認識結果の3分類結果（i），(ii)，(iii)”の内、特に、(ii）が成立し、

$$\mathrm{card}(\gamma_t)=1,\ \text{つまり、}\ N(t)=m_0-1 \tag{3.19}$$

であることが望ましい。一般に、非カテゴリ候補の除去を意味する不等式

$$N(0)\leqq N(1)\leqq N(2)\leqq N(3)\leqq\cdots\leqq N(t-1)\leqq N(t)\leqq\ \cdots \tag{3.20}$$

が成立するが、この計数過程

$$N(0), N(1), N(2), N(3),\cdots, N(t-1), N(t),\cdots\in\{0,1,2,\cdots\} \tag{3.21}$$

を非定常ポアッソン確率過程とみる訳である。

ここで、Poisson確率分布を説明しておこう。

1次元**Poisson分布**とは、離散確率分布であって、確率変数 x が非負整数値 $n\in\{0,1,2,\cdots\}$をとる確率 prob{x＝n} が

$$\mathrm{prob}\{x=n\}=\exp(-\lambda)\cdot\lambda^n/(n!) \tag{3.22}$$

と与えられるものを指す。

Poisson分布では、次の3性質(i)，(ii)，(iii)が成り立つことがよく知られている：

（i）$\sum_{n=0}^{\infty}\mathrm{prob}\{x=n\}=1.$

（ii）平均値(mean) E(x)

$=\sum_{n=0}^{\infty}\ n\cdot\mathrm{prob}\{x=n\}=\lambda$ with E being the expectation operator.

（iii）分散(variance) $E([x-E(x)]^2)$

$=\sum_{n=0}^{\infty}(n-E(x))^2\cdot\mathrm{prob}\{x=n\}$

$=\lambda.$

□

式（3.14）の$N(t;\varphi)$は、離散時刻（認識段階数）tを固定すると、φを確率径数（probability parameter;parameter of distribution）とする確率変数となる。また、φをφ_0と固定すると、時間tの関数$N(t;\varphi_0)$が得られる。$N(t;\varphi_0)$を$\varphi=\varphi_0$に対する見本過程（sample process）という。

3.4 非定常ポアソン過程の強度関数と単位時間当たりの削除された非カテゴリ数の発見率、平均値関数、カテゴリ候補の絞り込み指数

本節では、不動点探索形構造受精多段階パターン認識過程が非定常ポアソン過程で表される事実をまず指摘した後、そのポアソン過程の強度関数が単位時間当たりの削除された非カテゴリ数の発見率になること、更に、平均値関数の、強度関数による積分表現、並びに、カテゴリ候補の絞り込み指数などが説明される。

3.4.1 計数過程$\{N(t)\}_{t\geq 0}$の3性質

以下の(iii)で登場している

・ランダウの記号$f(x)=o(g(x))$は、変数xの値がある極限に近づくとき、$f(x)/g(x)$が0に収束することの意　(3.23)

である。

式（3.21）の計数過程$\{N(t)\}_{t\geq 0}$は、つぎの3性質（i），(ii)，(iii）を満たすとする：

(i)（$t=0$では、削除されるカテゴリ候補は無い）

$$N(0;\varphi)=0. \tag{3.24}$$

(ii) 区間$0<\tau<t$の分割

$$t_0\equiv 0<t_1<t_2<\cdots<t_{n-1}<t_n\equiv t \tag{3.25}$$

での任意の分割点の集合$\{t_k\}$について、$N(t;\varphi)$は、

$$N(t;\varphi)=\sum_{k=1}^{n}[N(t_k;\varphi)-N(t_{k-1};\varphi)]+N(t_0;\varphi) \tag{3.26}$$

と分解できるが、

$$N(t_k;\varphi)-N(t_{k-1};\varphi),\ k=1,2,\cdots,n \tag{3.27}$$

が統計的に独立な加法過程（differential process）[15] である。

(iii)（確率連続な加法過程$N(t;\varphi)$は、φを固定したときの見本過程$N(t;\varphi)$が殆ど確実に、高さ1の飛躍のみで増加する連続階段関数である）

$$\mathrm{Prob}\{N(t+\triangle t;\varphi)-N(t;\varphi)=1\}=h(t)\cdot\triangle t+o(\triangle t) \tag{3.28}$$

$$\wedge\ \mathrm{Prob}\{N(t+\triangle t;\varphi)-N(t;\varphi)\geqq 2\}=o(\triangle t). \tag{3.29}$$

□

3.4.2 非定常ポアソン確率過程としての、不動点探索形構造受精多段階パターン認識過程の強度関数、平均値関数

文献 [15] のp.129,定義27.1（§27）より、上述の3性質（i），(ii)，(iii）を満たす式（3.13）、或いは、式（3.21）の計数過程

$$\{N(t)\}_{t\geq 0} \tag{3.30}$$

は、非定常ポアソン確率過程（nonhomogenuous stochastic Poisson process）となる。

式（3.28）の密度関数$h(t)$を、

非定常ポアソン確率過程の強度関数（intensity function）　(3.31)
という。2式（3.1），(3.14）の設定の下で、2式（3.2），(3.3）の何れかを初期条件として採用し、不動点方程式（3.5）の成立を終了条件とする“式（3.4）の不動点探索形構造受精多段階パターン認識過程”

$$\{TA(\mu_s)T\langle\psi_s,\lambda_s\rangle\}_{0\leqq s\leqq t} \quad (3.32)$$

においては、強度関数 h(t) は単位時間当たりの削除される非カテゴリ候補数の発見率を表す。

N(t) の期待値（expectation）

$$H(t)\equiv E(N(t))\equiv\sum_{n=0}^{\infty} n\cdot \mathrm{Prob}\{N(t)=n\} \quad (3.33)$$

は、**平均値関数**（mean value function）と呼ばれる。

式（3.33）に登場している“N(t)＝n となる確率 Prob{N(t)＝n}”は、

$$\mathrm{Prob}\{N(t)=n\}=[H(t)^n/n!]\cdot\exp(-H(t))$$
$$,n=0,1,2,\cdots \quad (3.34)$$

と表される。ポアソン過程の性質から、

$$\text{平均値 } \mathrm{Mean}(N(t))\equiv E(N(t))=H(t) \quad (3.35)$$
$$\text{分散 } \mathrm{Var}(N(t))\equiv E([n-E(N(t))]^2)$$
$$=\sum_{n=0}^{\infty}[n-E(N(t))]^2\cdot \mathrm{Prob}\{N(t)=n\}$$
$$=H(t) \quad (3.36)$$

が成り立つ。

式（3.33）の平均値関数 H(t) は、式（3.28）の強度関数 h(t) の、時間区間

$$[0,t]\equiv\{s\mid 0\leqq s\leqq t\} \quad (3.37)$$

にわたる積分で表されるというのが、次の定理3.1である。

［定理3.1］**（平均値関数定理）**

上述の3.4.1項での3性質（i)，(ii)，(iii）の下で、

$$H(t)\equiv E(N(t))=\int_0^t d\tau\ h(\tau). \quad (3.38)$$

（証明）区間 $0<\tau<t$ を、

$$t_k\equiv t_0+(k/n)\cdot(t_n-t_0),\ k=0,1,2,\cdots,n \quad (3.39)$$

として、式（3.25）の如く、n等分する。

さて、$t=t_n$として、式(3.26) が成り立っているが、2式 (3.24), (3.26) から成り立っている (i)の $N(t_0)=0$ を考慮すると、

$$N(t_n;\varphi)$$
$$=\sum_{k=1}^{n}[N(t_k;\varphi)-N(t_{k-1};\varphi)] \quad (3.40)$$

と書ける。

ところで、(iii）より、式（3.39）の

$$\triangle t\equiv(k/n)\cdot(t_n-t_0) \quad (3.41)$$

の近似として、$d\tau$を十分小にとれば、

$$1\times \mathrm{prob}\{N(\tau+d\tau;\varphi)-N(\tau;\varphi)=1\}$$
$$=d\tau\cdot h(\tau) \quad (3.42)$$

が成り立つことがわかる。(ii）より、式（3.27）は確率的に独立な増分であるから、2式（3.40)，(3.42）より、

$$k\times \mathrm{prob}\{N(t_k;\varphi)=k\}=d\tau\cdot h(t_k) \quad (3.43)$$

が成り立つことがわかる。よって、

$$\sum_{k=1}^{n} k \times \mathrm{prob}[N(t_k;\varphi)=k]$$
$$=\sum_{k=1}^{n} d\tau \cdot h(t_k) \tag{3.44}$$

を得て、$d\tau$ は十分小ととってあるから、これは、式（3.38）の成立を意味する。 □

尚、文献［15］のp.131（§27）の定理27.2を適用して、$N(t)-N(s)$の分布は、

$H(t)-H(s)$ を平均とするPoisson分布である (3.45)

ことが示されるから、上述の定理3.1の成立が言えるのである。

3.4.3 非カテゴリ候補の発見率、カテゴリ候補絞り込みの指数

2式 (3.1), (3.14) の設定の下では、

$$H(\infty)(=E(N(\infty)))\equiv \lim_{t\to\infty} H(t)$$
$$=\int_0^{\infty} d\tau\ h(\tau)$$
$$\equiv m \tag{3.46}$$

は、不動点探索の最終段階までに削除された非カテゴリ候補の総数であり、式（3.19）との関係でいえば、

$$m=m_0-1 \tag{3.47}$$

が成り立つことが望ましい。式（3.46）の $H(\infty)$ は、mを平均とするポアソン分布に従う。

時刻 t において解候補となっているカテゴリの総数の期待値 $G(t)$ は、

$$G(t)\equiv E(N(\infty)-N(t))$$
$$=m-H(t) \quad \because\ 2式\ (3.46), (3.33)$$
$$=\int_0^{\infty} d\tau\ h(\tau)-\int_0^{t} d\tau\, h(\tau) \quad \because\ 2式\ (3.46), (3.38)$$
$$=\int_t^{\infty} d\tau\ h(\tau) \tag{3.48}$$

と表される。

さて、時刻 t でカテゴリ候補の削除が起こったという条件の下で、時間区間

$$(t, t+h]\equiv\{\tau \mid t<\tau\leqq t+h\} \tag{3.49}$$

においてカテゴリ候補の削除が起こらない確率 $R(h/t)$ は、h_1 を十分小にとれば、

$$R(h/t)\equiv R(h/t;\varphi)$$
$$=\mathrm{Prob}\{N(t)-N(t-h_1)=1 \wedge$$
$$N(t)-N(t+h)=0\}$$
$$/\mathrm{Prob}\{N(t)-N(t-h_1)=1\} \tag{3.50}$$

であるが、式（3.45）を考慮すると、また、

$N(t)-N(t-h_1)$, $N(t)-N(t+h)$ は独立

であるから、不動点探索形構造受精多段階パターン認識過程の信頼度（reliability of the recognition process）と呼ばれる式（3.50）の $R(h/t)$ は、

$$R(h/t)$$
$$=\mathrm{Prob}\{N(t)-N(t-h_1)=1\}\cdot \mathrm{Prob}\{N(t)-N(t+h)=0\}/\mathrm{Prob}\{N(t)-N(t-h_1)=1\}$$
$$=\mathrm{Prob}\{N(t)-N(t+h)=0\}$$
$$=\exp(-\{H(t+h)-H(t)\})\ (t\geqq 0, h\geqq 0)$$
$$\because\ 式\ (3.34) \tag{3.51}$$

と再表現される。

非定常ポアソン過程モデルに於いては、式（3.33）の平均値関数 $H(t)$、或いは、式（3.28）の強度関数 $h(t)$ を決定すれば、その確率的挙動は決まる。

$$D(t;\varphi)\equiv h(t)/G(t)$$

$$=h(t)/[m-H(t)] \quad \because \quad 式(3.48) \qquad (3.52)$$

は、

時刻 t に於ける残存カテゴリ候補1個当たりの、削除された非カテゴリ候補数の発見率（the rate of discovery per candidate-category） (3.53)

を表している。

最後に、

$$1-FND(h/t)$$

＝［時間区間 $(t, t+h]$ において解候補から削除された非カテゴリ候補の総数の期待値］

/［時間 t において解候補となっているカテゴリ候補の総数の期待値］

$$=[H(t+h)-H(t)]/[m-H(t)]$$

$$=[-G(t+h)+G(t)]/G(t)$$

$$=1-G(t+h)/G(t) \qquad (3.54)$$

$$\therefore \quad \lim_{h\to 0}(1/h)\cdot[1-FND(h/t)]$$

$$=-[dG(t)/dt]/G(t) \qquad (3.55)$$

が成り立つから、表現

$$FND(h/t)=G(t+h)/G(t) \qquad (3.56)$$

を得て、FND(h/t) は、

FND(h/t)

＝［時刻 $t+h$ において解候補となっているカテゴリの総数の期待値］

/［時刻 t において解候補となっているカテゴリの総数の期待値］ (3.57)

を表しており、カテゴリ候補絞り込みの指数（figure of narrowing category-candidate down）と呼ばれる。

以上で説明されたのは、

①式(3.38)の $H(t)$

②式(3.46)の $m=H(\infty)$

③式(3.28)の $h(t)=dH(t)/dt$

④式(3.48)の $G(t)$

⑤式(3.52)の $D(t)=h(t)/\int_t^\infty d\tau\ h(\tau)$

⑥式(3.51)の $R(h/t)$

⑦式(3.56)の $FND(h/t)$

$$=\int_{t+h}^\infty d\tau\ h(\tau)/\int_t^\infty d\tau\ h(\tau)$$

の7つの量である。この内、本研究で独自に導入されたのは、FND(h/t) である。

3.5 1例

単位時間当たりに発見される非カテゴリ候補の数は、その時刻の残存するカテゴリ候補の数に比例するとして、微分方程式

$h(t)=dH(t)/dt=b[m-H(t)]$ (3.58)

を導入してみよう。mは、

認識過程の開始前に潜在する総期待非カテゴリ候補の数

である。また、定数bは非カテゴリ候補の発見率を表す比例定数である。

H(t)の初期条件として、

$H(0)=0$ (3.59)

を採用し、この微分方程式（3.58）を解けば、

$H(t)=m[1-\exp(-bt)]$ (3.60)

$h(t)=mb\cdot\exp(-bt)$ (3.61)

となる。式（3.60）のH（t）は、付録Cの式（C1）に一致し、指数型SRGMに対応したものであることがわかる。

興味ある諸量を具体的に計算してみれば、

（イ）式(3.48)の$G(t)=\int_t^{\infty} d\tau\ h(\tau)$

$=m-H(t)=m\cdot\exp(-bt)$

（ロ）式(3.52)の$D(t;\varphi)\equiv h(t)/G(t)$

$=mb\cdot\exp(-bt)/[m\cdot\exp(-bt)]$

$=b$

（ハ）式(3.51)の$R(h/t)$

$=\exp(-\{H(t+h)-H(t)\})\ (t\geqq 0, h\geqq 0)$

$=\exp[-m\{\exp(-bt)-\exp(-b(t+h))\}]$

（ニ）式(3.56)の$FND(h/t)=G(t+h)/G(t)$

$=\exp(-bh)$

である。

3.6 パラメータの最尤推定

文献［29］の第3章に倣って、式（3.13）の非定常ポアソン過程のパラメータを最尤推定（the maximum-likelihood estimates of the model parameters）する手法を説明してみよう（付録Dを参照）。

式（3.13）の非定常ポアソン過程は、式（3.38）で表現される式（3.3）のH(t)は、

$w_0\equiv H(\infty)=m$と、その他のu個のパラメータ

$w_i\ (i=1,2,\cdots,u)$ (3.62)

があるものとする。

式(3.25)と同様に、

区間$(0,t_n]\equiv\{t\,|\,0<t\leqq t_n\}$を、

$t_0\equiv 0<t_1<t_2<\cdots<t_{n-1}<t_n$ (3.63)

と分割して得られた時間区間

$(0,t_k]\equiv\{t\,|\,0<t\leqq t_k\}$ (3.64)

において、時刻t_kでy_k個の非カテゴリ候補が観測されたとしよう。

$\underline{t}=\langle t_1,t_2,\cdots,t_n\rangle$, $\underline{w}=\langle w_0,w_1,w_2,\cdots,w_u\rangle$,

$\underline{y}=\langle y_1,y_2,\cdots,y_n\rangle$ (3.65)

として、

$$L \equiv L(\underline{w}/\underline{t},\ \underline{y})$$
$$= \prod_{k=1}^{n} [[\{H(t_k)-H(t_{k-1})\}^{y_k-y_{k-1}}/\ \{(y_k-y_{k-1})\ !\}]\cdot \exp[-\{H(t_k)-H(t_{k-1})\}]] \tag{3.66}$$
が尤度関数（the likelihood function for the unknown model-parameters）となり、パラメータ $\underline{w}$ はこのLを最大とするように、選ばれていると考えられる。このための手法が最尤推定法（method of maximum-likelihood）である。

最尤法を具体的に適用しよう。Lの対数をとった対数尤度

$$\ell \equiv \ell(\underline{w}/\underline{t},\ \underline{y})$$
$$= \log_e L(\underline{w}/\underline{t},\ \underline{y})$$
$$= \sum_{k=1}^{n} [y_k - y_{k-1}]\cdot \log_e [H(t_k)-H(t_{k-1})]$$
$$- H(t_n) - \sum_{k=1}^{n} \log_e[(y_k - y_{k-1})\ !] \tag{3.67}$$
について、Lが最大となるための必要条件としての尤度方程式（the simultaneous likelihood equations）

$$\partial \ell(\underline{w}/\underline{t},\ \underline{y})/\partial w_i = 0,\ i = 0,1,2,\cdots,n \tag{3.68}$$
を満たす w を求めれば良い。

観測データ 〈$\underline{t}$, $\underline{y}$〉が 与えられたとき、尤度方程式（連立非線形方程式）(3.68) を解き、パラメータ w を推定する手法が最尤推定法である。最尤推定法は、Lが全域で最大となる十分条件を満たすとは限らないことに注意しておく。

4. 並列認識過程の分析

本章では、第1章での ②カテゴリ候補の並列的絞り込み方法 に限って、ソフトウェアに残存するエラーの個数を推定するときに、経験的に中小規模のソフトウェアの信頼性モデルによく適合すると言われている **"遅れS字型ソフトウェア信頼度成長モデル"** に対応する結果を導く。

4.1 非カテゴリ候補の検出に関する遅れS字型モデル

前前章で非カテゴリ候補の検出に関する不動点方程式 (2.15), (2.40) の説明が与えられ、前章でこの検出に関し確率過程としてのソフトウェア信頼度成長モデルが適用された不動点探索型構造受精多段階パターン認識の働きを分析するにあたって、式 (3.38) の平均値関数 H(t) として、**遅れS字型モデル**を採用する。

「非カテゴリ候補の除去数は残存非カテゴリ候補数に比例するが、非カテゴリ候補検出から非カテゴリ候補の除去迄時間を要する」と解釈し、時刻 t 迄に検出された期待非カテゴリ候補数である式 (3.38) のH(t)を、

$$H(t) = m[1-(1+bt)\exp(-bt)]$$
$$= m - m\cdot\exp(-bt) - mbt\cdot\exp(-bt),\ m,b > 0 \tag{4.1}$$
で表したモデルである。ここで、m,b は、

m：最終的に検出される期待非カテゴリ候補数
(the expected number of non-candidates to be
eventually detected for the input pattern)　(4.2)

b：非カテゴリ候補出現率

(the occurrence rate of non-candidates for the input pattern) (4.3)

である。

4.2 遅れS字型モデルによる不動点探索型構造受精多段階パターン認識の働きの分析

前節の設定に従って、3.4.3項の①〜⑦の7量を求めよう。

①の $H(t)$ については、式(4.1)で与えられている。

②の式(3.46)の $m=H(\infty)$ については成立している。

③の式(3.28)の $h(t)=dH(t)/dt$ については、

$h(t)$

$=(d/dt)[m-m\cdot\exp(-bt)-mbt\cdot\exp(-bt)]$ ∵ 式(4.1)

$$=mb^2t\cdot\exp(-bt) \tag{4.4}$$

と求められる。

$H(0)=0$ が成り立っており、微分方程式

$$dH(t)/dt=b[m\{1-\exp(-bt)\}-H(t)] \tag{4.5}$$

が成り立っている。因みに、式(4.5)の証明は次の通りである。

$b[m\{1-\exp(-bt)\}-H(t)]$

$=mb-mb\cdot\exp(-bt)-b[m-m\cdot\exp(-bt)-mbt\cdot\exp(-bt)]$

$=mb^2t\cdot\exp(-bt)$

$$=h(t)=dH(t)/dt \tag{4.6}$$

□

④の式(3.48)の $G(t)$ については、

$G(t)=m-H(t)$

$$=m(1+bt)\exp(-bt) \quad ∵ 式(4.1) \tag{4.7}$$

と求められる。

⑤の式(3.52)の $D(t)=h(t)/\int_t^\infty d\tau\ h(\tau)$ については、

$D(t)=h(t)/G(t)$

$=mb^2t\cdot\exp(-bt)/[m(1+bt)\exp(-bt)]$ ∵ 2式(4.4), (4.7)

$$=b^2t/[1+bt] \tag{4.8}$$

と求められる。よって、

$$\lim_{t\to\infty} D(t)=b \tag{4.9}$$

も成り立つ。

⑦の式(3.51)の $FND(h/t)=G(t+h)/G(t)$ については、

$FND(h/t)=m[1+b(t+h)]\exp[-b(t+h)]/[m(1+bt)\exp(-bt)]$ ∵ 式(4.7)

$$=[1+bh/\{1+bt\}]\cdot\exp(-bh) \tag{4.10}$$

である。

⑥式(3.51)の $R(h/t)=\exp(-\{H(t+h)-H(t)\})\ (t\geqq 0, h\geqq 0)$ については、

$R(h/t)$

$=\exp[-m\{(1+bt)\exp(-bt)$

$$-(1+b(t+h))\exp(-b(t+h))\}] \quad ∵ 式(4.1) \tag{4.11}$$

4.3　パラメータ b の決定

式(4.1)の平均値関数 H(t)内のパラメータ m については、式(3.47)で与えられている。残りのパラメータ b を3.6節の手法に従って決定することを考えよう。

時刻 t に於ける残存カテゴリ候補1個当たりの、削除された非カテゴリ候補数の発見率 D(t)に関し、式(4.9)が成り立っているが、D(t)は直接観測可能な量ではない。それで、時刻 t において解候補となっているカテゴリの総数の期待値である式(3.48)のG(t)を使う。

$$G(t_1)=k_1, G(t_2)=k_2, \cdots, G(t_n)=k_n \tag{4.12}$$

が観測されたとする。各 $G(t_i)$が平均値 k_i、分散 σ^2 の正規分布をするとすれば、対数尤度

$$\log_e \prod_{i=1}^{n} (1/\sqrt{2\pi\sigma^2} \cdot \exp[-(2\sigma^2)^{-1} \cdot \{G(t_i)-k_i\}^2] \tag{4.13}$$

を最大とする b を決定すればよい。それには、

$$F(b)=(1/2) \cdot \sum_{i=1}^{n} [G(t_i)-k_i]^2 \rightarrow \min \tag{4.14}$$

となるように、b を決定すればよいことになる。

$$F(b)$$
$$=(1/2) \cdot \sum_{i=1}^{n} [m \cdot \exp(-bt_i)+mbt_i \cdot \exp(-bt_i)-k_i]^2 \quad \because \quad \text{式}(4.7) \tag{4.15}$$

であるから、

$$0=(\partial/\partial b) F(b)$$
$$=\sum_{i=1}^{n} [m \cdot \exp(-bt_i)+mbt_i \cdot \exp(-bt_i)-k_i] \cdot$$
$$[-mt_i \cdot \exp(-bt_i)+mt_i \cdot \exp(-bt_i)-mbt_i^2 \cdot \exp(-bt_i)]$$
$$=-\sum_{i=1}^{n} [m \cdot \exp(-bt_i)+mbt_i \cdot \exp(-bt_i)-k_i] \cdot mbt_i^2 \cdot \exp(-bt_i)] \tag{4.16}$$

から、b を決定すればよい。

1つの解を与えておけば、次のようになる：

$n=1$ のとき、t_1を十分小さく選んでいれば、

$$b$$
$$\fallingdotseq (2t_1)^{-1} \cdot \log_e(m/k_1) \cdot [1+[1+8/\log_e(m/k_1)]^{1/2}]. \tag{4.17}$$

式(4.16)を導いておこう。

$n=1$のとき、式(4.16)から、

$$m \cdot (1+bt_1)\exp(-bt_1)=k_1 \tag{4.18}$$

が得られる。式（4.19）を変形すれば、

$$\log_e(1+bt_1)-bt_1=\log_e(k_1/m) \tag{4.19}$$

となる。

ここで、a,z>0 であれば成り立つ対数近似式

$$\log_e(a+z)$$
$$=\log_e a+2 \cdot [\{z/(2a+z)\}+(1/3) \cdot \{z^3/(2a+z)^3\}+(1/5) \cdot \{z^5/(2a+z)^5\}+\cdots]$$
$$\fallingdotseq \log_e a+2z/(2a+z) \quad \cdots z\text{が十分小さいとき} \tag{4.20}$$

を、$a=1, z=bt_1$ として、式（4.19）に対し適用すれば、

$$\text{loge}1+2bt_1/(2+bt_1)-bt_1 \fallingdotseq \log_e(k_1/m) \tag{4.21}$$

が得られる。

式（4.21）を変形すれば、

$$x=bt_1 \tag{4.22}$$

として、

$x^2 + ax + c \fallingdotseq 0$ (4.23)

$, \text{where } a = \log_e(k_1/m) \text{ and } c = 2a$ (4.24)

となる。

変数 x の2次方程式 (4.23) を解けば、

$x = 2^{-1} \cdot [-a \pm \sqrt{a^2 - 4c}]$ (4.25)

である。

2式(4.22), (4.25)の関係を使って、b を求めれば、

$b \fallingdotseq (2t_1)^{-1} \cdot \log_e(m/k_1) \cdot [1 \pm [1 + 8/\log_e(m/k_1)]^{1/2}]$ (4.26)

が得られるが、3式 (3.52), (4.8), (4.9) を考慮して、式 (4.26) の1つを選んだものが式 (4.17) である。

5. むすび

処理の対象とする問題の入力パターン φ のカテゴリ依存構造を数理で表現すれば、パターン認識の働きの本質が現れると見て、パターン φ の帰属に関する間違った、φ 内のカテゴリ成分が式 (2.8)～式 (2.15) の多段認識過程を lock in してしまう前に、

パターン φ の帰属に関する、φ 内の正しいカテゴリ成分の優位を、初期の認識段階において慎重に助長していくことの重要性を取り入れることのできる連想形パターン認識技術

の基礎として、パターン φ のカテゴリ帰属知識 $\langle \varphi, \gamma \rangle$ が直交分解できることが証明されている (文献 [55] の定理7.1)。

残っている課せられた基本的な問題とは、

パターン φ の帰属に関する複数の可能な結果の内、その特定の1つがどのように、式(2.8)～式(2.15) の多段認識過程の進展と共に、選択されて来るのか？

が、順を追って説明されることだが (SS理論 [54])、この説明の前哨戦として、

本論文では、不動点探索形構造受精多段階変換によるパターン認識の働きを備えている認識システムRECOGNITRON の入力パターン φ を認識する過程が

φ の帰属するカテゴリ候補を１つ２つと削除し、

絞り込んでゆく計数過程 $N(t) \geqq 0$ である

と見做された。入力パターン φ を認識する過程とは、カテゴリ候補の絞り込みの働きが発現する知能現象と見て、カテゴリ候補探索に関するこの知能の働きを具体的には、非定常ポアソン確率過程として記述した。

ソフトウェア信頼度成長モデル（ソフトウェア開発プロセスの下流工程にあるテスト工程、或いは、運用段階において、その上流工程で潜入したソフトウェアの欠陥や誤りなどのフォールトの発見事象やソフトウェア故障の発生現象のモデル [35]）の理論での、

個数計測モデル（ソフトウェア故障発生数、或いは、フォールト発見数に着目した確率・統計モデル）[37]

を適用するため、

ソフトウェア故障が引き起こすフォールト

⟷取り除かれるカテゴリ候補

という対応を想定し、ソフトウェアに残存するエラーの個数を推定する時、経験的に中小規模のソフトウェアの信頼性のモデルによく適合するといわれている“遅れS字型ソフトウェア信頼度成長モデル”に対応した分析を行った。

認識段階数 t が非負整数値をとるときの、処理の対象とする問題の入力パターンφに関わるカテゴリ帰属知識の変換方程式 (2.14) は、パターンについては、

$$[\psi_{t+1}-\psi_t]/[(t+1)-t]$$
$$=_{\triangle}-\psi_t+TA(\mu_t\cap\lambda_t)T\psi_t\in\Phi \tag{5.1}$$

となり、カテゴリ候補集合の決定方程式については、

$$CSF(\psi_{t+1},\mu_{t+1}\cap\lambda_{t+1})\in 2^J \tag{5.2}$$

が有効なカテゴリ候補の番号リストである。

カテゴリ帰属知識の和“+”[55] と差“−”とが適切に定義できるとすると、カテゴリ帰属知識の変換方程式 (2.14) は、2式 (5.1), (5.2) から、

$$[\langle\psi_{t+1},\lambda_{t+1}\rangle-\langle\psi_t,\lambda_t\rangle]/[(t+1)-t]$$
$$=_{\triangle}-\langle\psi_t,\lambda_t\rangle+TA(\mu_t)T\langle\psi_t,\lambda_t\rangle\in\langle\Phi,2^J\rangle \tag{5.3}$$

と書き直される。よって、認識段階数 t が実数値をとるときは、パターンの変換方程式

$$d\psi_t/dt=-\psi_t+TA(\mu_t\cap\lambda_t)T\psi_t\in\Phi \tag{5.4}$$

と、カテゴリ候補集合の決定方程式

$$CSF(\psi_t,\mu_t\cap\lambda_t)\in 2^J \tag{5.5}$$

とは、カテゴリ帰属知識$\langle\psi_t,\lambda_t\rangle$の変換方程式

$$d\langle\psi_t,\lambda_t\rangle/dt$$
$$=_{\triangle}-\langle\psi_t,\lambda_t\rangle+TA(\mu_t)T\langle\psi_t,\lambda_t\rangle\in\langle\Phi,2^J\rangle \tag{5.6}$$

にまとめられる。

認識段階数 t が実数値をとるこの場合の、不動点探索型構造受精多段階パターン認識の働きの挙動の解析、並びに、非カテゴリ候補を除去していくときのその確率論的性質を研究することは、残された課題である。

文　献

[1] 青木利夫, 高橋渉：“集合・位相空間要論”, 培風館, Sept.1979

[2] Angus E. Taylor, David C. Lay: “Introduction to function analysis”, John Wiley & Sons, Inc., 1980

[3] Edited by W. K. Estes : “Handbook of learning and cognitive processes (Volume 4 Attention and memory)”, Lawrence Erlbaum Associates, Inc.,1976

[4] Nils J. Nillson : “Problem-solving methods in artificial intelligence”, McGraw-Hill Book Company, Inc.,1971

[5] Elaine Rich and Kevin Knight : “Artificial intelligence (Second edition)”, McGraw-Hill, Inc.,1991

[6] Abhijit S. Pandya and Robert B. Macy: “Pattern recognition with neural networks in C++”, A CRC Book Published in Cooperation with IEEE,1996

[7] M.A.アービップ：“脳ー思考と行動の源を探るー”, 金子隆芳訳, サイエンス社, Apr.1980

[8] ルーメルハート：“人間の情報処理（新しい認知心理学へのいざない）”, 御領謙訳, サイエンス社, Sept.1980
[9] 長尾真：“パターン情報処理（電子通信学会大学シリーズⅠ-4）”, コロナ社, Mar.1983
[10] 太原育夫：“認知情報処理”, オーム社, Mar.1991
[11] Jeffrey Wood：“Invariant pattern recognition: A review”, Pattern Recognition, vol.29, no.1, pp.1-17, 1996
[12] 池田克夫, 田村秀行, 全炳東：“知能情報メディアーマルチメディアの進化形ー”, 電子情報通信学会誌, vol.79, no.8, p.p.788-792, Aug.1996
[13] 斎藤嘉博：“信頼性の基礎数学（信頼性工学講座第2巻）”, 東京電機大学出版局, May 1975
[14] Shigeru Yamada：“Software quality/reliability measurement and assessment : Software reliability growth models and data analysis”, Journal of Information Processing, vol.14, no.3, pp.254-266, 1991
[15] 伊藤清：“確率論（現代数学14）”, 岩波書店, Nov.1966
[16] S.Yamada, M.Ohba and S.Osaki: “S-shaped reliability growth modeling for software error detectipon”, IEEE Trans. Reliability R-32, pp.475-478, 1983
[17] 丸山儀四郎：“確率および統計（基礎数学講座10）”, 共立出版, Feb.1963
[18] 高井峰生, 山城登久二, 成田誠之助：“Synchronous Conservative Algorithm を用いた離散事象並列シミュレーションにおける性能予測”, 電子情報通信学会論文誌(D-Ⅰ), vol.J80-D-Ⅰ, no.3, pp.237-246, Mar.1997
[19] 坂本慶行, 石黒真木夫, 北川源四郎：“情報量統計学（情報科学講座A･5･4）”, 共立出版, June 1993
[20] 中川豊：“ソフトウェア信頼度成長曲線の分析に基づく連結指数形ソフトウェア信頼度成長モデル”, 電子情報通信学会論文誌（D-Ⅰ), vol.J77-D-Ⅰ, no.6, pp.433-442, June 1994
[21] 山田茂, 半谷知久, 尾崎俊治：“一般化された遅延S字形ソフトウェア信頼度成長モデルとその適合性評価に関する考察”, 電子情報通信学会論文誌（D-Ⅰ), vol.J76-D-Ⅰ, no.11, pp.613-620, Nov. 1993
[22] 高田義広, 松本健一, 鳥居宏次：“ニューラルネットを用いたソフトウェア信頼性予測モデル”, 電子情報通信学会論文誌 (D-Ⅰ), vol. J77-D-Ⅰ, no.6, pp.454-461, June 1994
[23] 尾崎俊治：“非定常ポアソン過程モデル”, 情報処理（情報処理学会誌), vol.31, no.12, pp.1631-1640, Dec.1990
[24] Hisanao Ogura：“Orthogonal functionals of the Poisson process”, IEEE Trans. on Information Theory, vol.IT-18, no.4, July 1972
[25] Masami Noro, Kunio Goto and Katsushige Sawaki：“A new software reliability growth predicted on counting processes for instruction execution”, Transactions of Information Processing Society of Japan, vol.35, no.12, pp.2623-2630, Dec.1994
[26] 当麻喜弘：“超幾何分布にもとづくソフトウェア残存フォールト数推定モデル”, 情報処理（情報処理学会誌), vol.31, no.12, pp.1641-1646, Dec.1990
[27] 大場充：“ソフトウェア信頼モデル入門”, 情報処理（情報処理学会誌), vol.31, no.12, pp.1623-1630, Dec.1990
[28] 松本健一, 菊野亨, 鳥居宏次：“S字型ソフトウェア信頼度成長モデルの大学環境における実験的評価ー推定精度の比較と習熟係数の決定ー”, 電子情報通信学会論文誌（D-Ⅰ), vol.J73-D-Ⅰ,

no.2, pp.175-182, Feb.1990
[29] 大寺浩志, 山田茂, 成久洋之："テスト空間を考慮したソフトウェア信頼度成長モデル", 電子情報通信学会論文誌 (D-I), vol.J73-D-I, no.2, pp.170-174, Feb.1990
[30] 大寺浩志, 山田茂, 成久洋之："ソフトウェア信頼度成長モデルによるテスト工程管理", 電子情報通信学会論文誌 (D), vol.J70-D, no.5, pp.889-895, May 1987
[31] 大橋守, 松田啓介："nタイプのエラーを考慮したソフトウェア最適リリース時間の性質について", 情報処理 (情報処理学会誌), vol.36, no.2, pp.463-469, Feb.1995
[32] 澤田清, 山道弘明, 藤井進："Kullback-Leiblerの情報量に基づくソフトウェアの信頼性実証試験に関する離散型モデル", 電子情報通信学会論文誌(A), vol.J79-A, no.3, pp.830-833, Mar.1996
[33] 木村光宏, 山田茂："テスト項目の消化過程に基づくソフトウェアテスト進捗度評価モデルに関する考察", 電子情報通信学会論文誌 (D-I), vol.J79-D-I, no.12, pp.1211-1217, Dec.1996
[34] 島和之, 松本健一, 鳥居宏次："ソフトウェアフォールごとの故障生起頻度の違いについての考察", 電子情報通信学会論文誌 (D-I), vol.J79-D-I, no.12, pp.1203-1210, Dec.1996
[35] 山田茂："潜入フォールトによる不完全デバッグを考慮したソフトウェア信頼度成長モデル", 電子情報通信学会論文誌(A), vol.J80-A, no.2, pp.363-370, Feb.1997
[36] 今泉充啓, 安井一民, 中川とし夫："処理完了時限をもつマイクロプロセッサシステムの信頼性評価", 情報処理学論文誌, vol.38, no.2, pp.370-376, Feb.1997
[37] 得能貢一, 山田茂："ソフトウェア可用性評価のためのアベイラビリティモデル", コンピュータソフトウェア, vol.14, no.2, pp.142-148, Mar.1997
[38] 池田浩二, 小松隆行, 長岡浩司："ボルツマンマシンを用いた天気データのモデリング", 電子情報通信学会論文誌 (D-II), vol.J77-D-II, no.6, pp.1169-1172, June 1994
[39] 和田安弘, 川人光男："新しい情報量基準とCross Validationによる汎化能力の推定", 電子情報通信学会論文誌 (D-II), vol.J74-D-II, no.7, pp.955-965, July 1991
[40] 鈴木昇一："誤差確率分布を考慮した誤差逆伝播学習", 情報研究 (文教大学・情報学部), vol. 13, pp.173-202, Dec.1992
[41] 鈴木昇一："認識工学 (上)", 柏書房, Feb.1975
[42] 鈴木昇一： "マルチメディア情報処理のための一般化誤差逆伝播学習ニューラルネットの新数理", 近代文芸社, Sept.1996
[43] 鈴木昇一："抽出された特徴による手書き漢字構造の再生", 情報処理学会誌, vol.18, no.11, pp.115-1122, Nov.1977
[44] 鈴木昇一："パターンのエントロピーモデル", 電子情報通信学会論文誌 (D-II), vol.J77-D-II, no.10, pp.2220-2238, Nov.1994
[45] 鈴木昇一, 斉藤静昭, 奥野治雄, 太田芳雄："画像の復元とその計算機シュミレーション", 工学院大学研究報告, no.39, pp.198-206. Jan.1976
[46] 鈴木昇一："回転群と画像の分解・強調・構造化再構成に関する計算機シュミレーション", 情報研究 (文教大学情報学部), vol.4, pp.36-56, Dec.1983
[47] 鈴木昇一："連想形記憶器 MEMOTRON と日本語母音系列の再生に関する計算機シミュレーション", 情報研究 (文教大学情報学部), vol.7, pp.14-29, Dec.1986
[48] 鈴木昇一："収縮写像の構成用空間回路とその計算機シミュレーション", 情報研究 (文教大学・情報学部), vol.9, pp.17-29, Dec.1988

[49] 鈴木昇一："多変量解析に基づく大分類関数の決定とその計算機シミュレーション", 情報研究(文教大学・情報学部), vol.10, pp.35-49, Dec.1989
[50] 鈴木昇一："帰属係数法に基づく類似度、帰属関係あいまい度、認識情報量の計算機シミュレーション"、情報研究(文教大学・情報学部), vol.11, pp.51-68, Dec.1990
[51] 鈴木昇一, 前田英明："パターンの変形理論", 情報研究(文教大学・情報学部), vol.16, pp.209 267, Dec.1995
[52] 鈴木昇一, 佐久間拓也, 前田英明："数理形態学における諸演算とモデル構成作用素", 情報研究(文教大学・情報学部), vol.17, pp.133-170, Dec.1996
[53] 鈴木昇一："Radial-basis function networks, wavelet-based networks を用いたモデル構成作用素の構成法", 情報研究(文教大学・情報学部), vol.17, pp.71-131, Dec.1996
[54] 鈴木昇一："パターン認識の数学的理論", 電子情報通信学会技術研究報告[パターン認識と学習, パターン認識と理解], PRL84-6(第1部), PRL84-30, PRL84-38, PRL85-27, PRU86-8, PRU86-35, PRU86-69, PRU87-1, PRU87-28, PRU88-30, PRU89-1, PRU89-27, PRU89-40, PRU89-66, PRU89-77, PRU89-136, PRU90-5, PRU90-15, PRU90-29, PRU90-125, PRU91-1, PRU91-29, PRU91-42, PRU92-1, PRU92-18, PRU92-25, PRU92-89, PRU92-102(第28部), May 1984～Jan.1993
[55] 鈴木昇一："知能情報メディア情報処理のためのカテゴリ帰属知識を用いたパターン認識問題の数理的一般解決", 近代文芸社, June 1997
[56] 河田敬義, 丸山文行："基礎課程 数理統計", 裳華房, p.59, p.88, Mar.1963
[57] S. Kullback and R. A. Leibler: "On information and sufficiency", Annals of Mathematical Statistics, , vol.22, pp.79-86, 1951
[58] Hirotugu Akaike: "A new look at the statistical model identification", IEEE Trans. on Automatic Control, vol.AC-19, no.6, Dec.1974
[59] 竹内啓："AIC基準による統計的モデルの選択をめぐって", 計測と制御, vol.22, no.5, pp.445-453, May 1983
[60] H. Tong: "Determination of the order of a Markov chain by Akaike's information criterion", J. Applied Probability, vol.12, 1975
[61] I. Csiszar: "I-Divergence geometry of probability distributions and minimization problems", The Annals of Probability, vol.3, no.1, pp.146-158, 1975
[62] Gideon Schwarz: "Estimating the dimension of a model", The Annals of Statistics, vol.6, no.2, pp. 461-464, 1978
[63] Amrit L.Goel: "Software reliability models: Assumptions , limitations, and applicability", IEEE Trans. on Software Engineering, vol.SE-11, no.12, Dec.1985
[64] 堀越力, 鈴木智, 中根一成："AIC(赤池の情報量基準)を用いた最少部品による形状記述", 電子情報通信学会論文誌(D-II), vol.J77-D-II, no.9, pp.1691-1700, Sept.1994
[65] Koichi Tokuno and Shigeru Yamada: "A Markovian Software Availability Measurement with a geometrically decreasing failure-occurrence rate", IEICE Trans. Fundamentals, vol.E78-A, no.6, June 1985
[66] 高橋将人, 鈴木和幸："複数修理保全モデルにおける最適保全方策に関する一考察", 電子情報通信学会論文誌A, vol.J80-A, no.4, pp.677-683, Apr.1997

[67] 村田厚生："長期記憶からの検索過程における活性化に関する考察", 電子情報通信学会論文誌 A, vol.J80-A, no.4, pp.712-715, Apr.1997

付録A．信頼性モデルと、逐次的認識

本付録Aでは、不等式（2.42）を満たすカテゴリ番号の列

$\lambda_0, \lambda_1, \lambda_2, \cdots, \lambda_{t-2}, \lambda_{t-1}, \lambda_t (\subseteq J)$ (A1)

がランダムに選ばれ、パターン認識過程に残存するカテゴリ候補の個数を推定するためのモデルを、第1章の①の場合に限って、構築する場合などについて、信頼性モデルに関連した事柄が説明される。

A1. 指数型確率分布

指数(exponential)型分布[17]の確率密度 f(t)は、

$$f(t) = \begin{cases} \lambda \exp(-\lambda t) \cdots t \geqq 0 \text{ のとき} \\ 0 \qquad \cdots t < 0 \text{ のとき} \end{cases}$$

ここに、$\lambda > 0$ (A2)

であり、分布関数F（t）は、

$$F(t) \equiv \int_0^t d\tau\ f(\tau) = \begin{cases} 1 - \exp(-\lambda t) \cdots 0 \leqq t < \infty \text{ のとき} \\ 0 \qquad \cdots t < 0 \text{ のとき} \end{cases} \tag{A3}$$

と計算される[17]。平均値（mean value）E, 分散（variance）Var は、各々、

$$E \equiv \int_0^\infty d\tau\ \ \tau \cdot f(\tau) = (1/\lambda) \tag{A4}$$

$$Var \equiv \int_0^\infty d\tau\ \ [\tau - E]^2 \cdot f(\tau) = (1/\lambda)^2 \tag{A5}$$

と計算される。

平均寿命が $1/\lambda$ である偶発故障が期間（0,t）の間に発生する確率は、実は、式（A3）の F(t)のことである。

A2. 不動点探索形構造受精認識の探索確率

A2.1 平均探索時間

情報探索モデルでは、通常、図書館で n 冊の本からある本を探索時間 $t \geqq 0$内に発見する確率 q(t)は、式(A3)を勘案して、

$$q(t) = 1 - \exp(-(\lambda/n) \cdot t) \tag{A6}$$

とすることがある。長期記憶からの検索過程の数理モデルも、この式（A6）で表されるとした研究があり、検索過程の前半と後半で活性化の状況が異なるとして、より複雑な数理モデルも研究されている［67]。同様に、14式（2.5）～（2.18）から成る**不動点探索形**構造受精変換**認識過程**においても、処理の対象とする問題のパターン φ の帰属するカテゴリ（カテゴリ候補）を、1個だけに絞って時刻 $t \geqq 0$ 内に発見する確率（**探索が終了する確率**）p(t)を、$a > 0$ というパラメータを導入し、

m_0：カテゴリ総数 (A7)

として、式（A2）のタイプの指数型確率分布を想定し、

$$p(t) = \int_0^t d\tau\,(a/m_0)\cdot\exp(-(a/m_0)\cdot\tau)$$
ここに、$a>0$ は定数 (A8)

としてみよう。この $p(t)$ を具体的に計算すれば、

$$p(t) = \begin{cases} 1-\exp(-(a/m_0)\cdot t) & \cdots 0 \leqq t < \infty \text{ の場合} \\ 0 & \cdots t<0 \text{ の場合} \end{cases} \quad \text{(A9)}$$

となる。この式（A9）における t は探索時間である。

以下の（i),（ii),（iii）が成り立つが、（ii）によれば、平均探索時間 $t=m_0/a$ に探索が終了する確率は、

$$p(t)\,|_{t=m_0/a} = 1-\exp(-(a/m_0)\cdot m_0/a)$$
$$= 1-\exp(-1) \fallingdotseq 0.63 \quad \text{(A10)}$$

であることがわかる。式（A8）内の正定数 a は、経験的に決めなければならないが、この（ii）からも、およその値を決定できる：

（i)（十分長い探索時間の後には、入力パターン φ の帰属するカテゴリを必ず発見できる）

$$\int_0^\infty d\tau\,(a/m_0)\cdot\exp(-(a/m_0)\cdot\tau) = 1.$$

（ii)（入力パターン φ の帰属するカテゴリを探索するのに必要とされる平均探索時間 t は m_0/a である）

$$\int_0^\infty d\tau\ \ \tau\cdot(a/m_0)\cdot\exp(-(a/m_0)\cdot\tau)$$
$$= [1/(a/m_0)] = m_0/a.$$

（iii)（探索時間の分散は、$(m_0/a)^2$ である）

$$\int_0^\infty d\tau\,[\tau - m_0/a]^2\cdot(a/m_0)\cdot\exp(-(a/m_0)\cdot\tau)$$
$$= [1/(a/m_0)]^2 = (m_0/a)^2.$$

□

A2.2 探索持続時間

式（A8）の $p(t)$ を使って定義される1より大きくない非負量

$$Q(t) \equiv 1 - P(t) \quad \text{(A11)}$$

は、探索が少なくとも、t 単位時間持続する確率である。具体的に計算すると、

$$Q(t) = \int_t^\infty d\tau\ (a/m_0)\cdot\exp(-(a/m_0)\cdot\tau) = \begin{cases} \exp(-(a/m_0)\cdot t) & \cdots 0 \leqq t < \infty \text{ の場合} \\ 1 & \cdots t<0 \text{ の場合} \end{cases} \quad \text{(A12)}$$

となり、**指数型持続時間**（exponential holding time）といわれる。

よって、式（A8）の指数型探索終了確率分布 $p(t)$ の下では、探索が t 単位時間持続したという条件の下で、探索が t から t+h の間に終了する確率

$$\mathrm{prob}\{t \leqq \tau < t+h\}/\mathrm{prob}\{t \leqq \tau\}$$
$$= \int_t^{t+h} d\tau\,(a/m_0)\cdot\exp(-(a/m_0)\cdot\tau) \Big/ \int_t^\infty d\tau\ (a/m_0)\cdot\exp(-(a/m_0)\cdot\tau)$$
$$= -[Q(t+h) - Q(t)]/Q(t) \quad \text{(A13)}$$
$$= 1 - Q(t+h)/Q(t) \quad \text{(A14)}$$

,where the notation prob {A} means the probability of event A (A15)

は、式（A8）を代入して、

$$\mathrm{prob}\{t \leqq \tau < t+h\}/\mathrm{prob}\{t \leqq \tau\}$$
$$=1-\exp(-(a/m_0)\cdot h) \qquad \text{(A16)}$$

と表される。

この際、式（A13）から、

$$\lim_{h\to 0} \mathrm{prob}\{t \leqq \tau < t+h\}/\mathrm{prob}\{t \leqq \tau\}/h$$
$$=-[dQ(t)/dt]/Q(t) \qquad \text{(A17)}$$
$$=-(a/m_0) \ \text{if} \ t>0 \qquad \text{(A18)}$$

が成立している事実に注意しておこう。

A3. 逐次認識過程の分析

第1章の ①カテゴリ候補の逐次的絞り込み方法 について、分析してみよう。

以下では、任意の入力パターンφをとるが、固定されているので、$f(t;\varphi)$ を簡単に、$f(t)$と書く。

A3.1 カテゴリ$\mathfrak{C}_j$の寿命の総和

任意の1つのカテゴリ候補が残存する持続時間（holding time）の分布が、式（A2）の指数型分布 $f(t)$であるとしよう。ここに、

$$\lambda=a/m_0 \qquad \text{(A19)}$$

とおく。

$t(j)$：第$j \in J$番目のカテゴリ$\mathfrak{C}_j$の寿命（持続時間）

を用意し、その総和

$$\sum_{j=1}^{n} t(k_j) \text{ ここに、} k_j \geqq 0 \qquad \text{(A20)}$$

の密度を、

$$f_n(t) \qquad \text{(A21)}$$

とすれば、先ず、

$$f_1(t)=f(t) \qquad \text{(A22)}$$

であることがわかる。

k_1番目のカテゴリ候補が実際にパターンφの帰属するカテゴリであるかどうかを時間$t(k_1)$をかけて確かめ、そうでなければ、次に、k2番目のカテゴリ候補が実際にパターンφの帰属するカテゴリであるかどうかを時間$t(k_2)$をかけて確かめ、…という **“並列でない逐次的認識過程”** では、

不等式

$$\sum_{j=1}^{u+1} t(k_j) \geqq t \qquad \text{(A23)}$$

を満たす時刻tでは、u個のカテゴリ候補が除外されて、m_0-u個のカテゴリ候補が残存しており、$n(t)$を時刻tにおいて候補外に置かれるカテゴリ候補の総数の意として、

$$n(t) \leqq u \qquad \text{(A24)}$$

が成り立っている

ということになる。言い替えれば、

$$\mathrm{prob}\{n(t) \leqq u\}=\mathrm{prob}\{\sum_{j=1}^{u+1} t(k_j) \geqq t\} \qquad \text{(A25)}$$

が成り立っている。

A3.2　複数のカテゴリ候補が除外される確率の密度

以後、2つの確率変数

$$t(k_i),\ t(k_j)\quad (i \neq j) \tag{A26}$$

は確率的に独立であるとしよう。

p_1, p_2 は2連続確率変数 X,Y, Z の確率密度であるとすれば、X, Yが独立であれば、その和 $Z=X+Y$ の分布の確率密度 $q(z)$ は、p_1, p_2 の畳み込み（convolution）で与えられ、

$$q(z)=\int_{-\infty}^{+\infty} dx\, p_2(z-x)\cdot p_1(x) \tag{A27}$$

であることに、注意しておく。

よって、具体的に計算すれば、

$$f_2(t)$$

$$=\int_{-\infty}^{+\infty} dx\, f(t-x)\cdot f(x) \tag{A28}$$

$$=\int_0^t dx\, f(t-x)\cdot f(x) \tag{A29}$$

$$=\lambda^2\cdot t\cdot \exp(-\lambda t) \tag{A30}$$

$$f_3(t)$$

$$=\int_0^t dx\, f_2(t-x)\cdot f(x)$$

$$=(\lambda^3\cdot t^2/2)\exp(-\lambda t) \tag{A31}$$

…

$$f_{n+1}(t)$$

$$=\int_0^t dx\, f_n(t-x)\cdot f(x)$$

$$=(\lambda^{n+1}\cdot t^n/n!)\exp(-\lambda t) \tag{A32}$$

$$n=0, 1, 2, \cdots$$

が得られる。式（A32）を用いれば、式（A25）の確率について、

$$\mathrm{prob}\{n(t)\leqq u\}$$

$$=\mathrm{prob}\{\sum_{j=1}^{u+1} t(k_j)\geqq t\}\qquad \because\ 式（A25）$$

$$=\int_t^{\infty} d\tau\ f_{u+1}(\tau)$$

$$=\int_t^{\infty} d\tau\ (\lambda^{u+1}\,\tau^u/u!)\exp(-\lambda\tau)\qquad \because\ 式（A32）$$

$$=\int_t^{\infty} d\tau\ (\lambda^{u+1}/u!)\cdot\tau^u\cdot(d/d\tau)\cdot\{\exp(-\lambda\tau)/(-\lambda)\}$$

$$=(\lambda^u\cdot t^u/u!)\exp(-\lambda t)+\int_t^{\infty} d\tau\, f_u(\tau)$$

$$=(\lambda^u\cdot t^u/u!)\exp(-\lambda t)+\mathrm{prob}\{\sum_{j=1}^{u} t(k_j)\geqq t\}$$

$$=(\lambda^u\cdot t^u/u!)\exp(-\lambda t)+\mathrm{prob}\{n(t)\leqq u-1\}\qquad \because\ 式（A25） \tag{A33}$$

を得て、従って、

$u\geqq 1$ のとき、

$$\mathrm{prob}\{n(t)=u\}$$

$$=\mathrm{prob}\{n(t)\leqq u\}-\mathrm{prob}\{n(t)\leqq u-1\}\qquad \because\ 式（A33）$$

$$=\{(\lambda\cdot t)^u/u!\}\cdot\exp(-\lambda t) \tag{A34}$$

が成り立つ。一方、

$$\mathrm{prob}\{n(t)=0\}$$

$$=\mathrm{prob}\{t(k_1)\geqq t\}$$

$$=\int_t^{\infty} d\tau\ f(\tau)$$

$=\exp(-\lambda t)$　　∵　2式(A2), (A22)　　(A35)

であるから、n(t)(=0,1,2,⋯)　の分布は u=0 を含めて、式 (A34) である。結局、注目した1つのカテゴリが正しいカテゴリであるかどうかを、各認識段階で確かめ、正しいカテゴリでなかったら、このカテゴリを捨て、次の認識段階へ進む逐次的認識過程方式（第1章の ①カテゴリ候補の逐次的絞り込み方法）での、**時刻 t において候補外に置かれるカテゴリ候補の総数 n(t)∈{0,1,2,⋯} の分布は、平均値、分散が共に、λt であるPoisson分布である。**

付録B．非カテゴリ候補の発見確率 z(t), 信頼度R(t), 平均寿命 θ, アベイラビリティA(t)

本付録Bでは、確率過程としての、カテゴリ候補の絞り込みの働きとしてのパターン認識過程に関し、

発見された1つのシステム（例えば、ソフトウェア）故障
=発見された1つの非カテゴリ候補　　(B1)

と見立て、システム（例えば、ソフトウェア）信頼度の理論を適用する基礎が論じられる。

先ず、時刻 t 迄にシステムが故障しない確率として定義される信頼度 R(t)がシステム内の故障を発見できる確率密度 z(t)で表されることを指摘し、その後、**システムが故障する確率密度**（処理の対象とする問題のパターン φ に関し、φ が帰属しないと思われるその非カテゴリ候補が発見される確率密度）**が指数形分布である場合**、平均寿命（非カテゴリ候補が発見される平均時間)、システムが特定の時点で要求仕様通りの機能を維持している度合いとしてのアベイラビリティ（system availability; 認識システムが特定の時刻にカテゴリ候補の絞り込み機能を維持している度合い）などを介して、パターン認識の働きを分析する。

B1．発見率 z(t) による信頼度 R(t) の表現

本研究では、**非定常ポアッソン過程**としてのパターン認識過程（Pattern-recognizing process as a nonhomogeneous poisson process）が説明されている。この場合、パターン認識の働きとは、確率過程としての、**カテゴリ候補の絞り込み**と見ている。

任意に固定したパターン $\varphi\in\Phi$ は、認識の時刻 t に、

カテゴリ集合 $\{\mathfrak{C}_j \mid j\in\gamma_t\}$ のいずれか1つに帰属している　　(B2)

と判明しているものとする。次の時刻 t+(△t) には、パターン $\varphi\in\Phi$ は、

カテゴリ集合 $\{\mathfrak{C}_j \mid j\in\gamma_{t+(\triangle t)}\}$ のいずれか1つに帰属している　　(B3)

と決定されたとする。このとき、

$\mathfrak{C}_j,\ j\in\gamma_{t+(\triangle t)}$　　(B4)

は、パターン $\varphi\in\Phi$ の、認識の時刻 t におけるカテゴリ候補であったということになっている。また、時刻 t+(△t) において除去されたカテゴリ集合

$\{\mathfrak{C}_j \mid j\in\gamma_{t+(\triangle t)}-\gamma_t\}$　　(B5)

は、パターン $\varphi\in\Phi$ の、認識の時刻 t における**非カテゴリ候補**であったと、称することにしよう。

時刻 t において、

$f(t)\cdot(\triangle t)\ (=f(t;\varphi)\cdot(\triangle t))$　　(B6)

は、処理対象としている問題のパターン $\varphi\in\Phi$ について非カテゴリ候補が発見される確率であると

しよう。例えば、**指数形分布**を選ぶものとすれば、$\lambda>0$ を正定数として、

$$f(t)=\lambda\cdot\exp(-\lambda t)\quad(t\geqq 0,\lambda>0) \tag{B7}$$

となる。当然ながら、

$$[\forall t\in[0,\infty]\equiv\{s\,|\,0\leqq s<\infty\}, f(t)\geqq 0] \tag{B8}$$

$$\wedge\int_0^\infty dt\, f(t)=1 \tag{A9}$$

が成立している。

以後、式（B6）での$\triangle t$を無限小 dt にとろう。

ここで、非カテゴリ候補をシステムの故障に対応させ、システムの信頼性技術[13]を適用しよう。

時刻 $t\geqq 0$ 迄に**非カテゴリ候補を発見できる確率 F(t)**は、

$$F(t)=\int_0^t d\tau\ f(\tau) \tag{B10}$$

と表され、時刻 $t\geqq 0$ 迄に**非カテゴリ候補を発見できない確率 R(t)**は、

$$R(t)=1-F(t)=\int_t^\infty d\tau\ f(\tau) \tag{B11}$$

と表現される。ここに、2式(B10),（B11）から、

$$\forall t\in[0,\infty],\ (d/dt)F(t)=f(t) \tag{B12}$$

$$\wedge-(d/dt)R(t)=f(t) \tag{B13}$$

が成立していることに注意しておこう。

式（B11）の R（t）は、システムの信頼性技術においては、

時刻 t 迄にシステムが故障しない確率

であって、いわゆる**信頼度**（reliability）と呼ばれている。

今、

$$(t)\equiv f(t)/\int_t^\infty d\tau\ f(\tau)=f(t)/R(t) \tag{B14}$$

という量 **z(t)**を考えると、

時刻 $t(\geqq 0)$ 迄に非カテゴリ候補を発見できなかったが、その直後の dt 時間内に非カテゴリ候補を発見できる“**発見率**としての確率密度（条件付き非カテゴリ候補の発見確率密度）” (B15)

を表しており、式(B11)の R(t)を用いて、式(B13)から、z(t)は、

$$z(t)=-[R(t)]^{-1}\cdot(d/dt)R(t) \tag{B16}$$

と再表現されることがわかる。

［定理A1］**（発見率 z(t)による信頼度 R(t)の表現定理）**

式(B16)の z(t)を用いて、3式(B11),（B10),（B6）の R(t), F(t), f(t)は、

(i) $R(t)=\exp(-\int_0^t du\, z(u))$

(ii) $F(t)=1-R(t)=1-\exp(-\int_0^t du\, z(u))$

(iii) $f(t)=(d/dt)F(t)$

$=z(t)\cdot R(t)=z(t)\cdot\exp(-\int_0^t du\, z(u))$

と表現できる。

（証明）(i)の証明：

$z(u)\cdot du=-[R(u)]^{-1}\cdot dR(u)$　　∵　式（B16）

$=-d\log_e R(u)$

$$\therefore\quad -z(t)\cdot du=d\log_e R(u) \tag{B17}$$

を得て、両辺を u=c から u=t 迄積分すれば、

$$-\int_c^t z(u)\cdot du = \log_e R(t) - \log_e R(c) \quad \text{(B18)}$$

が成り立つことがわかる。

ここで、

$c=0$ ととれば、$R(0)=1$ $\because$ 式（B11）

$$\therefore \quad \log_e R(0)=0 \quad \text{(B19)}$$

であるから、式 (B19) を式 (B18) に代入して、

$$-\int_c^t z(u)\cdot du = \log_e R(t)$$

を得て、結局、(i) が証明された。

(ii)の証明：(i)を式(B11) に代入したものである。

(iii)の証明：

$f(t)=(d/dt)F(t)$ $\because$ 式（B12）

$=z(t)\cdot R(t)$ $\because$ 式（B14）

$=z(t)\cdot \exp\left(-\int_0^t du\, z(u)\right)$ $\because$ (i) □

ここで、式 (B6) の意味するところより、

$$\theta \equiv \int_0^\infty d\tau\ f(\tau)\cdot\tau$$

$$= -\int_0^\infty dR(\tau)\cdot\tau \quad \because \text{式 (B13)} \quad \text{(B20)}$$

は、非カテゴリ候補が発見される平均時間（非カテゴリ候補の平均寿命）を表している。

B2. 指数形分布での信頼度 R(t)の表現

式（B7）の指数形分布 $f(t)$ の場合、

(イ)式(B20)の「平均寿命」θ

と、定理B1での3つの量である

(ロ)「式(B10)の非カテゴリ候補の発見確率」 $F(t)$

(ハ)「式(B11)の非カテゴリ候補の発見不能確率としての 信頼度」 $R(t)$

(ニ)「式(B14)の条件付き非カテゴリ候補の発見確率密度」 $z(t)$

とを計算してみよう。

先ず、

$$\theta \equiv \int_0^\infty d\tau\ f(\tau)\cdot\tau$$

$$= \int_0^\infty d\tau\ \lambda\cdot\exp(-\lambda\tau)\cdot\tau = 1/\lambda \quad \text{(B21)}$$

$$\therefore \quad \lambda = 1/\theta \quad \text{(B22)}$$

であることがわかる。以後、

$$f(t) = (1/\theta)\cdot\exp(-t/\theta) \quad \text{(B23)}$$

とおく。次に、

$$F(t) = \int_0^t d\tau\ f(\tau)$$

$$= \int_0^t d\tau\ (1/\theta)\cdot\exp(-\tau/\theta)$$

$$= 1-\exp(-t/\theta) \quad \text{(B24)}$$

$$R(t) = 1-F(t) = \exp(-t/\theta) \quad \text{(B25)}$$

$$z(t) = f(t)/R(t)$$

$$= (1/\theta)\cdot\exp(-t/\theta)/\exp(-t/\theta)$$

$$= 1/\theta = \lambda \quad \text{(B26)}$$

と、求められる。

尚、時刻 t が平均寿命 θ になる時点で、

$$R(\theta)=\exp(-1)\fallingdotseq 0.368 \quad \text{(B27)}$$

$$F(\theta)=1-R(\theta)\fallingdotseq 0.632 \quad \text{(B28)}$$

に注意しておこう。

本研究内容は、以下の①,②,③の分析を今少し、精密に構成し直したものである：

①カテゴリ総数が m_0 であり、m_0-1 個の非カテゴリ候補が存在する場合を考えよう。

平均寿命 $t=\theta$ の時点で、発見できる非カテゴリ候補の総数は、

$$(m_0-1)\cdot F(\theta)=(m_0-1)\times 0.632 \quad \because \text{ 式 (A28)} \quad \text{(B29)}$$

である。例えば、

$m_0=100$ の場合、

$$(m_0-1)\cdot F(\theta)=99\times 0.632\fallingdotseq 62.528 \quad \text{(B30)}$$

である。

②また、或る不動点認識過程においては、$t=7000$ [時間単位] にわたって観測したところ、その間に10個の非カテゴリ候補が発見されたとしよう。このとき、平均寿命 θ は、

$$\theta=7000/10[\text{時間単位}]=700[\text{時間単位}] \quad \text{(B31)}$$

になり、

時刻 $t=\theta$ 迄に非カテゴリ候補を発見できなかったが、その直後の dt 時間内に非カテゴリ候補を少なくとも1個発見できる発見率 $z(\theta)$ は、

$$z(\theta)=1/\theta=1/700\fallingdotseq 0.143 \quad \because \text{ 式 (A26)} \quad \text{(B32)}$$

である。

③認識を開始して、1000時間単位後において非カテゴリ候補を発見できる確率 $F(1000)$ は、

$$F(1000)=1-\exp(-1000/700)\fallingdotseq 1-\exp(-1.429)$$

$$\fallingdotseq 1-0.239=0.761 \quad \text{(B33)}$$

である。

B3. 平均寿命θと認識の終了時刻tとの間の関係

非カテゴリ候補の発見が、その確率密度 $f(t)$ が式(B23)の指数形分布に従う場合、

ある時刻 $t\geqq 0$ 迄に非カテゴリ候補を発見できる確率 $F(t)$ は、式(B24)で与えられる。カテゴリ総数が m_0 であり、m_0-1 個の非カテゴリ候補が存在する場合、

式(B21)の θ は、m_0-1 個の非カテゴリ候補が

発見される平均時間（非カテゴリ候補の平均寿命）

である。

時刻 t 迄に発見された非カテゴリ候補の個数が c 以下である確率 $Bt(c)$ は、2項分布（binominal distribution）を使って、

$$Bt(c)=\sum_{r=0}^{c} {}_{m_0}C_r\cdot F(t)\cdot(1-F(t))^{m_0-r} \quad \text{(B34)}$$

ここに、

$${}_{m_0}C_r\equiv m_0!/[r!\cdot(m_0-r)!] \quad \text{(B35)}$$

となる。

数 s_1 を十分1に近く選定し、かつ、$c=m_0-1$ と選んで、不等式

$B_t(m_0-1) \geqq s1$ (B36)

を解いて、時刻tを求めれば、このtが

認識過程がほぼ、終了する時刻

である。

(i) 終了時刻tが与えられたとき、このtに合致する式(B22)の平均寿命θを求めること

(ii) 式(B22)の平均寿命θが与えられた場合、終了時刻tを予測すること

が必要となる。尚、認識過程においては、

(イ)有望なカテゴリ候補を不合格な非カテゴリ候補として排除する誤り(第1種の誤り)

(ロ)当然、非カテゴリ候補として排除すべきものを残存させ、合格とする誤り(第2種の誤り)

という2種類の誤りがある。

B4. 認識システムの、カテゴリ候補絞り込み過程における保全度と可稼働度

認識のある段階で、あるカテゴリが非カテゴリ候補として取り扱われていたが、いわゆる"修理(repair)"を行って、このカテゴリをカテゴリ候補に復帰させる働きは、これまでの不動点探索形認識[54]では取り入れられていないが、この場合を想定して見よう。

認識システムが或る入力パターンに関し、カテゴリ候補が不足していたため、その不足のカテゴリ候補を復帰させる一連の動作を保全(maintenance)と呼ぶことにする。保全度(maintenability)とは、

非カテゴリ候補に対し保全が実行されるとき、規定の時間内に保全を終了する確率

として定義される。

時間t内に保全を完了する確率としての保全度(時刻t迄に、非カテゴリ候補がカテゴリ候補として修理できる確率)を$M(t)$で表せば、

$M(t)=1-\exp(-\mu t)$ (B37)

と表せるだろう。$1/\mu$は**平均修理時間**(mean time to repair; MTTR)と呼ばれる。この$M(t)$に対し、式(B24)における$F(t)$における、式(B22)の$1/\lambda=\theta$は平均寿命であるが、このθは、信頼性理論[13]における"修理の効がない非カテゴリ候補の平均寿命"としての

MTTF(meantime to failure)

に対応する。

動作状態と、非動作状態とを繰り返すa Markov processを持ったソフトウェアシステムでのinstantaneous software availabilityが求められているけれども[65]、認識システム全体の可稼働状況を表す尺度として、

認識システムが特定の瞬時にカテゴリ候補に関する

絞り込み機能を維持している確率(可稼働度) $A(t)$

を、

瞬時アベイラビリティ(instantaneous availability)

ということにする。

この$A(t)$を、

$A(t)=[\lambda/(\lambda+\mu)]+[\mu/(\lambda+\mu)]\cdot\exp(-(\lambda+\mu)/t)$ (B38)

と与えるのがよい。ここに、

$\lim_{t\to+0}A(t)=\lambda/(\lambda+\mu)$ (定常項) (B39)

$\lim_{t\to+\infty}A(t)$

$$=[\lambda/(\lambda+\mu)]+[\mu/(\lambda+\mu)]$$
$$=1 \qquad \text{(B40)}$$
の成立に注意する。

保全動作は、カテゴリ候補の探索において**back-trackがある認識過程**を考えているが、これまでの不動点探索形認識[54]では、このようなbacktrack動作は考慮の外にある。将来の認識理論に考慮しなければならない。

付録C．まえがきの(イ)以外のソフトウェア信頼度成長モデル

ソフトウェア信頼度成長モデル**SRGM**（software reliability growth model）の主な用途は、開発途中、並びに、開発終了時点における残存フォールト数の見積りであるが［20］、まえがきの（イ）の他に、次の3種類(ロ)，(ハ)，(ニ)のSRGMがある[20]：

時刻 t 迄に検出された期待フォールト数を H(t)

と表す。

（ロ）**指数型SRGM**（exponential SRGM）

単位時間当りのフォールト検出数は残存フォールト数に比例すると解釈したものであり、

$$H(t)=a\ [1-\exp\ (-bt)]. \qquad \text{(C1)}$$

（ハ）**習熟S字型SRGM**（inflection S-shaped SRGM）

フォールトには他のフォールトと関係なく検出可能なものと、他のフォールトが検出されてから検出可能なものがあるとし、障害検出率は検出可能なフォールト数に比例すると解釈したものであり、

$$H(t)=a[1-\exp(-bt)]/[1+c\exp(-bt)] \qquad \text{(C2)}$$

ここに、

$$c=(1-r)/r \qquad \text{(C3)}$$

であり、この H(t)は、r が小の時 S字型を形成し、大きくなると、指数型を形成する。

（ニ）**複合SRGM**

習熟S字型の解釈に加えて、試験初期は独立フォールトのみが検出されるという解釈を採用したものであり、

$$H(t)=a[1-\exp(-bt)]/[1+c\cdot\exp(-b\cdot w(t,d))] \qquad \text{(C4)}$$

ここに、

$$\begin{cases} w(t,d)=t-d \cdots t-d\geqq 0\ \text{のとき} \\ 0 \qquad\qquad\quad \cdots t-d<0\ \text{のとき} \end{cases} \qquad \text{(C5)}$$

□

付録D．最尤推定と赤池情報量基準 AIC

統計モデル構成の第1歩は、先ず、与えられたデータ（第i(＝1～n）番目の試行結果）x_i の組 $\underline{x}$ を或る分布に従う確率変数ベクトル $\underline{X}$ の実現値と見なすことから始まる。

本付録Dでは、統計的なAIC（Akaike Information Criterion; 赤池情報量基準）［19]，［58]，［60］に

ついて最尤推定の観点から解説され、信頼度モデル内の諸パラメータを最小自乗法で推定する場面へ、AIC を適用する手法 [22] が説明される。

採用する自由パラメータの異なる各 AIC の値の差が 1～2 程度以上ならば、真の確率分布を近似する"標本統計分布で表された各モデル"の内に、データ数（$\underline{x}$ 内のデータ x_i の総数）を n として、通常 $2\sqrt{n}$（高々、n/2）個の自由パラメータ数 k の範囲で、良いモデルが存在することを指摘出来ること [19] に注意しておこう。

D1. 平均対数尤度

2つの関数 $g(x)$, $f(x)$ を、

$g(x)$：真の確率分布の確率密度関数

$f(x)$：設定するモデルに対応する確率密度関数

とする。$g(x)$, $f(x)$ の違いを表し、$g(x)$ が $f(x)$ に一致すると零になる非負量としての K-L 情報量 (Kullback-Leibler Information) [32], [61] $I(g;f)$ は、

$$I(g;f) \equiv \int_{-\infty}^{+\infty} dx\, g(x) \cdot \log_e [g(x)/f(x)] \tag{D1}$$

$$= \int_{-\infty}^{+\infty} dx\, g(x) \cdot \log_e g(x) - \int_{-\infty}^{+\infty} dx\, g(x) \cdot \log_e f(x) \geqq 0 \tag{D2}$$

と定義される。式（D2）の $I(g;f)$ の右辺第1項は固定定数であり、従って、

$$\textbf{平均対数尤度}\ \int_{-\infty}^{+\infty} dx\, g(x) \cdot \log_e f(x) \tag{D3}$$

を大きくすれば、K-L 情報量 $I(g;f)$ は小さくなる。 n 個の独立な観測値

$$\{x_1, x_2, \cdots, x_n\} \tag{D4}$$

が得られると、その対数尤度

$$\sum_{i=1}^{n} \log_e f(x_i) \tag{D5}$$

の n 分の 1

$$(1/n) \cdot \sum_{i=1}^{n} \log_e f(x_i) \tag{D6}$$

で、式 (D3) の平均対数尤度が近似される。

よって、式 (D2) の符号に注意すると、

式(D5)の対数尤度が大きいほど、設定したモデルは

真の分布の様相に近い

という結論が得られる。

D2. AIC の定義

最尤法で当てはめられたモデルが複数個存在するとき、その中の1つを選択する基準として、AIC を導入する。

一般に、母集団特性量 θ に対し、統計量 T が

$$E[T]\,(\text{T の期待値}) = \theta \tag{D7}$$

を満たすとき、T を θ の**不偏推定量**（unbiased estimate）という。

以下の式（D9）は、最大対数尤度が同程度のモデルがある時、その中で実際に推定しなければならないパラメータの数が最も少ないものを選ぶべきであることを示しており、"節約の原理"の1つの具体化と言える。

複数個のモデルがある時、各モデルの善し悪しを評価する基準として、最尤モデルの平均対数尤度である式 (D6) の、式 (D4) のデータに関する期待値（**期待平均対数尤度**）を導入する。

期待対数尤度の値が大きいほどそのモデルは良いと言える。モデルの最大対数尤度を期待対数尤度の1つの推定量と考えることができるが、詳しく調べると、**最大対数尤度そのものは期待平均対数尤度の不偏推定量に他ならない**ことがわかる。

一般に、最大対数尤度は、期待平均対数尤度の本当の値に比べて大きく出やすいという偏りを持つ。この傾向はモデル内の自由パラメータの数が大きいほど著しい。これは、最大対数尤度の比較によってモデルを選択すると、自由パラメータの数の大きいモデルほど、選ばれやすいことを示している。

最大対数尤度の期待平均対数尤度に対する偏りの程度と、モデル内の自由パラメータの数との間の関係を調べると、

(モデルの最大対数尤度)－(モデル内の自由パラメータの数)　(D8)

が近似的に、期待平均対数尤度の不偏推定量となることが導かれる。歴史的経緯を考慮して、この式 (D8) を (－2) 倍した量

AIC≡(－2)×(期待平均対数尤度の不偏推定量)

＝(－2)×[(モデルの最大対数尤度)－(モデル内の自由パラータの数)]

＝(－2)×(モデルの最大対数尤度)＋2×(モデル内の自由パラータの数)　(D9)

がモデル選択の基準となる。AIC 内の2項(－2)×(モデルの最大対数尤度), 2×(モデル内の自由パラータの数)は各々、モデルのデータからのズレ、モデルの複雑さを表していると考えることが出来る［59］。

An extension of the maximum likelihood principle is suggested by Akaike for the slightly more general problem of choosing among different models with different numbers of parameters［62］.

AIC を最小とするモデル（最小AIC推定値;MAICE）が最適なモデルと考えられる。

式（D9）の AIC は、

AIC

＝(－2)×(モデルの最大対数尤度)＋2×(モデル内の、推定すべきパラメータの数)　(D10)

とも書ける。

"an alternative approach to the problem of choosing the appropriate dimensionality of a model that will fit a given set of observations"

も研究されている［62］。

D3. モデル推定に最小自乗法を使用する場合の AIC

測定可能な多くの変数の中から信頼性に強く影響する変数の組み合わせを選択し、それらを説明変数とする予測モデル

$$\log_e Y = f(X_1, X_2, \cdots, X_n) + E \quad \text{(D11)}$$

を構築することを考えよう。ここに、

Y：は目的変数

X_k ($k=1 \sim n$)：説明変数

E：変数 Y と独立な（予測値と実測値との間の）残差

である。簡単な場合は、各 a_k を回帰係数として、

$$f(X_1, X_2, \cdots, X_n) = a_0 + \sum_{k=1}^{n} a_k \cdot X_k \tag{D12}$$

である。一般には、目的変数と各説明変数との非線形な関係を表現するためには、階層型ニューラルネット［42］を使用すればよい［22］。このときは、式（D11）の f は、階層型ニューラルネットによって表現される非線形関数である。

モデル内の、推定すべきパラメータの推定

には、最小自乗法を使用する場合には、式（D10）は、具体的には、

$$\begin{aligned}&\mathrm{AIC}\\&=(\text{データ数})\times\log_e(\text{残差平方和})+2\times(\text{モデル内の、推定すべきパラメータの数})\end{aligned} \tag{D13}$$

となる [22]。

式（D13）において、第1項は尤度の1つの推定値の（−2）倍を表し、第2項はその推定値と実際の尤度との差の（−2）倍を表している。第1項の推定値は不偏推定量ではなく、パラメータ数が多いほど、実際の尤度の期待値より大きくなる偏りがある。第2項は、その偏りを補正する役割を持っている。この特徴により、AIC は説明変数の個数の異なるモデル同士やパラメータ数の異なるモデル同士の比較に使用できる [22]。

D4. 推定量の不偏性、有効性、一致性

D4.1　不偏性、有効性、一致性に関する7定理

n回の試みの結果を示す記録を、

$$\underline{x} = (x_1, x_2, \cdots, x_n) \tag{D14}$$

と表す。こうして、毎回の実験の結果を代表する確率変数（random variable）X_i の組 $\underline{X}$ として、n次元の確率ベクトル

$$\underline{X} = (X_1, X_2, \cdots, X_n) \tag{D15}$$

を想定する。

$X_1, X_2, \cdots, X_n$ が独立であること、並びに、同一確率分布に従うことを仮定してよい。実測値の供給源として考えた試みの無数の標識集集合を母集団（population）といい、式（D14）の $\underline{x}$ をこの母集団から抽出（sampling）によって得られた標本（sample）とよぶ。実験回数 n はこの標本の大きさ（size）と呼ばれる。

問題の性質や経験から母集団に想定された分布、例えば、多項分布、正規分布などを母集団分布と呼ぶ。抽出値を代表する各変数 X_j は母集団分布に従い、互いに独立である。一般に、母集団分布を定める助変数、例えば、平均、分散などを母集団特性量（母集団パラメータ、母数）と呼ぶ。また、標本値の関数

$$T = T(\underline{X}) = T((X_1, X_2, \cdots, X_n) \tag{D16}$$

を統計量（statistic）と呼ぶ。統計量は確率変数であり、その分布は母集団分布によって定まる。

母集団分布 $f(x, \theta)$ の特性量 θ の具体的な値を統計量によって推定する手法を、点推定（point estimate）と呼ぶ。点推定を行う場合、どのような特性量を採用すべきかが問題になる。

その採用（判定）の基準になる性質として、

（i）不偏性(unbiasedness)

（ii）有効性(efficiency)

（iii）一致性(consistency)

がある。この内、不偏性については、式（D7）を導入し、既に説明されている。確率変数 X の平均値を E（X）と表すことにすれば、確率変数 X の分散（variance）

$$\sigma^2 \equiv \sigma^2(X) \equiv \mathrm{Var}(X) \equiv E((X-E(X))^2) \tag{D17}$$

の正の平方根を標準偏差（standard deviation）と呼ぶが、一定の大きさの標本から作った不偏推定量の中で、集中の度合いを計量する分散が出来るだけ小さい者を探すことが重要である。何故ならば、次の定理D1の(ii)が成り立ち、分散が小さいとき推定量 $\theta^\wedge$ が母集団の特性量 θ の近傍に集中する確率が大きいことが知られているからである［17］。

［定理D1］(事象確率の評価定理)

(i) 関数 $h(x) \geqq 0$ に対し、$E(h(x))$ が存在すれば、任意の正実定数 c に対し、不等式

$$\mathrm{prob}\{h(x) \geqq c\}\ (h(x) \geqq c \text{ が成立する事象の確率}) \leqq c^{-1} \cdot E\ (h(x)) \tag{D18}$$

が成り立つ。

(ii) (Bienaymé-Tchebycheffの不等式) 特に、X の平均 $m \equiv E(X)$,分散 σ^2 が存在すれば、

$$\mathrm{prob}\{\,|x-m| \geqq t \cdot \sigma\} \leqq t^{-2}. \tag{D19}$$

□

分散が小さいとき推定量 $\theta^\wedge$ を求めようとするのは、無意味ではない。それは、その最小分散を与える不偏推定量が唯1つ存在することを指摘する次の定理D2が成り立つからである。

［定理D2］(分散の最小な不偏推定量の一意的存在定理)

標本の大きさ n が一定のとき、母集団の特性量 θ の不偏推定量の内、分散の最小なものは、もし存在すれば、1つに限る。 □

今後、x の関数はすべて $(-\infty, +\infty)$ において区分的に連続とする。

$$I_1 \equiv \{x \mid -\infty < x < +\infty\},\ I_2 \equiv \{\theta \mid \theta_1 < \theta_0 < \theta_2\} \tag{D20}$$

を定義すると、$x \in I_1, \theta \in I_2$ で定義された“θ を助変数とする x の母集団確率密度関数 $f(x, \theta)$”が与えられるとする。大きさ n の式（D14）の標本 $\underline{x} = (x_1, x_2, \cdots, x_n)$ を得たとすれば、一般に標本値がこの標本値 $\underline{x}$ を中心とする近傍の値を実現する確率は、

$$L(\underline{x}, \theta)\, dx_1 dx_2 \cdots dx_n$$
$$\equiv f(x_1, \theta) \cdot f(x_2, \theta) \cdots\cdots f(x_n, \theta)\, dx_1 dx_2 \cdots dxn \tag{D21}$$

によって与えられる。この式（D21）の形を確率要素といい、係数

$$L(\underline{x}, \theta) \equiv f(x_1, \theta) \cdot f(x_2, \theta) \cdots\cdots f(x_n, \theta) \tag{D22}$$

を**尤度関数** (likelihood function) と呼ぶ。

母数 θ を含まない統計量

$$\theta^\wedge = \theta^\wedge(x_1, x_2, \cdots, x_n) \tag{D23}$$

に対して、次の4条件（イ）～（ニ）を仮定する：

（イ）$\partial\theta/\partial x_i\ (1 \leqq i \leqq n)$ は存在して、$\underline{x}$ の区分的連続関数である。

（ロ）$\theta^\wedge$ の他に、n－1 個の関数

$$\xi_1 = \xi_1(\underline{x}),\ \xi_2 = \xi_2(\underline{x}), \cdots, \xi_{n-1} = \xi_{n-1}(\underline{x}) \tag{D24}$$

が選べて、$\xi_1, \xi_2, \cdots, \xi_{n-1}$ は $x_1, x_2, \cdots, x_n$ につき区分的連続な第1次導関数を持つように出来る。

（ハ）$(x_1, x_2, \cdots, x_n)$ と

$$(\theta^\wedge, \xi_1, \xi_2, \cdots, \xi_{n-1}) \tag{D25}$$

との対応は1対1になっている。従って、$\theta^\wedge$ の密度関数

$$g(\theta^\wedge; \theta) \tag{D26}$$

と、$\theta^{\wedge}$で条件付けた（$\xi_1, \xi_2, \cdots, \xi_{n-1}$）の条件付き密度関数

$$h(\xi_1, \xi_2, \cdots, \xi_{n-1} / \theta^{\wedge}; \theta) \tag{D27}$$

は共に、θに関係しているが、2つの表現

$$L(\underline{x}, \theta) \cdot \partial(x_1, x_2, \cdots, x_n) / \partial(\theta^{\wedge}, \xi_1, \xi_2, \cdots, \xi_{n-1}) = g(\theta^{\wedge}; \theta) \cdot h(\xi_1, \xi_2, \cdots, \xi_{n-1} / \theta^{\wedge}; \theta) \tag{D28}$$

$$f(x_1, \theta) \cdot f(x_2, \theta) \cdot \cdots \cdot f(x_n, \theta)\, dx_1 dx_2 \cdots dx_n = g(\theta^{\wedge}; \theta) \cdot h(\xi_1, \xi_2, \cdots, \xi_{n-1} / \theta^{\wedge}; \theta)\ d\theta^{\wedge}\, d\xi_1 d\xi_2 \cdots d\xi_{n-1} \text{（確率要素に対する不変式）} \tag{D29}$$

が成り立つ。

（ニ）3つの関数

$$f(x, \theta), g(\theta^{\wedge}; \theta), h(\xi_1, \xi_2, \cdots, \xi_{n-1} / \theta^{\wedge}; \theta) \tag{D30}$$

をθで微分すれば、$x, \theta^{\wedge}, \xi_1, \xi_2, \cdots, \xi_{n-1}$に関して、一様微分可能性の条件を満たし、不等式

$$|\partial f(x, \theta) / \partial \theta| \leqq F(x) \wedge |\partial g(\theta^{\wedge}; \theta) / \partial \theta| \leqq G(\theta^{\wedge}) \wedge |\partial h(\xi_1, \xi_2, \cdots, \xi_{n-1} / \theta^{\wedge}; \theta) / \partial \theta| \leqq H(\xi_1, \xi_2, \cdots, \xi_{n-1} / \theta^{\wedge}) \tag{D31}$$

を満足する3つの関数F,G,Hが存在する。但し、F,G,Hは、条件

$$\int_{-\infty}^{+\infty} dx\, F(x) < \infty \wedge \int_{-\infty}^{+\infty} d\theta^{\wedge} |\theta^{\wedge}|\, G(\theta^{\wedge}) < \infty \wedge \int_{-\infty}^{+\infty} d\theta^{\wedge} G(\theta^{\wedge}) \cdot \int_{-\infty}^{+\infty} d\xi_1 \int_{-\infty}^{+\infty} d\xi_2 \cdots \int_{-\infty}^{+\infty} d\xi_{n-1} H(\xi_1, \xi_2, \cdots, \xi_{n-1}, \theta^{\wedge}) < \infty \tag{D32}$$

を満たすものとする。 □

このとき、不偏推定量$\theta^{\wedge}$の分散の限界に対する下の限界式（D36）を与える次の定理D3が成り立つ。

［定理D3］（有効推定量定理）

上の4条件（イ）～（ニ）の下で考える。

$$m(\theta) \equiv E(\theta^{\wedge}) = \theta + b(\theta) \tag{D33}$$

とおけば、不等式

$$E((\theta^{\wedge} - \theta)^2) \geqq [dm(\theta)/d\theta]^2 / [n \cdot \int_{-\infty}^{+\infty} dx\, f(x, \theta) \cdot (\partial \log_e f(x, \theta) / \partial \theta)^2] \tag{D34}$$

が成り立つ。式（D34）の等号は、$\theta \in I2$ に対して、次の2条件（i），（ii）が満たされるとき、然もその時に限って成り立つ：$g(\theta^{\wedge}; \theta) > 0$ならば、

（i）$h(\xi_1, \xi_2, \cdots, \xi_{n-1} / \theta^{\wedge}; \theta)$は$\theta$に無関係

（ii）$\theta^{\wedge}$に無関係であるが、一般にθには関係する定数cが存在して、

$$\partial g(\theta^{\wedge}; \theta) / \partial \theta = c \cdot (\theta^{\wedge} - \theta). \tag{D35}$$

特に、$\theta^{\wedge}$がθの不偏推定量ならば、

$$Var(\theta^{\wedge}) \geqq 1 / [n \cdot \int_{-\infty}^{+\infty} dx\, f(x, \theta) \cdot (\partial \log_e f(x, \theta) / \partial \theta)^2] = n^{-1} \cdot [E((\partial \log_e f(x, \theta) / \partial \theta)^2)]^{-1} \tag{D36}$$

□

定理D3の（i），（ii）が成り立つとき、つまり、式（D36）の等号が成り立つとき、$\theta^{\wedge}$は有効

(efficient) 推定量であるという。有効推定量は式 (D36) の右辺で示される同一の分散を持つのであるから、有効推定量は定理D2によって、唯1つに限ることになる。

さて、多くの例について、母集団特性量 θ の推定量 $\theta^{\wedge}$ の分散 $\mathrm{Var}(\theta^{\wedge})$ は、σ を標本の大きさ n に無関係な定数として、

$n \to \infty$ のとき、$\mathrm{Var}(\theta^{\wedge})$ は σ^2/n で近似される (D37)

という漸近性を持つ。また、分散 $\mathrm{Var}(\theta^{\wedge})$ が存在しないような推定量 $\theta^{\wedge}$ に対しても、$\theta^{\wedge}$ が漸近的に平均値 θ, 分散 σ^2/n の正規分布 $N(\theta, \sigma^2/n)$ に従う場合がある。この場合、$\theta^{\wedge}$ の θ の回りの特性は、分散 σ^2/n の正規分布と同様と考えてよい。

次の定理D5の等式 (D38) を満たす $\theta^{\wedge}$ を一致推定量 (consistent estimate) という。

［定理D4］(一致推定量定理)

任意の正数 δ に対し、

$$\lim_{n\to\infty} \mathrm{prob}\{|\theta^{\wedge} - \theta| > \delta\} = 0 \tag{D38}$$

が成立し、$\theta^{\wedge}$ は θ に確率収束する。 □

更に、定理D3によって、$\theta^{\wedge}$ が有効推定量になるためには、定理D3の2条件 (i), (ii) が成り立つことが必要十分条件であった。定理D3の条件 (i) を満たす $\theta^{\wedge}$ を充足推定量 (sufficient estimate) という。このとき、次の定理D5が成り立つ。

［定理D5］(充足推定量定理)

式(D21)の尤度関数 $L(\underline{x}, \theta)$ が

$$L(\underline{x}, \theta) = g(\theta^{\wedge}, \theta) \cdot k(\underline{x}) \tag{D39}$$

の形に書けることが、$\theta^{\wedge}$ が充足推定量であるための必要十分条件である。ここに、$k(\underline{x})$ は θ に無関係な関数である。 □

点推定の最も重要な方法の1つは、フィッシャーによって導入された最尤法 (method of maximum likelihood) がある。

大きさ n の式 (D14) の標本 $x_1, x_2, \cdots, x_n$ に対して、式 (D22) の尤度関数 $L(\underline{x}, \theta)$ は、標本値 $\underline{x}$ が1点の近傍をとる確率に比例するから、θ の真値は $L(\underline{x}, \theta)$ に大きな値を与える可能性を持つ。そこで、$L(\underline{x}, \theta)$ にする θ を推定量とみなす立場を取ろう。そのような $\theta^{\wedge} = \theta$ は、最尤方程式と呼ばれる

$$\partial \log_e L(\underline{x}, \theta)/\partial\theta$$
$$(= L(\underline{x}, \theta)^{-1} \cdot \partial L(\underline{x}, \theta)/\partial\theta) = 0 \tag{D40}$$

を満たす。方程式 (D40) を満たし、最尤推定量 (maximum likelihood estimate) と呼ばれる最尤解

$$\theta^{\wedge} = \theta^{\wedge}(x_1, x_2, \cdots, x_n) \tag{D41}$$

と、有効推定量、充足推定量との間の関係を指摘している次の定理D6が成り立つ。

［定理D6］(最尤解定理) もしも、母集団の特性量 θ の有効推定量 $\theta^{\wedge}$ が存在すれば、最尤解が1つ唯1つ存在して、それは $\theta^{\wedge}$ と一致する。

もしも、θ の充足推定量 $\theta^{\wedge}$ が存在すれば、どんな最尤解も $\theta^{\wedge}$ の関数である。 □

以上の論は、標本の大きさ n を一定にして最尤解を考えたが、次の定理D7は、$n \to \infty$ のときの最尤解の漸近性 (asymptotic property) について指摘したものである。

［定理D7］(最尤解の漸近性定理)

母集団密度 $f(x, \theta)$ は、次の (i), (ii), (iii) の性質を満たすことを前提とする。

(i) $\log_e f(x, \theta)$ は、任意の $x \in I_1$ を固定したとき、すべての $\theta \in I_2$ に対して、3回微分可能である。

(ii) $f(x, \theta)$は、I_1の有限個の点を除いて、xにつき一様に、θの関数として2回微分可能であり、

$$|\partial f/\partial\theta| \leqq F_1(x) \wedge |\partial^2 f/\partial\theta^2| \leqq F_2(x) \wedge$$
$$|\partial^3 \log_e f/\partial\theta^3| \leqq F_3(x) \tag{D42}$$

を満たすxのみの3関数F1, F2, F3があり、その上にこれらの関数は、積分可能条件

$$\int_{-\infty}^{+\infty} dx\, F_1(x) < \infty \wedge \int_{-\infty}^{+\infty} dx\, F_2(x) < \infty \wedge$$
$$\int_{-\infty}^{+\infty} dx\, F_3(x) \cdot f(x, \theta) < \infty \text{ for any } \theta \in I_2 \tag{D43}$$

を満たすものとする。

(iii) すべての$\theta \in I_2$に対して、

$$0 < \tau^2 = \int_{-\infty}^{+\infty} dx (\partial \log_e f/\partial\theta)^2 \cdot f(x, \theta) < \infty. \tag{D44}$$

このとき、任意の$\varepsilon > 0$に対して、次の2性質 (a), (b) を満たすθの、式 (D23) の推定量 = $\theta^{\wedge} = \theta^{\wedge}(x_1, x_2, \cdots, x_n)$ が存在する。

(a) $n_0(\varepsilon)$が存在して、$n > n_0(\varepsilon)$ならば、$1-\varepsilon$よりも大きい確率で、式 (D23) の$\theta^{\wedge}$は最尤解になっている。

$n \to \infty$ に対して、式 (D23) の$\theta^{\wedge}$は漸近的に正規分布 $N(\theta_0, 1/(n\tau^2))$ に従い、θの漸近的に有効な推定量になっている。そして、

$$(b)\ (1/n) \cdot \log_e L(\underline{x}, \theta_0)$$
$$= (1/n) \cdot \sum_{i=1}^{n} \partial \log_e f(x_i, \theta)/\partial\theta \,|_{\theta=\theta_0}$$
$$= c_0 + c_1 \cdot (\theta - \theta_0) + (1/2) \cdot c_2 \cdot (\theta - \theta_0)^2 \tag{D45}$$

のテイラー展開内の第1項

$$c_0 \equiv (1/n) \cdot \sum_{i=1}^{n} \partial \log_e f(x_i, \theta)/\partial\theta \,|_{\theta=\theta_0} \tag{D46}$$

は平均0, 分散がτ^2に等しい独立な変数の相加平均になっている。ここに、不等式$\theta_1 < \theta_0 < \theta_2$を満たす$\theta_0$は未知の助変数$\theta$の真の値である。 □

以上の定理D1～D7の証明については、文献 [17] にある。

D4.2 最尤推定量に関する結論

前節から言えることは次の通りである（文献 [19], p.41 (3.3節最尤法, E節)）

助変数$\underline{\theta}$の真値が$\underline{\theta^*}$である確率密度関数

$f(x, \underline{\theta^*})$ を持つ統計的母集団を考えよう。この母集団が大きいために、標本$\underline{x} = (x_1, x_2, \cdots, x_n)$を抜き出す過程において母集団の構成は変わらないと考え、その第i (=1～n) 番目の実現値がx_iである確率変数X_iの列$\underline{X} = (X_1, X_2, \cdots, X_n)$の各$X_i$は、互いに独立で、同じ分布$f(x, \underline{\theta})$に従い、各$X_i$の統計分布密度は$f(x, \underline{\theta})$であると考えるのである。

$f(x, \underline{\theta})$に関するある条件の下では、対数尤度を最大とするような助変数を選ぶことによって(最尤法)、近似的には真の分布に近いモデル（最尤モデル）を得ようとして推定された推定量としての、最大対数尤度を与える最尤推定量$\underline{\theta^{\wedge}}$は、大きな実験回数（母集団のサイズ）nに対しては、

$$\text{平均値ベクトル } \underline{\theta^*}, \text{ 分散行列} (1/n) \cdot J^{-1} \text{の正規分布 } N(\underline{\theta^*}, (1/n) \cdot J^{-1}) \tag{D47}$$

に近似的に従う（漸近正規性）。但し、Jは、その第i行第j列の成分が

$$Ex(\partial \log_e f(x, \underline{\theta})/\partial\theta_i$$
$$\cdot \log_e f(x, \underline{\theta})/\partial\theta_j) \,|_{\underline{\theta}=\underline{\theta^*}} \tag{D49}$$

で与えられるFisher行列である。これより、

$n \to \infty$ のとき、

（イ）(**一致性**) $\underline{\theta}^{\wedge}$, は真の値 $\underline{\theta}^{*}$,に収束すること

（ロ）(**漸近不偏性**) $\underline{\theta}^{*}$,の推定量として偏りが無くなること

（ハ）(**漸近有効性**) 不偏推定量の中で最も分散が小さいこと

などが判明する。

以上で、最尤推定量としてパラメータを推定した後、初めて計算可能な AIC を次のD4.2節で導出する準備的考察が得られた。

D5. 最尤推定量を用いた AIC

AIC の導出方法について、文献［19］に従って説明しよう。

D5.1 正規分布、カイ自乗分布への漸近性

モデルは確率分布で表現されるとして、自由パラメータの数が k であるモデルを MODEL(k) と表す。

式 (D14) の標本値 $\underline{x}$ が与えられたときの、MODEL(k) の多次元パラメータ $\underline{\theta}$ の**最尤推定量** $\underline{\theta^{\wedge}{}_{k}}$ は、対数尤度

$$\ell(\underline{\theta}) \equiv \sum_{i=1}^{n} \log_e f(xi, \underline{\theta}) \qquad \text{(D50)}$$

を、$\underline{\theta} \in \underline{\Theta_k}$ の範囲で最大化することによって得られる。$\underline{\Theta_k}$ は多次元パラメータ $\underline{\theta}$ のなす空間である。この最大対数尤度 $\ell(\underline{\theta^{\wedge}{}_{k}})$ は、

$$(\underline{\theta^{\wedge}{}_{k}}) \equiv \max_{\theta \in \Theta_k} \ell(\underline{\theta}) \qquad \text{(D51)}$$

で与えられる。

或る確率変数列 $\{Y_n\}_{n=1,2,\cdots}$ と或る確率変数 Y とについて、任意の $\varepsilon > 0$ に対し、

$$\lim_{n\to\infty} \mathrm{prob}\{|Y_n - Y| < \varepsilon\} = 1 \qquad \text{(D52)}$$

が成り立つとき、Y_n は Y に**確率収束**する (converge in probability) という。

Ex(…)を … の平均値の意として、分布 $f(x, \underline{\theta})$ の平均対数尤度 $\mathrm{Ex}(\log_e f(x, \underline{\theta}))$ は、

$$\mathrm{Ex}\ (\log_e f(x, \underline{\theta})$$
$$\equiv \int dx\, f(x, \underline{\theta^*}) \cdot \log_e f(x, \underline{\theta}) \qquad \text{(D53)}$$

で定義される。この $\mathrm{Ex}(f(x, \underline{\theta})$ のn倍

$$\ell^*(\underline{\theta}) \equiv n \cdot \mathrm{Ex}(\log_e f(x, \underline{\theta})) \qquad \text{(D54)}$$

において、$\underline{\theta}$ として、最尤推定量 $\underline{\theta^{\wedge}{}_{k}}$ を代入して得られる

$$\ell^*(\underline{\theta^{\wedge}{}_{k}}) \equiv n \cdot \mathrm{Ex}(f(x, \underline{\theta^{\wedge}{}_{k}})) \qquad \text{(D55)}$$

を、

$$n \to \infty \text{ のとき、} \underline{\theta^{\wedge}{}_{k}} \to \underline{\theta^*} \text{ (確率収束)} \qquad \text{(D56)}$$

が保証される真のパラメータ $\underline{\theta^*}$ の周辺でテーラー展開すると、$(\cdots)^t$ を … の転置の意として、近似的に、

$$\ell^*(\underline{\theta^{\wedge}{}_{k}})$$
$$\fallingdotseq \ell^*(\underline{\theta^*}) + n \cdot (\underline{\theta^{\wedge}{}_{k}} - \underline{\theta^*}) \cdot \mathrm{Ex}(\partial \log_e f(x, \underline{\theta}) / \partial \underline{\theta})\,|_{\underline{\theta} = \underline{\theta^*}}$$
$$+ (1/2) \cdot n \cdot (\underline{\theta^{\wedge}{}_{k}} - \underline{\theta^*}) \cdot \mathrm{Ex}(\partial^2 \log_e f(x, \underline{\theta}) / \partial \underline{\theta^2})\,|_{\underline{\theta} = \underline{\theta^*}} \cdot (\underline{\theta^{\wedge}{}_{k}} - \underline{\theta^*})^t \qquad \text{(D57)}$$

となるが、

$$\mathrm{Ex}(\partial \log_e f(x, \underline{\theta}) / \partial \underline{\theta})\,|_{\underline{\theta} = \underline{\theta^*}} = 0$$

$\because$ $f(x, \underline{\theta})$ は $\underline{\theta} = \underline{\theta^*}$ で、真の分布となり、

式 (D54) の $\ell^*(\underline{\theta})$ は $\underline{\theta}=\underline{\theta}^*$ で最大 (D58)

であるから、Fisher行列

$J_* \equiv -Ex(\partial^2 \log_e f(x, \underline{\theta})/\partial \underline{\theta}^2)\,|_{\underline{\theta}=\underline{\theta}^*}$ (D59)

を用いて、

$\ell^*(\underline{\theta^\wedge_k})$

$\fallingdotseq \ell^*(\underline{\theta}^*) - (1/2)\cdot\sqrt{n}\cdot(\underline{\theta^\wedge_k}-\underline{\theta}^*)\cdot J_*\cdot\sqrt{n}\cdot(\underline{\theta^\wedge_k}-\underline{\theta}^*)^t$ (D60)

と書ける。ここで、

(i) (最尤推定量の漸近正規性)

$\underline{\theta^\wedge_k}$ が式 (D14) の $\underline{x}$ に基づく最尤推定量であるから、

一般的な条件の下で、$n\to\infty$ のとき、$\sqrt{n}\cdot(\underline{\theta^\wedge_k}-\underline{\theta}^*)$ は平均値ベクトル,

分散行列を各々 $0, J_*^{-1}$ とする正規分布 $N(0, J_*^{-1})$ に漸近する (D61)

注意1：モデルの自由パラメータの数が大きくなると、この (i) は成立しなくる。それで、データ数 ($\underline{x}$ 内のデータ x_i の総数) を n として、データに当てはめる自由パラメータ数 k は、通常 $2\sqrt{n}$ (高々、n/2) 個の範囲に選ぶのがよい [19]。

(ii) (カイ自乗分布)

$n\to\infty$ のとき、$\sqrt{n}\cdot(\underline{\theta^\wedge_k}-\underline{\theta}^*)\cdot J_*\cdot\sqrt{n}\cdot(\underline{\theta^\wedge_k}-\underline{\theta}^*)^t$ は近似的に、

自由度 k のカイ自乗分布に従う (D62)

を考慮すると、

n が十分大きいとき、近似的に、

$\ell^*_n(k) \equiv Ex(\ell^*(\underline{\theta^\wedge_k}))$

$= \ell^*(\underline{\theta}^*) - k/2$ (D63)

が成り立つ。

D5.2 AIC の導出

式 (D54) で定義される $\ell^*(\underline{\theta})$ において、$\underline{\theta}$ の代りに ℓ^* を代入して得られる $\ell^*(\underline{\theta}^*)$ と、k 個の自由パラメータを持つ MODEL(k) の、式 (D51) の最大対数尤度 $\ell(\underline{\theta^\wedge_k})$ との間の関係を調べよう。

式 (D50) の $\ell(\underline{\theta})$ を式 (D51) に登場している最尤推定量 $\underline{\theta^\wedge_k}$ の周辺でテーラー展開すれば、近似式

$\ell(\underline{\theta})$

$\fallingdotseq \ell(\underline{\theta^\wedge_k}) + (\underline{\theta}-\underline{\theta^\wedge_k})\cdot\partial\ell(\underline{\theta})/\partial\underline{\theta}\,|_{\underline{\theta}=\underline{\theta^\wedge_k}}$

$+ (1/2)\cdot(\underline{\theta}-\underline{\theta^\wedge_k})$

$\cdot\partial^2\ell(\underline{\theta})/\partial\underline{\theta}^2)\,|_{\underline{\theta}=\underline{\theta^\wedge_k}}\cdot(\underline{\theta}-\underline{\theta^\wedge_k})^t$ (D64)

が得られる。以後、この展開式 (D64) は厳密に成り立つとして、論を進めよう。

対数の法則：平均 μ, 分散 σ^2 の同一分布に従う確率変数の列

$\{X_1, X_2, \cdots\}$ (D65)

について、

$S_n \equiv \sum_{i=1}^{n} X_i$ (D66)

を定義すれば、

$n\to\infty$ のとき、$(1/n)\cdot S_n \to \mu\,(=E(X))$ (D67)

からわかるように、

$(1/n)\cdot\partial^2\ell(\underline{\theta})/\partial\underline{\theta}^2)\,|_{\underline{\theta}=\underline{\theta}^*}$

$= (1/n) \cdot \sum_{i=1}^{n} \partial^2 \log_e f(x_i, \underline{\theta}) / \partial \underline{\theta}^2) \mid_{\underline{\theta} = \underline{\theta}^*}$

$\to \mathrm{Ex}(\partial^2 \log_e f(x, \underline{\theta}) / \partial \underline{\theta}^2) \mid_{\underline{\theta} = \underline{\theta}^*}$

$= -J_*$　∵　式（D59）　(D68)

が成り立つ。この式（D68）と、式（56）とから、

$(1/n) \cdot \partial^2 \ell(\underline{\theta}) / \partial \underline{\theta}^2) \mid_{\underline{\theta} = \underline{\theta}^{\wedge}{}_k{}^*} \to -J_*$　(D69)

が得られる。この式（D69）と、

$\partial \ell(\underline{\theta}) / \partial \underline{\theta} \mid_{\underline{\theta} = \underline{\theta}^{\wedge}{}_k} = 0$　(D70)

∵　式(D51)が成り立つ、即ち、$\underline{\theta} = \underline{\theta}^{\wedge}{}_k$で、

$\ell(\underline{\theta})$は最大値を取る

とを考慮すると、式(D64)の$\ell(\underline{\theta})$は、

nが十分大きいとき、

$\ell(\theta)$

$\fallingdotseq \ell(\underline{\theta}^{\wedge}{}_k) - (1/2) \cdot \sqrt{n} \cdot (\underline{\theta} = \underline{\theta}^{\wedge}{}_k) \cdot J_* \cdot \sqrt{n} \cdot (\underline{\theta} = \underline{\theta}^{\wedge}{}_k)^t$　(D71)

と表現される。この式（D71）において、$\underline{\theta} = \underline{\theta}^*$とおいて得られる式の、両辺の期待値を取れば、式（D63）を導いたのと同様に考えて、

$\mathrm{Ex}(\ell(\underline{\theta}^*)) = \mathrm{Ex}(\ell(\underline{\theta}^{\wedge}{}_k)) - k/2$　(D72)

が得られる。ところで、各X_iの独立性の仮定から、$\underline{\theta} = \underline{\theta}^*$とおいて得られる式(D54)の$\ell^*(\underline{\theta}^*)$は、

$\ell^*(\underline{\theta}^*) = n \cdot \mathrm{Ex}(\log_e f(x, \underline{\theta}^*))$

$= \mathrm{Ex}(\sum_{i=1}^{n} \log_e f(X_i, \underline{\theta}^*))$

$= \mathrm{Ex}(\ell(\underline{\theta}^*))$　∵　式（D50）　(D73)

となるから、この式(D73)と、式(D72)から、

$\ell^*(\underline{\theta}^*) = \mathrm{Ex}(\ell(\underline{\theta}^{\wedge}{}_k)) - k/2$　(D74)

が得られる。

この式(D74)と式(D63)の$\ell^*{}_n(k)$とから、

$\mathrm{Ex}(\ell(\underline{\theta}^{\wedge}{}_k)) - k$

$= \ell^*(\underline{\theta}^*) + k/2 - k$　∵　式（D74）

$= \ell^*(\underline{\theta}^*) - k/2$

$= \ell^*{}_n(k)$　∵　式（D63）

∴　$\ell^*{}_n(k) = \mathrm{Ex}(\ell(\underline{\theta}^{\wedge}{}_k)) - k$　(D75)

$= \mathrm{Ex}(\ell(\underline{\theta}^{\wedge}{}_k) - k)$　(D76)

なる所要の関係が導かれる。

この式(D75)は、式(D51)の最大対数尤度$\ell(\underline{\theta}^{\wedge}{}_k)$を式(D63)の期待平均対数尤度$\ell^*{}_n(k)$の推定量とみたとき、自由パラメータ数kに等しい偏りを持っていること、並びに、この式(D76)は、この偏りを修正した量

$\ell(\underline{\theta}^{\wedge}{}_k) - k$

=(モデルの最大対数尤度)−(モデル内の自由パラメータの数)　(D77)

が、式（D63）の期待平均対数尤度$\ell^*{}_n(k)$の不偏推定量と成っていることを示している（式(D7)を参照)。

この式（D77）を（−2）倍した量が、自由パラメータの数がkであるMODEL(k)の赤池情報量基準

AIC(k)

$\equiv (-2) \times$ (期待平均対数尤度 $\ell^*_n(k)$ の不偏推定量) (D78)

$= (-2) \times [\ell(\underline{\theta^\wedge_k}) - k]$

$= (-2) \times \ell(\underline{\theta^\wedge_k}) + 2 \times k$ (D79)

$= (-2) \times$ (モデルの最大対数尤度) $+ 2 \times$ (モデル内の自由パラータの数) (D80)

である。

(鈴木昇一, 佐久間拓也, 釈氏孝浩, 前田英明, 下平丕作士, 文教大学・情報学部・情報システム学科, "文教大学・情報学部・情報研究　no.18" 投稿論文, 論文題目　不動点探索形構造受精変換多段階認識の、確率過程論的取り扱い, 投稿年月日　1997年9月9日)