

確 率 モ デ ル の 学 習 方 式 と 診 断 へ の 応 用

中 莖 洋 一 郎 古 関 義 幸 田 中 み ど り

日 本 電 気 (株) C&C シ ス テ ム 研 究 所

本報告では、確率的に発生する事象に対して、その観測結果から確率モデルの推定を行い、問題解決時に応用するための手法を提案する。確率モデルの表現方法として推定木を用い、モデル選択の基準としてMDL基準を用いることにより、適切な確率モデルを帰納的に学習できることを示す。また、本学習方式を診断問題に適用し、過去に発生した故障のデータをもとに学習される経験的な知識を用いることで、効果的な診断が可能となることを示す。

An Inductive Learning Method and Its Application to Diagnostic Systems

Yoichiro Nakakuki, Yoshiyuki Koseki and Midori Tanaka
C&C Systems Research Laboratories
NEC corporation

This paper describes an inductive learning method. It acquires an appropriate probabilistic model as a *Presumption Tree* based on given observation data. In order to select the most appropriate presumption tree, we adopted the MDL criterion. This learning capability enables the efficiency of certain kinds of performance systems, such as diagnostic system, that deal with probabilistic problems. The experimental results show that a model-based diagnostic system performs efficiently by making good use of the learning mechanism.

1 はじめに

本研究は、過去の経験を利用しながら新たに与えられた問題の解決を行う知的システムの実現を目的としているものである。一般に「経験」といった場合、様々な種類・レベルのものが考えられるが、本報告では、特に装置の故障等の物理現象（確率事象）に対する経験を対象とし、過去に観測された事象をもとに帰納学習を行い、その事象の確率モデルを推定し、問題解決に利用する方式について述べる。

帰納学習に関しては、さまざまな研究が行われているが、それらの多くは分類規則の学習を目的としている。確率的な分類問題を扱った研究もあるが[9, 11]、確率モデルそのものを学習し、利用する方式に関する研究に関する報告はない。本報告では、そのような確率モデルの帰納学習の必要性を示し、さらに推定木を用いた学習（確率モデル推定）方式を提案し、故障診断問題への応用を示す。

通常、故障診断の専門家は対象装置の構造や動作に関する知識の他に、各故障原因の起こりやすさ等の経験から得られる知識を用いてできるだけ少ないテストで故障原因を特定することが可能である。このような専門家の診断手順を診断ルールとして直接獲得するエキスパートシステムが提案されているが、多くの症状をカバーできるだけの診断ルールの獲得、知識ベースの管理には非常に多くの工数が必要となる欠点がある。これに対し、対象装置の構造や動作に関する知識を持ったモデルベースの診断では、様々な症状に対応が可能である反面、膨大な計算量を減らさなければならず、「故障確率の高い部分から先にテストする」等の経験的な知識の利用が不可欠である[3, 7, 2]。

de Kleerらは経験的に得られる知識として各故障原因の生起確率に着目する方式を提案した。その確率を用いて診断時のエントロピーを計算することで、各テストを行った時に得られる情報量を推定し、最も良いと思われるテストを選択する方法を提案している[3, 4]。しかし、各故障の生起確率を正確に推定できなければ適切なテストを選択することも不可能である。特に故障が頻繁には

起こらないような場合、十分な数のデータを集めることができないため、全く誤った故障確率を推定してしまい、適切でないテストを選択してしまう恐れがある。このように、与えられた過去の観測データから、故障発生の確率モデルを適切に推定することが必要である[10]。

本報告では、確率モデルを推定木の形で表現し、与えられた観測データを基に最も適切な確率モデル（推定木）を見つけるための評価基準として推定木の記述長を採用する。この記述長最小（MDL）基準[14, 15]は、Rissanenによって提案された原理であり、彼の主張によれば、過去の観測事例を基に「将来起こる事象を最も良く推定可能なモデル」は、その記述長が最小となる。この指標は、分類問題における分類規則の帰納学習等にも利用されているものである[12, 18]。

以下、2章では診断問題における確率モデル学習の必要性について述べ、3章では、確率モデルの推定問題を定義し、推定木によるモデルの表現方法、及びその記述長の計算方法を示す。4章では、本学習方式を故障診断に応用した適応型の故障診断システムについて、その診断方式を示す。

2 診断問題と確率モデルの学習

2.1 診断問題

一般に故障診断システムは、外部から与えられる観測情報によって診断を進め、必要であればさらにテストを繰り返し、故障原因を絞り込んでいく。

また、通常は提示されたテストを実際に行って結果を観測する操作にはコスト（時間等）がかかるため、できるだけ少ないコストで正確に故障箇所を見つけることが要求される。

従来の故障診断システムのアプローチは、大きく2つの方式に分けることができる。ひとつは、専門家の持つ「このような症状がでたらこのテストを行う」とか「この場合はこの部品が壊れている」といった、いわゆる浅い知識をルール等の形で獲得し、適用する方式である。この方式では、

全ての症状に対する診断知識を獲得することは事実上不可能であるため、システムの持つ知識が適用できるような場合には効果的であるが、それ以外の場合には、全く対処できないといった欠点がある。

もうひとつの方式は、装置の構造や動作に関する、いわゆる深い知識を用いた診断方式である[1, 3, 5, 13]。この方式は、装置が本来どのように動作すべきかという論理的な知識をもとに診断を行うために、さまざまな症状に対して適用できるという利点がある反面、経験的な知識が利用できないため、「よく壊れる部品を先にテストする」といったヒューリスティックが適用できず、診断が終るまでに多くのコストが必要とされる。また、装置の規模の増大に伴って計算時間が膨大になるといった欠点がある。

筆者らは両者の長所を兼ね備えた診断エージェント[7, 8, 17, 6]について研究を進めているが、本報告では、推定木による学習によって経験的知識を獲得し、効率の良い問題解決を実現する方式について述べる。

2.2 Minimum Entropy Technique

診断時に、最適なテストを選択するためには、各部品の故障確率を考慮する必要がある。

例えば図2-1に示すように、故障の疑いのある部品が8個（部品1から部品8）ある場合を考える。

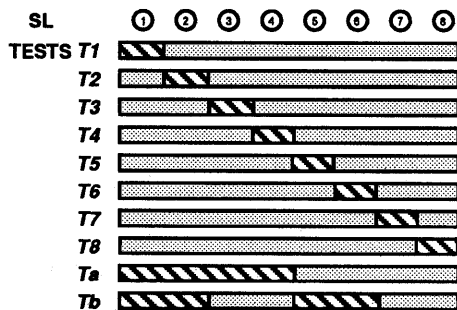


図 2-1 テストの選択

この時に可能なテストは T_1 から T_8 , T_a , T_b の 10 種類あり、各テストは斜線で示した部品が壊れて

いるのか、それ以外の部品が壊れているのかを調べることができるものとする。例えば T_1 は

- (1) 部品 1 が壊れているのか、
- (2) それ以外の部品（部品 2 から部品 8）が壊れているのか

を判定できることを示している。

ここで、もし 8 個の部品の故障確率が全て等しい ($1/8$) 場合には、 T_a , T_b の順にテストを行い、被疑部品の数を 4 個、2 個と減らしていく方式が良いと考えられる。しかし、もし部品 1 の故障確率が 99% である場合には、まず T_1 を行うことにより 99% の場合に 1 回のテストで故障原因を特定することが可能となる。このように、テストの有効性を評価するためには各被疑部品の故障確率を考慮することが重要であることがわかる。

de Kleer らは各部品の故障確率から診断の途中状態におけるエントロピーを定義し、テスト後の状態のエントロピーが最小となるようなテストを選択する方式を提案した[4]。以下、簡単にその方式の概要を述べる。

まず、診断の途中状態においてその被疑部品の持つ故障確率の推定値をもとに、その状態のエントロピーを計算する。ここで、被疑部品の集合を SL (suspect list) を

$$SL = \{S_1, S_2, \dots, S_n\}$$

とし、各被疑部品 $S_1, S_2, \dots, S_n$ の疑わしさ（故障確率）を $p_1, p_2, \dots, p_n$ ($\sum p_i = 1$, $p_i > 0$) とする。このとき、エントロピー $E(SL)$ を

$$E(SL) = - \sum_i p_i \log p_i$$

と定義する。さらに、あるテスト T の結果 $t_1, t_2, \dots, t_k$ が得られた時の SL を SL_i 、それぞれのテスト結果 t_i が得られる確率を q_i としたとき、そのテストの効果 $gain(T)$ を次のように定める。

$$gain(T) = \sum_i q_i (E(SL) - E(SL_i))$$

これは Quinlan の決定木の学習アルゴリズム ID3 [11] で用いられている $gain$ 関数と同じ考え方によるも

のである。各テスト T について $gain(T)$ を求めた時に、最も大きな $gain$ を持つテストが最も有効なテストと考える。

しかし、この方法においては、各部品の故障確率 p_i が正しいことを前提としている。従って、誤った p_i の推定に基づいてテストの選択を行うと、かえって効率の悪いテストを選択してしまう恐れもある。例えば、次のような通信網の診断を考える。

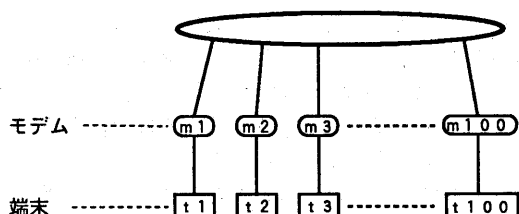


図 3-1 通信網の例

この装置は、ネットワークに 100 個のモデムを通して端末 100 台が接続されている装置である。ここで、過去の故障 10 回の全てがあるモデムの故障（全て異なるモデム）であったとする。ここで、各々の部品に着目 して故障確率の推定を行うと、合計 200 個の部品のうち、過去に壊れた 10 個と、1 回も壊れていない 190 個の部品に分けて、前者の方の故障確率が高いという推定結果が得られてしまう。一方、部品の種類に着目 すると、10 回とも故障はモデムであり、端末やネットワークよりはモデムが故障している可能性が高いと判断することができる。このように着目の仕方によって推定確率に大きな違いが発生するため、故障確率の推定の方法が診断の効率に大きく影響することがわかる。

従って、上記のような故障に関する情報（過去の経験）から適切な故障確率の推定を行う必要がある。その具体的な方法について次章で考察する。

3 確率モデルの学習

ここで、ある装置の故障を例として、その故障確率モデルの学習について考察していく。まず、

その装置が 16 個の部品から構成されているとし、各部品の過去の故障頻度とその部品の種類・新しさに関する各属性値が表 3-1 に示すように与えられた場合に、その情報からどのようなことがいえるのかを考えてみる。

表 3-1 観測データの例

部 品	属 性		故障頻度 (回)
	種 類	新しさ	
1	a	new	1
2	a	old	0
3	b	new	13
4	b	old	9
5	c	new	1
6	c	old	1
7	d	new	0
8	d	old	0
9	e	new	0
10	e	old	0
11	f	new	1
12	f	old	0
13	g	new	0
14	g	old	5
15	h	new	1
16	h	old	0

まず、部品の種類と故障頻度の関係に着目すると、以下に示すような表を作成することができる。

表 3-2 部品種類と故障頻度の関係

種 類	部品数	故 障 頻 度
a	2	1
b	2	22
c	2	2
d	2	0
e	2	0
f	2	1
g	2	5
h	2	1

この表を見ると、種類 b の部品の故障確率が非常に高いことが判る。また、種類 g も比較的故障の頻度が高いと判断するのが自然である。一方、他の種類の部品は過去の故障回数が 0 回から 2 回となっているが、例えば種類 c は種類 a の 2 倍壊れやすいとか、種類 d は種類 a より壊れにくいと判

断してしまうことは危険である。なぜならば、これだけのデータからでは、もともと種類 a, c, d 等の故障確率に大差はなく、その中のあるものが偶然壊れて、それが観測されたということが十分に考えられるからである。

一方、属性として「新しさ」を考えてみると表 3-3 のような結果となる。

表 3-3 部品の場所と故障頻度の関係

新しさ	部品数	故障頻度
new	8	17
old	8	15

この結果からは、新しさが old であるか new であるかによって故障頻度に有意な差があるとは認められない。言い換えればある部品の故障確率を推定する際にその「新しさ」の情報だけではほとんど何もいえないことになる。これに対して前述の「種類」に関しては、その情報によって例えば b ならば比較的その故障確率が高いといった情報が得られる。このように各部品の故障確率を推定する際に有効である属性とそうでない属性がある。また、今の例でも「種類」が g の場合を考えると、表 3-4 から判るようにこの場合には「新しさ」が重要な情報となると考えられる。

表 3-4 部品の種類・場所の組み合わせ

種類	新しさ	故障頻度
g	new	0
g	old	5

このように、物理事象の生起確率の推定を行う際にどの属性（の組み合わせ）に着目するか、また、その属性に関する情報をどのように解釈すべきかを決定する必要がある。これは、与えられた観測事象から最も適切な確率モデルを推定する問題であり、一種の帰納学習であるといえる。以下、その学習方式について述べる。

3.1 準備

ここで、前記のような問題を推定問題として定式化する。まず、排他的かつ網羅的な事象の集合

$X = \{x_1, x_2, \dots, x_m\}$ を考える。つまり、一回の試行で X の中のあるひとつの事象 x_i のみが生起するという状況を考える。さらに $A = \{a_1, a_2, \dots, a_n\}$ を考え、各属性 a_j ($j = 1, 2, \dots, n$) の取り得る値を有限集合 $Dom(a_j)$ とする。表 3-5 に示すように各 x_i には、それぞれ属性 a_j ($j = 1, 2, \dots, n$) に対する属性値 v_{ij} が与えられる。本報告では、どの事象 x_i についても属性値の欠落はなく、必ず各属性 a_j に対する値を持つものとする。

表 3-5 与えられる観測結果

事 象	属 性				頻 度 (回)
	属性 a_1	属性 a_2	...	属性 a_n	
x_1	v_{11}	v_{12}	...	v_{1n}	n_1
x_2	v_{21}	v_{22}	...	v_{2n}	n_2
x_3	v_{31}	v_{32}	...	v_{3n}	n_3
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_m	v_{m1}	v_{m2}	...	v_{mn}	n_m

このような事象の集合 X に対して、各事象 x_i が過去の試行で生起した頻度 n_i が与えられた時に、各事象 x_i の生起確率の推定値 ($\hat{p}_i$) を求める問題を推定問題と呼ぶことにする。

このような推定を行う場合、十分に多くの観測データが得られている場合には、どのような方法を用いてもかなり正確に推定を行うことが可能であると考えられる。しかし、観測データが少ない場合には、偶然発生した事象によるノイズの影響が無視できなくなる。例えば 2 つの事象に対して、発生回数に 1 回でも差があればそれらの事象の発生確率に差があるものと判断する方法が考えられるが、この方法はノイズに弱く、誤った推定をしてしまう恐れがある。一方、逆に観測数に多少の差があってもそれをノイズによるものであるとして、同じ発生確率であると判断してしまう方式も考えられる。しかしこの方法には、少ない観測データからは何も言えないという問題点がある。従って、与えられたデータを基に、ノイズを避けながら有効な情報を抽出することが必要である。

さらに、各事象を個別に見るのではなく、ある属性に注目していくつかの事象のグループを考え、

各グループ毎に発生確率を推定すべき場合も考えられる。例えば、過去に壊れた部品に関して、その設置場所の温度に注目することで、温度の高い場所にある部品の故障確率が温度の低い場所の部品より高いといった傾向が読み取れる可能性がある。

このように、与えられたデータを基に各事象の発生確率に関係のある属性を探し、さらに各事象の確率の分布に関する知識（確率モデル）を学習する必要がある。この問題を解決する方式として、次節においては、確率モデルを表現するために「推定木」を導入し、さらに最適な推定木（確率モデル）を選択するための基準について述べる。

3.2 推定木

本節では、確率モデルを表現する手段として推定木を導入する。下図に示すように、推定木は●で示される分岐点と、○で示される葉とから構成される。

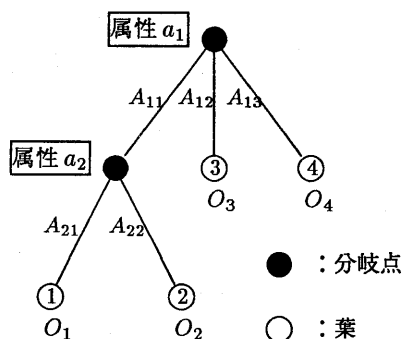


図 3-2 推定木

各分岐点●にはそれぞれある属性 a_i が対応し、（図では属性 a_1, a_2 ）、その子への枝にはそれぞれ属性 a_i の取り得る値の部分集合 $A_{i1}, A_{i2}, \dots, A_{il}$ (l はその分岐点からその子への枝の本数) が対応する。ここで、各 A_{ij} は次のような条件を満足するものとする。

$$\begin{aligned} A_{ij} \cap A_{ik} &= \phi \quad (j \neq k) \\ \bigcup_j A_{ij} &= \text{Dom}(a_i) \end{aligned}$$

推定木は事象の分類を行うために使用される。例えば上図の例では○で示される4つの葉1, 2, 3, 4があるが、これは事例の集合 X を4つのグループに分けることを示している。ここで、例えば葉1は属性 a_1 の値が A_{11} に属し、かつ属性 a_2 の値が A_{21} に属する事象の集合に対応し、その集合内の任意の要素 x_i, x_j に対してはそれぞれ生起確率を同じものと推定すべきであることを示している。また、各葉 k に対応する事象を X_k とし、 $x_i \in X_k$ なる全ての事象 x_i の観測数 n_i の和を O_k とする。この推定木を用いることで、各事象 $x_i \in X_k$ の生起確率 $\hat{p}_i$ の推定を次のように行うことができる。

$$\hat{p}_i = \frac{1}{|X_k|} \cdot \frac{\sum_{x_i \in X_k} n_i}{\sum_{x_i \in X} n_i}$$

このように、推定木によって確率モデルを表現することが可能である。次節では、与えられた事例をもとに適切な推定木を得るための基準となる、MDL 基準について述べる。

3.3 MDL 基準

与えられた観測データを基に、適切な確率モデルを獲得するという問題は、いわゆる分類問題における事例からの帰納学習、つまり、「いくつかの属性で記述される事例とそれがどのクラスに属するか」という情報をもとにその分類規則を帰納学習する問題を考える際にも生じてくる問題である。

その場合、どのレベルまで特殊化 / 一般化すれば将来与えられるであろう新しい事例をうまく分類することができるかを判断しなければならないことになる。この問題に対して Quinlan らは分類規則を決定木として表現し、その決定木の記述長をそのモデルの評価基準としている [12]。これは Minimum Description Length criterion, MDL 基準といわれるもので Rissanen [14, 15, 16] によって提案された手法である。

推定問題を解決するために我々は前節において、「推定木」による確率モデルの表現方法を採用した。そこで、与えられたデータに対して考えられ

る推定木の中で、「将来起こる事象の予測に最適な」推定木（確率モデル）を選択するための基準として MDL 基準を用いる。一般に、確率モデルの記述長は、

1. そのモデル自身の記述長と、
2. そのモデルを用いた場合に、与えられたデータを記述するのに必要な記述長

の和として計算される（単位は共に bit）。この考え方に基づいて、推定木の記述長を計算する。つまり、

1. 木の記述自体に必要な情報量と
2. その木の表わすモデルの、与えられたデータに対する対数尤度

との和が推定木の記述長となる。以下に、推定木の記述長の具体的な計算方法・計算例について述べる。

推定問題においては、考えるべき属性の数や各属性の定義域は有限集合であるため、考え得る推定木の数も高々有限個である。しかし、属性数や各属性の定義域が大きくなるにつれて、その組み合わせの数は爆発的に増えることになる。そこで、本報告では推定木の形に以下に述べるような制限を加える。これにより、計算量を減らすことができる。当然ながらこの場合、必ずしも最小の推定木ではなく、極小の推定木が得られる可能性もあるが、殆どの場合に問題ないものと考えられる。

制限 推定木の全ての分岐点において、その対応する属性を a_i 、分岐数を l 、各分岐条件 A_{ij} とするとき、 A_{ij} ($j = 1, 2, \dots, l-1$) は単集合 (singleton set) に限定する。従って、 A_{il} は「その他の値」の集合を示す。つまり、

$$|A_{ij}| = 1 \quad (j = 1, 2, \dots, l-1)$$

$$A_{il} = \text{Dom}(a_i) - \bigcup_{j=1}^{l-1} A_{ij}$$

を満足するものとする。この制約を満たす分岐点の例を図 3-3 に示す。図の分岐点には属性 a_i が対応している。 $\text{Dom}(a_i)$ は a,b,c,d,e の 5 個の要素か

らなる集合であり、 $A_{i1} = \{a\}$ 、 $A_{i2} = \{c\}$ 、 A_{i3} は「その他の要素」の集合、つまり $\{b, d, e\}$ となる。

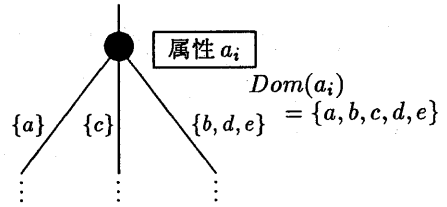


図 3-3 分岐点の例

このような制約を満たす推定木に対して、そのモデル自身の記述長 $L1$ は、次のような式によって計算される（付録参照）。

$$L1 = \sum_{x \in P \cup Q} \log(n - d_x + 1) + \sum_{x \in Q} \frac{1}{2} \log O_x$$

$$+ \sum_{x \in P} \{\log(k_x - 1) + \log \text{cl}(k_x, l_x)\}$$

ここで、 P は全ての分岐点の集合、 Q は全ての葉の集合であり、分岐点 x に対して l_x は分岐数、 d_x は根からの深さ（根を 0 とする）を示す。さらに $k_x = |\text{Dom}(a_i)|$ （ただし、 a_i は分岐点 x に対応する属性）である。

一方、あるモデル（推定木）を用いた時の、与えられたデータに関する記述長 $L2$ は、次の式で計算される。

$$L2 = - \sum_i n_i \log \hat{p}_i$$

ただし、 $\hat{p}_i$ は、そのモデルによって推定される事象 x_i の生起確率である。従って、ある推定木の記述長は、上記の $L1+L2$ によって計算することができる。

3.4 例題

本節では、MDL 基準を用いた推定木選択の例を示す。表 3-1 のようなデータが与えられた場合、何種類もの推定木（確率モデル）が考えられるが、その代表的なものを図 3-4 に示す。

A
○
32

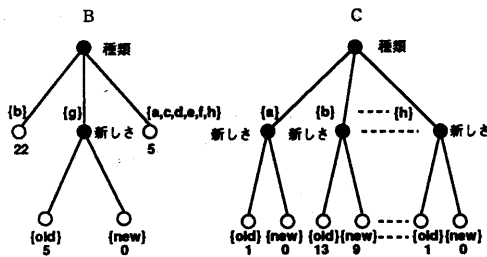


図 3-4 推定木の例

図中に示した3つの推定木の中で、AとCは両極端なモデルとなっている。Aの方は、16個の事象全てをひとつのグループに分類する推定木であるのに対して、Cの方は、各々の事象を異なる16個のグループに分けるものとなっている。当然、モデル自身の記述量はモデルCの方が大きくなるが、与えられたデータの記述量は逆にCのモデルを用いた時の方が小さい。これは、Aはノイズには強いが有効な情報をうまく抽出できないモデルであり、Cの方はその逆であることを示している。

また、Bはそれらの中間のモデルのひとつである。これらのモデルに対して実際に前記の計算式を用いて各々の記述長を計算した結果を表3-6に示す。

表 3-6 各モデルの記述長

モデル	L1	L2	合計 L1+L2 (bits)
A	4.1	128.0	132.1
B	17.8	78.6	96.4
C	29.9	71.8	101.7

この表からわかる通り、Bのモデルの記述長が最も小さい。従って、この3つのモデルに関しては、Bの推定木で表わされるモデルが最も良いモデルであると考えられる。

3.5 実験結果

本方式の有効性を確認するため、図3-1に示すような簡単なネットワークの診断を例題として、

実験を行った。モデムは古いものと新しいものが各50台、端末は全て新しいものとした。故障が起こった時に、それが古いモデムである確率を66%、新しいモデムである確率を33%、端末である確率を1%と仮定した。

まず、上記の分布に基づいて故障をいくつか発生させて学習用のデータセットを作成し、そのデータセットに基づいて学習を行わせた。次に、新たに（前記の分布に基づいて）100個の故障を発生させ、診断に要した平均のテスト回数を測定した。また、比較のために各部品毎にその故障頻度から単純に故障確率を推定し、診断に利用する方法についても同様の実験を行った。学習のために与えたデータの数と平均のテスト回数の関係を図3-5に示す。

平均テスト回数

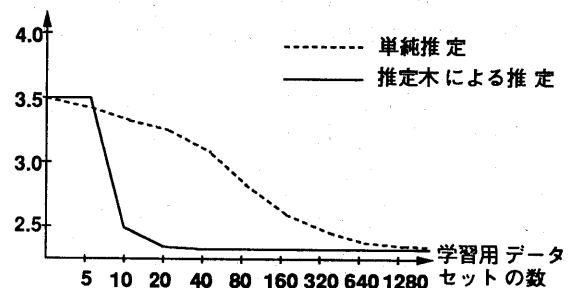


図 3-5 実験結果

単純な推定を行う方式では、経験による学習効果が出るまでに多数の事例が必要であるのに対して、提案する方式を用いて学習を行うことで、より少ない事例を見ただけでも多数の事例が与えられた場合と同等の効果をあげることが可能であることが判明した。

4 適応型故障診断システム

本章では、提案した学習方式を用いて適応型のモデルベース診断を行うための実験システムについて述べる。診断の対象となる装置については、

1. 複数の部品を接続したものとしてモデル化が可能であり、
2. 各部品は、入力・出力ピンを持ち、かつ
3. 部品間の接続記述と各部品の動作記述が与えられる

ことを前提としている。また、故障の発生の仕方として、単一故障及び固定故障を仮定している。

システムは、診断対象装置の動作や構造に関する「設計知識」を用いて、症状や観測情報をもとに被疑部品リスト (SL) の計算を行う。さらに、次に行うテストの選択時には、「設計知識」をもとに生成されるいくつかのテストに対して、「経験的知識」(故障発生の確率モデル)を用いて、その有効性の評価を行い、最も効果的と思われるテストを選択・実行する(図4-1参照)。

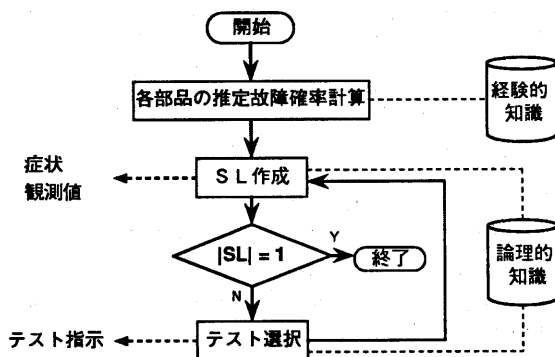


図 4-1 診断の流れ

実験システムは、推論マシン PSI-II 及び並列マシン Multi-PSI 上に、ESP 及び KL1 言語を用いて開発を行っている。図 4-2 にその実行画面例を示す。

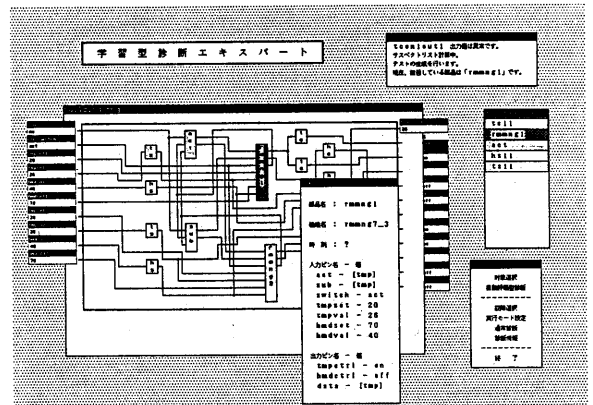


図 4-2 実験システム画面例

5 まとめ

確率的に発生する事象に対して、その観測結果から確率モデルの推定を行い、問題解決へ応用するための手法を提案した。確率モデルの表現方法として推定木を用い、モデル選択の基準として MDL 基準を用いることで、効果的な学習、及びその学習結果を利用した効果的な問題解決が可能となることを示した。

本報告で提案した方法は、診断以外の多くの問題にも適用可能であると考えられるため、今後はさらに幅広い応用分野への適用を考えていく予定である。

謝辞

本研究は、第5世代コンピュータプロジェクトの一環として行われたものである。日頃御世話になっている(財)新世代コンピュータ技術開発機構新田室長に感謝いたします。

参考文献

- [1] Davis, R., "Diagnostic reasoning based on structure and behavior," *Artificial Intelligence*, Vol. 24, pp. 347-410, 1984.
- [2] de kleer, J., K., Mackworth A., and R., Reiter, "Characterizing diagnoses," *Proc. AAAI-90*, Vol. 1, pp. 324-330, 1990.

- [3] de Kleer, J. and Williams, B. C., "Diagnosing multiple faults," *Artificial Intelligence*, Vol. 32, pp. 97-130, 1987.
- [4] de Kleer, J. and Williams, B. C., "Diagnosis with behavioral modes," *Proc. IJCAI-89*, Vol. 2, pp. 1324-1330, 1989.
- [5] Genesereth, M. R., "The use of design descriptions in automated diagnosis," *Artificial Intelligence*, Vol. 24, pp. 411-436, 1984.
- [6] Koseki, Y., Nakakuki, Y., and Tanaka, M., "An adaptive model-Based diagnostic system," *Proc. PRICAI'90*, Vol. 1, pp. 104-109, 1990.
- [7] Koseki, Y., "Experience learning in model-based diagnostic systems," *Proc. IJCAI-89*, Vol. 2, pp. 1356-1361, 1989.
- [8] 古関義幸 「モデルベース診断における経験的知識の学習」 第3回人工知能学会全国大会予稿集 pp. 235-238 1989.
- [9] Mingers, J., "An empirical comparison of pruning methods for decision tree induction," *Machine Learning*, Vol. 4, pp. 227-243, 1989.
- [10] Nakakuki, Y., Koseki, Y., and Tanaka, M., "Inductive learning in probabilistic domain," *Proc. AAAI-90*, Vol. 2, pp. 809-814, 1990.
- [11] Quinlan, J. R., "Induction of decision trees," *Machine Learning*, Vol. 1 (1), pp. 81-106, 1986.
- [12] Quinlan, J. R. and Rivest, R. L., "Inferring decision trees using the minimum description length principle," *Information and Computation*, Vol. 80 (3), pp. 227-248, 1989.
- [13] Reiter, R., "A theory of diagnosis from first principles," *Artificial Intelligence*, Vol. 32, pp. 57-95, 1987.
- [14] Rissanen, J., "Modeling by shortest data description," *Automatica*, Vol. 14, pp. 465-471, 1978.
- [15] Rissanen, J., "A universal prior for integers and estimation by minimum description length," *Ann. of Statist.*, Vol. 11, pp. 416-431, 1983.
- [16] Rissanen, J., "Complexity of strings in the class of Markov sources," *IEEE Trans. on Information Theory*, Vol. 32 (4), pp. 526-531, 1986.
- [17] 田中みどり、古関義幸、中莖洋一郎 「論理的知識と経験的知識を併用した故障原因絞り込み手法」 第39回情報処理学会全国大会予稿集 pp. 243-244 1989.
- [18] Yamanishi, K., "Inductive inference and learning criterion of stochastic classification rules with hierarchical parameter structures," *Proc. SITA-89*, 1989.

付録

推定木による確率モデルの記述長 $L1$ は次の3つの記述長の和として計算される。

1. 推定木の各節点の情報
2. 各分岐枝の情報
3. 推定確率の情報

まず、(1) についての記述長を考える。各節点は分岐点か葉であり、さらに分岐点ではどの属性に対する分岐点であるかという情報が必要となる。さらに分岐点の場合、全体の属性数が n であるから $n - d_x$ 個の属性の可能性がある。従って、これらの情報の記述には $\log(n - d_x + 1)$ の記述長が必要となる。

次に(2)の記述長を計算する。分岐点 x に対応する属性を a_x とすると分岐の数 l_x は2から k_x の間のいずれかであるので、その記述長は $\log(k_x - 1)$ となる。さらに各枝に対応する属性値のとりかたは $c1(k_x, l_x)$ 通りとなる。従って、その記述長は $\log(c1(k_x, l_x))$ となる。ただし、ここで $c1(k_x, l_x)$ は、 $l_x < k_x$ の時には $k_x C_{l_x-1}$ 、それ以外の場合には1である。

最後に(3)の記述長を計算する。各葉 x に対応する頻度を O_x とすると、その記述長は、 $\frac{1}{2} \log O_x$ となる。従って、推定木による確率モデルの記述長 $L1$ は、

$$L1 = \sum_{x \in P \cup Q} \log(n - d_x + 1) + \sum_{x \in Q} \frac{1}{2} \log O_x + \sum_{x \in P} \{\log(k_x - 1) + \log c1(k_x, l_x)\}$$

となる。