

〔非公開〕

TR-C-0118

定性的システム同定法と
データ共有技術との関連について

安部 伸治
Shinji ABE

1 9 9 4 . 4 . 1 4

A T R 通信システム研究所

定性的システム同定法とデータ共有技術との関連について

安部 伸治

ATR 通信システム研究所

1995年4月14日

もくじ	2
-----	---

もくじ

1 はじめに	3
2 定性的システム同定法について	3
3 変数間の因果構造の同定	3
3.1 手法の概要	3
3.2 実験例	5
4 実時間因果構造の同定	7
4.1 手法の概要	7
4.2 状態推定規則とその適用例	7
4.3 制約規則の推定	9
5 データ共有技術について	10
5.1 数量的なデータの扱いについて	10
5.2 従属関係の同定	10
5.3 シンボリックなデータの扱いについて	13
5.4 平坦な構造のデータの扱いについて	14
5.5 データのタイプが混在する場合の扱いについて	14
6 応用：創造的な活動の支援環境	16
7 おわりに	16

1 はじめに

本報告では、定性推論の枠組を利用した数量的な一貫性の発見に関する研究を紹介し、データ共有技術への応用に関する提案を行なう。ここで言うデータ共有技術とは、様々な目的で蓄積されたデータ群をある目的のために一括して利用したり、蓄積されたときとは別の目的でデータを利用して、今着目している何らかの現象を予測するために役立つ技術を目指すものとする。

今、自分が(あるいは着目しているエージェントが)知りたい事柄について、ネットワーク上分散的な複数のデータ(データベース)群を利用して調査を行なう、という状況について考えてみよう。このとき、知りたい事柄がどのデータベースにも明示的に記述されていなかったり、あるいはデータ構造が異なるがために参照できないという状況が比較的頻繁に発生するものと思われる。このような状況に対し、過去に蓄積された沢山のデータ群を利用して今知りたい事柄にどこまで迫れるか、と言った一種の予測技術についての議論は極めて有意義である。

本研究では、数量的な予測に必要なモデル獲得手法のひとつとして現在まで研究を行ってきた、定性推論をベースとしたシステム同定法について紹介し、これを足掛かりとしたデータ共有技術について、その要素研究課題を明らかにする。

2 定性的システム同定法について

定性的システム同定法は、モデルの知られていない物理系に対して、観測結果の大雑把な記述から規則性を見い出して、数学的なモデルを推定する手法である。定性推論は、大雑把な言い方をすると、変量の因果的な構造に関する推論の方法で、物理系を構成する変数の間に因果を伝播させて、系の構造や状態に関する説明を生成したり、挙動を予測したりといった方法が提案されている [de Kleer 84, Forbus 84, Kuipers 86, Nishida 87, Nishida 89, Weld 89]。これらは、系の構造を定性的なモデルとして与えて、説明を生成したり挙動を予測したりといったような過程を実現するが、本研究はこれらの逆問題、つまり、系を観測して得られる挙動を与えることによりもとのモデルを同定する手法に関するものである。

このような課題に関して、現在までに研究を行ってきた2つのテーマについて概略を述べる。1つめは、変数間の因果構造の同定に関するもので、ブラックボックスの入出力の関係から内部の因果的な構造を定式化する手法である。もうひとつは、実時間因果構造の同定に関するもので、力学系を対象とし、実時間変化を含む因果構造を定性的常微分方程式としてモデル化する手法である。

3 変数間の因果構造の同定

3.1 手法の概要

最初の研究である変数間の因果構造の同定に関して述べる。

まず、 n 個の独立変数と1つの従属変数の間の従属関係を仮定する。たとえば、表1のように、 x_1 から x_n までの n 個の条件属性と P という1つの決定属性をもつ決定表を考える。このような決定表を、この図3.1のようなブラックボックスとみなし、入力と出力の間の増減関係を与えることにより、内部構造を推定する。

ここで、出力を入力の詳細な関数であると仮定する。これを式で書くと次のようになる、

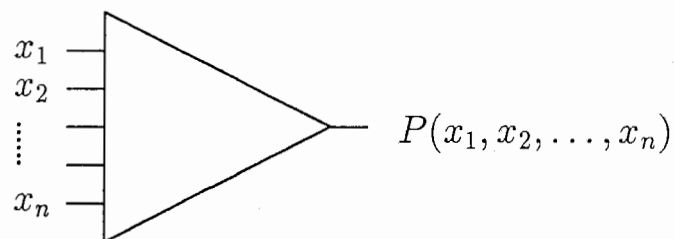
$$dP/dt = \sum_{j=1}^n \{ \partial P / \partial x_j \} \{ dx/dt \} \quad (1)$$

2番目の仮定として、出力は入力に関する次のような有理式でモデル化できるものとする。

$$P(x_1, x_2, \dots, x_n) = \frac{\sum_{K=0}^{\text{order}} a_{k_1, k_2, \dots, k_n} \prod_{i=1}^n x_i^{k_i}}{\sum_{L=0}^{\text{order}} b_{l_1, l_2, \dots, l_n} \prod_{i=1}^n x_i^{l_i}} \quad (2)$$

表 1: n 個の条件属性と 1 つの決定属性の間の従属関係を仮定した決定表

Data	x_1	x_2	x_3	\dots	x_n	P
$data_1$	•	•	•	\dots	•	•
$data_2$	•	•	•	\dots	•	•
$data_3$	•	•	•	\dots	•	•
$data_4$	•	•	•	\dots	•	•
$data_5$	•	•	•	\dots	•	•
\dots						

図 1: n 入力 1 出力を持つブラックボックス

$$\begin{aligned} \text{ここで、} K &= k_1 + k_2 + \cdots + k_n \\ L &= l_1 + l_2 + \cdots + l_n \end{aligned}$$

これにより、観測された入出力の関係を満たすような係数列 a と b を見い出せば良い。さて、ここで、2番目の式を x_j で偏微分して1番目の式に代入すると、

$$dP/dt = \sum_{j=1}^n (dx_j/dt) \cdot \frac{\sum_{m_s=0}^{2*order} \sum_{k_s+l_s=0}^{m_s} a_{k_1, \dots, k_n} b_{l_1, \dots, l_n} (k_j - l_j) \prod_{i=1}^n x_i^{k_i+l_i}}{x_j \{ \sum_{l_1, \dots, l_n}^{order} b_{l_1, \dots, l_n} \prod_{i=1}^n x_i^{l_i} \}^2} \quad (3)$$

となる。これは入出力の増減関係を表す式であり、この式を定性値からなる定性値空間に量子化して、観測結果を代入することにより、それらを満たす係数列を推定することができる。

3.2 実験例

2つの入力を持ったブラックボックスに関する実験例を御紹介する。ブラックボックスの入出力関係について次の表2のようなデータを本手法に対して与える。このデータは、

$$P(x_1, x_2) = x_1 \cdot x_2 + x_1 + x_2$$

という因果構造を仮定したデータである。

そうすると次のようなモデルが、与えた関係を満たす構造として推論される。

$$p(x_1, x_2) = \frac{-constant + x_1 \cdot x_2}{constant + x_1 + x_2 + x_1 \cdot x_2}$$

$$p(x_1, x_2) = \frac{-constant}{constant + x_1 + x_2 + x_1 \cdot x_2}$$

$$p(x_1, x_2) = \frac{-constant + x_1 \cdot x_2}{constant + x_1 + x_2}$$

$$p(x_1, x_2) = \frac{constant + x_1 + x_2 + x_1 \cdot x_2}{constant}$$

$$p(x_1, x_2) = \frac{x_1 + x_2 + x_1 \cdot x_2}{constant}$$

$$p(x_1, x_2) = \frac{-constant + x_1 + x_2 + x_1 \cdot x_2}{constant}$$

$$p(x_1, x_2) = \frac{constant + x_1 + x_2 + x_1 \cdot x_2}{constant - x_1 \cdot x_2}$$

$$p(x_1, x_2) = \frac{constant + x_1 \cdot x_2}{constant - x_1 \cdot x_2}$$

4番目の式が、与えた構造と同じもの。それ以外は、構造は異なるが、与えた入出力関係を満たすモデルとすることができる。

定性推論を使ったモデル生成の特徴のひとつに、変量の値の不確定さを残したまま推論を進めることができる、ということが挙げられる。定性推論における計算のメカニズムは、基本的には値と区間に関するものであり、まったく値のわからない変量については、その値の範囲を実数の全区間とするような記述が許されるのである。

尚、本研究の詳細については [Abe 91a, Abe 91b] を参照されたい。

表 2: ブラックボックスの入出力関係に関するテストデータ

データの値はすべて定性値で、 $x_i > 0 \rightarrow [x_1] = [+]$, $x_i = 0 \rightarrow [x_1] = [0]$, $x_i < 0 \rightarrow [x_1] = [-]$, $x_i \in (-\infty, \infty) \rightarrow [x_i] = [?]$ の様に表記するものとする。

また、 ∂x_i は、 dx_i/dt を定性値の空間に量子化した値を意味する。

Data	$[x_1]$	$[x_2]$	∂x_1	∂x_2	∂P
<i>data</i> ₁	[+]	[+]	[+]	[+]	[+]
<i>data</i> ₂	[+]	[+]	[+]	[0]	[+]
<i>data</i> ₃	[+]	[+]	[+]	[-]	[?]
<i>data</i> ₄	[+]	[+]	[0]	[+]	[+]
<i>data</i> ₅	[+]	[+]	[0]	[0]	[0]
<i>data</i> ₆	[+]	[+]	[0]	[-]	[-]
<i>data</i> ₇	[+]	[+]	[-]	[+]	[?]
<i>data</i> ₈	[+]	[+]	[-]	[0]	[-]
<i>data</i> ₉	[+]	[+]	[-]	[-]	[-]
<i>data</i> ₁₀	[+]	[0]	[+]	[+]	[+]
<i>data</i> ₁₂	[+]	[0]	[+]	[0]	[+]
<i>data</i> ₁₃	[+]	[0]	[+]	[-]	[?]
<i>data</i> ₁₄	[+]	[0]	[0]	[+]	[+]
<i>data</i> ₁₅	[+]	[0]	[0]	[0]	[0]
<i>data</i> ₁₆	[+]	[0]	[0]	[-]	[-]
<i>data</i> ₁₇	[+]	[0]	[-]	[+]	[?]
<i>data</i> ₁₈	[+]	[0]	[-]	[0]	[-]
<i>data</i> ₁₉	[+]	[0]	[-]	[-]	[-]
...
<i>data</i> _{<i>j</i>}	[-]	[-]	[+]	[+]	[?]
<i>data</i> _{<i>j</i>+1}	[-]	[-]	[+]	[0]	[?]
<i>data</i> _{<i>j</i>+2}	[-]	[-]	[+]	[-]	[?]
<i>data</i> _{<i>j</i>+3}	[-]	[-]	[0]	[+]	[?]
<i>data</i> _{<i>j</i>+4}	[-]	[-]	[0]	[0]	[0]
<i>data</i> _{<i>j</i>+5}	[-]	[-]	[0]	[-]	[?]
<i>data</i> _{<i>j</i>+6}	[-]	[-]	[-]	[+]	[?]
<i>data</i> _{<i>j</i>+7}	[-]	[-]	[-]	[0]	[?]
<i>data</i> _{<i>j</i>+8}	[-]	[-]	[-]	[-]	[?]

4 実時間因果構造の同定

4.1 手法の概要

ここでは、実時間因果構造の同定に関する研究の概略を述べる。表3に例を示す。表1の例とは異なり、変量には独立変数、従属変数という区別がない。ただし、これらの n 個の変量は互いに完全に独立ではなく、相関を持つと仮定する。以下の議論ではそれがどのような相関関係、つまり因果構造となっているかを見い出すための手法について述べる。今度の手法は、データ間の相関を仮定してそれを利用することにより因果構造の同定を行なう方法である。データ間の相関としては、ここでは時間的な相関を利用する。つまり、データが時系列として与えられる場合を想定する（必ずしも、時間相関でなければならないということはない）。

表 3: n 個の属性の間の従属関係を仮定した決定表

Data	x_1	x_2	x_3	...	x_n
$data_1$	•	•	•	...	•
$data_2$	•	•	•	...	•
$data_3$	•	•	•	...	•
$data_4$	•	•	•	...	•
$data_5$	•	•	•	...	•
...					

4.2 状態推定規則とその適用例

ここでは、次のような簡単な例を用いて説明する。ある物理系が存在するとしてこの系は、 x_1, x_2 という2つの変量から成り、これらの時系列が表4のように観測されたとする。この例では、位相の異なった振動がやがてゼロに収束してゆくような、時系列をデータとして与えた。ここで、まず、定性的な値と時間相関を表す定性微分値からなる定性的状態を推定するという問題について考えてみよう。系を構成する変量に関する定性値自身は観測されているの値をそのまま用いるとして、ここでは、データの時間相関を利用して定性微分値を求めれば良い。一般に、データから時間相関を求める為には、となりあう2つの時刻のデータの値の差分を求めたり、連続した何点かを用いて着目点の微分値を推定するということが行なわれるが、変量が定性値の空間に量子化されているので、時間相関を計算する新たな枠組が必要ということになる。

そこで、量子化された定性値の空間で成立する次の3つの基本規則を導いた。

1. 変数変化の連続性の仮定

- (a) 境界標を跨ぐ変化の禁止則
- (b) 境界標から別の境界標への変化の禁止則

2. 瞬間/時区間判定則

時間的前後関係から、着目する時間量子が瞬間であるか時区間であるかを判定する規則

3. 定性的積分規則

着目する時間量子の状態から、次の状態を推定するための規則

表 4: 時系列データと状態推定規則の適用

観測された時系列 $[x_1][x_2]$	適用規則	推定された定性的状態 $([x_1], \partial x_1, [x_2], \partial x_2)$
$[-][+]$	} <i>Rule.1</i> →	$([0], [0+], [+], [+0-])$
$[0][+]$		
$[+][+]$	} <i>Rule.2</i> →	$([+], [+], [+], [+0-])$ $([+], [0+], [+], [+0-])$ $([+], [+0-], [+], [-0])$ $([+], [+0-], [+], [-])$
$[+][0]$		
$[+][-]$	} <i>Rule.2</i> →	$([+], [+0-], [0], [-0])$ $([+], [+0-], [-], [-])$ $([+], [-0], [-], [+0-])$ $([+], [+0-], [-], [-0])$ $([+], [-], [-], [+0-])$
$[0][-]$		
$[-][-]$	} <i>Rule.2</i> →	$([0], [-0], [-], [+0-])$ $([-], [-], [-], [+0-])$ $([-], [-0], [-], [+0-])$ $([-], [+0-], [-], [0+])$ $([-], [+0-], [-], [+])$
$[-][0]$		
$[-][+]$	} <i>Rule.2</i> →	$([-], [+0-], [0], [0+])$ $([-], [+0-], [+], [+])$ $([-], [0+], [+], [+0-])$ $([-], [+0-], [+], [0])$ $([-], [+], [+], [-])$
$[0][0]$		
$[0][0]$	<i>Rule.1</i> →	$([0], [0], [0], [0])$

これらの規則の導出および詳細については、[Abe 92a, Abe 92b, Abe 93a, Abe 93b, Abe 93c]で述べてあるのでここでは省略するが、基本的には変化の連続性を前提として、実数の空間で成立する関係を定性値空間に量子化したものである。ひとつめは、変数の変化の連続性を満たすために必要な規則。2つめは、観測された時間量子のデータが瞬間的な状態を表すものか、時区間を表すものかを判別するための規則。3つめは、変数の時間的な状態変化を記述する定性的積分規則である。これら3つの基本規則によって、先ほどの例の状態を推定したものが表4で示した定性的状態である。

4.3 制約規則の推定

モデル推定は前節の方法で推定された状態から、時間相関を含めたこれら4の変数 ($[x_1]$, ∂x_1 , $[x_2]$, ∂x_2) の間に成立する規則性を見いだすという問題に帰着できたことになる。そこで、そのような過程を次のような多項式を用いて実現する。

$$[0] \subset \sum_{m=0}^{order} [C_{k_1, \dots, k_n, l_1, \dots, l_n}] \prod_{i=1}^n [x_i]^{k_i} \prod_{j=1}^n \partial x_j^{l_j} \quad (4)$$

$$m = k_1 + \dots + k_n + l_1 + \dots + l_n$$

Σ : 定性的な加算.

Π : 定性的な乗算.

$[C_{k_1, \dots, k_n, l_1, \dots, l_n}]$: 多項式の係数列.

$order$: 多項式の次数.

この多項式 (量子化された多項式) に先ほどの定性的状態を次々と代入し、常にこの関係が成立するような係数列 C を推定する。この関係式は、複数量変数とその時間相関について定式化されたものとなりまして、定性的な常微分方程式に他ならない (order が1のときは線形常微分方程式、2以上は非線形となる)。表4の例の場合には、次のような6通りの力学モデルとして定式化される。

$$\begin{aligned} \text{Model No.1} & \left\{ \begin{array}{l} [0] \subset [x_1] + \partial x_1 + [x_2] + \partial x_2 \\ [0] \subset \partial x_1 - [x_2] \end{array} \right. \\ \text{Model No.2} & \left\{ \begin{array}{l} [0] \subset [x_1] + \partial x_1 + \partial x_2 \\ [0] \subset \partial x_1 - [x_2] \end{array} \right. \\ \text{Model No.3} & \left\{ \begin{array}{l} [0] \subset [x_1] + [x_2] + \partial x_2 \\ [0] \subset \partial x_1 + \partial x_2 \end{array} \right. \\ \text{Model No.4} & \left\{ \begin{array}{l} [0] \subset [x_1] + [x_2] + \partial x_2 \\ [0] \subset \partial x_1 - [x_2] \end{array} \right. \\ \text{Model No.5} & \left\{ \begin{array}{l} [0] \subset [x_1] - \partial x_1 + [x_2] + \partial x_2 \\ [0] \subset \partial x_1 + \partial x_2 \end{array} \right. \\ \text{Model No.6} & \left\{ \begin{array}{l} [0] \subset [x_1] - \partial x_1 + [x_2] \\ [0] \subset \partial x_1 + \partial x_2 \end{array} \right. \end{aligned}$$

5 データ共有技術について

前章までに述べてきた研究を足掛かりとした、データ共有技術についてその概要と要素研究課題について述べる。

5.1 数量的なデータの扱いについて

ここでは、まずデータが数量的な値から成る場合について考える。取り敢えず、次のような決定表形式のデータを仮定する。

$$DB = (\text{対象オブジェクトの集合, 条件属性の集合, 決定属性の集合, 値の集合})$$

$$\text{値の集合} = \{ \text{条件属性の値集合, 決定属性の値集合} \}$$

異なる目的で構築された n 個のデータベースをそれぞれ、

$$DB_d = (OB_d, at_d, AT_d, \{VAL_{at_d}, VAL_{AT_d}\})$$

$$d = 1, \dots, n$$

(ただし、 DB_d は共通の条件属性 $G \in AT_d : d = 1, \dots, n$ を持つものとする。) とする。このとき、これらのデータベースに一部共通する条件属性の値が与えられたときに、決定属性 G の値を予測し (たとえば区間, 近似値, 確率的表現などの形式でも良い)、第 $n+1$ 番目のデータベース DB_{n+1} を作ることは出来ないだろうか。ここで、

$$DB_{n+1} = (OB_{n+1}, at_{n+1}, \tilde{G}, \{VAL_{n+1}, \tilde{VAL}_{\tilde{G}}\})$$

(ただし、 $at_{n+1} \cap at_d \neq \phi : d = 1, \dots, n$

また、 \tilde{G}, \tilde{VAL} は予測値を表すものとする。)

たとえば、表5, 表6, 表7のような3つのデータベースが存在して、これらのデータベースに一部共通する条件属性 {IQ, 研修の成績, 職場適合度,} の値が与えられたときに社員評価値を推定して、表8のような第4のデータベース DB_4 をつくりたい。

表5: データベース1

社員集合	IQ	学力試験点数	口頭試問点数	社員評価
社員 ₁	150	800	900	0.8
社員 ₂	160	760	800	0.6
...				

4つのデータベースをマージしたデータベース DB_{merge} は、表9のような部分的に値の不明な“不完全情報システム [Lipski, JR 79]”となる。不完全情報システムの取り扱いについては明らかになっていない部分が多いが、データ共有におけるここでの問題は \tilde{VAL}_{AT} に関する予測問題に帰着させることができるのである。

5.2 従属関係の同定

前節の例を用いると、第4のデータベースの条件属性がその他のデータベースのものと完全に一致する場合には、それぞれの条件属性と決定属性の間の従属関係が、次の f_1, f_2, f_3 ように同定されていれば、社員評価の値は比較的容易に予測することができる。

$$\text{社員評価} = f_1(\text{IQ, 学力試験点数, 口頭試問点数})$$

表 6: データベース 2

社員集合	大学の成績	大学院の成績	研修の成績	社員評価
社員 ₃	400	500	400	0.9
社員 ₄	350	550	430	0.7
...				

表 7: データベース 3

社員集合	年齢	勤続年数	職能等級	職場適合度	社員評価
社員 ₅	25	1	2	80	0.5
社員 ₆	26	8	2	90	0.6
...					

表 8: データベース 4

社員集合	IQ	研修の成績	職場適合度	社員評価
社員 ₇	160	450	85	?
社員 ₈	180	480	92	?
...				

表 9: マージされたデータベース

社員集合	IQ	学力試験 点数	口頭試 問点数	大学の 成績	大学院 の成績	研修の 成績	年齢	勤続年 数	職能等 級	職場適 合度	社員評 価
社員 ₁	150	800	900	?	?	?	?	?	?	?	0.8
社員 ₂	160	760	800	?	?	?	?	?	?	?	0.6
...	•	•	•	?	?	?	?	?	?	?	•
社員 ₃	?	?	?	400	500	400	?	?	?	?	0.9
社員 ₄	?	?	?	350	550	430	?	?	?	?	0.7
...	?	?	?	•	•	•	?	?	?	?	•
社員 ₅	?	?	?	?	?	?	25	1	2	80	0.5
社員 ₆	?	?	?	?	?	?	26	8	2	90	0.6
...	?	?	?	?	?	?	•	•	•	•	•
社員 ₇	160	?	?	?	?	450	?	?	?	85	?
社員 ₈	180	?	?	?	?	480	?	?	?	92	?
...	•	?	?	?	?	•	?	?	?	•	?

社員評価 = f_2 (大学の成績, 大学院の成績, 研修の成績)

社員評価 = f_3 (年齢, 勤続年数, 職能等級, 職場適合度)

このような従属関係の同定については先に紹介したこれまでの研究を利用することが出来る。

これら f_1, f_2, f_3 という3つの関係式を“or”で結んだものをマージされたデータベースのモデルと言っても間違いであると言うことはできないが、社員評価値の予測のために用いることは困難である。結局は、マージされたデータベースのすべての条件属性と決定属性の間の従属関係を仮定して、その構造を同定しなければならないものと思われる（あるいは、何らかの統計的な手段により第4のデータベースの条件属性と決定属性の間の関係式としても良いかも知れない）。

社員評価 = f_{merge} (IQ, 学力試験点数, 口頭試問点数, 大学の成績,

大学院の成績, 研修の成績, 年齢, 勤続年数, 職能等級, 職場適合度)

f_{merge} の条件属性の値を為す変量に関する前提がなにもなしというのでは手法的にかなり苦しいので、たとえば各データベースの不明な属性についてはすべてのオブジェクトについて一定の値であると仮定する、などのように不明な値に関するある程度の制約を設ける必要があると思われる。

5.3 シンボリックなデータの扱いについて

ここでは、データが変量ではなくシンボリックなものである場合について述べる。ここでは、先ほどと同じように不完全情報システムについてと、平坦な構造のデータの取り扱いについて、考えてみたい。

次のような例題“頭痛患者のデータベース”(表10)を仮定する。このとき、患者6と患者

表 10: 頭痛患者のデータベース

患者集合	部位	性状	様式	吐き気	所属クラス
患者 ₁	全体	持続性	持続性	?	筋収縮性頭痛
患者 ₂	側頭部	拍動性	発作性	?	偏頭痛
...	•	•	•	?	•
患者 ₃	?	持続性	持続性	なし	筋収縮性頭痛
患者 ₄	?	持続性	持続性	あり	偏頭痛
患者 ₅	?	持続性	持続性	なし	心因性頭痛
...	?	•	•	•	•
患者 ₆	全体	?	?	なし	?
患者 ₇	側頭部	?	?	なし	?
...	•	?	?	•	?

7が訴えている頭痛がどのクラスに属するかを、部位と吐き気の有無から予測しようという訳である。この場合、先ほどのような数量的なデータを対象とした因果構造の同定のような方法を用いることは困難である。従属関係 f は、条件属性に関する複雑な論理式になることは明らかである。このような課題に対して、集合の近似的な取り扱いに関する“ラフ集合”という考え方があり[Rough Set 94]。ラフ集合を扱った研究の中には、決定規則の生成

という課題を扱ったものもあるので、それらを参考にしたい。ただし、不完全情報システムについては現在のところ足掛かりとなるような技術はなく、具体的アプローチは現状明確になっていない。

5.4 平坦な構造のデータの扱いについて

今度は、先ほどのような決定表のような構造を持たないデータについて考えてみる(例題“スクラップの分類”表11)。これは、長さ不定の文章によって記述された記事が蓄積され

表 11: スクラップの分類

記事集合	記事内容	興味の有無
記事 1	本報告では、仮想現実感に関する考え方を整理するとともに、仮想現実システムを構築するうえで表面化する人工知能技術と共通の課題について解説する。	有
記事 2	本稿は、発送支援ツールのなかで特に発散的思考を支援するツールや技術に関して解説する。……………。	有
記事 3	周期的に大爆発を引き起こす激変星は、公転周期がわずか数時間という非常に短い近接した連星系である……………。	無
記事 4	近年、図形を用いた推論の研究が注目を集めている。特に Barwise らによる異種推論の研究の興味は……………。	有
記事 5	放電管の中で発生させたプラズマ中の電子密度の空間分布を測定する手法に関する研究で……………。	無
記事 6	データベースとして蓄積された動画像から、シーン中の物体の運動などのような動的変化をキーとして……………。	無
記事 7	磁気ディスクや光ディスクなどの記録媒体に蓄積された情報をランダム且つ高速に……………。	無
…	……………	•

たものであるが、これらの記事群の中から、利用者の興味にあった記事を集めてスクラップブックのようなものを作りたいものとする。問題は、データ群のサブセットから記事を選んで作られた小さなスクラップブックを参考にして、それを作った人が喜びそうな記事をデータ群の中から見つけ出すことができるか、ということである。これは、平坦な構造のデータに対して興味の有無という新たな決定属性を付加して、その値を予測するという問題と見ることができる。これらのような記事は、最終的には単語のレベルまでブレイクダウンすることができる、先ほどまでの話題の単純な拡張を考えてみると。表12のように、条件属性としてあらゆる単語を取り、その値は、記事の中にその単語が含まれるか否かの2値とする、あるいは、記事の中にその単語が現れる回数とすることも可能かもしれない。また、決定属性については興味の有無を表す2値とすると、やや強引ではあるが、先ほどの変量に関する決定表と見ることができるだろう。ただし、変量の値の連続性は保証されないので、先ほどの因果構造の同定法はそのまま利用することは困難であると思われる。

5.5 データのタイプが混在する場合の扱いについて

最後は、今まで述べてきたすべてのケースがデータの中に混在する場合について述べる。一例として表13のような健康診断結果に関するデータベースを仮定する。これは様々なタイプのデータを統合的に利用して、健康に関する総合的な判断をするという予測問題と考えることができよう。

表 12: データを数量化したスクラップデータベース

記事集合	仮想現実感	整理	仮想現実システム	発想支援	発散的思考	ツール	周期的	大爆発	激変星	凶形	推論	研究	放電管	プラズマ	電子密度	データベース	動画像	シーン	磁気ディスク	光ディスク	記録媒体	...	興味の有無	
記事 ₁	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	1
記事 ₂	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	1
記事 ₃	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	...	0
記事 ₄	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	...	1
記事 ₅	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	...	0
記事 ₆	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	...	0
記事 ₇	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	...	0	
.....

表 13: 健康診断結果に関するデータベース

被検者集合	総蛋白	血糖	血圧	胸部 x 線	心電図	内科所見	総合判定
被検者 ₁	7.0	84	126/76	異常なし	正常範囲	アレルギー性鼻炎の疑いあり。	耳鼻咽喉科にて再検査の必要あり。
被検者 ₂	7.2	79	120/80	措置不要	正常範囲	健康者扱い。	異常なし
被検者 ₃	7.1	84	131/86	治癒所見	正常範囲	結核治癒所見もしくは第3肋骨に過形成所見あり。	胸部 x 線については半年毎のフォローが望ましい。
被検者 ₄	6.9	75	128/82	異常なし	正常範囲	健康者扱い。	異常なし
被検者 ₅	6.8	77	122/81	異常なし	正常範囲	健康者扱い。	異常なし
...

6 応用：創造的な活動の支援環境

将来的な応用例として、計算機環境を利用して蓄積された巨大なデータ群を前提として、次のような支援環境を考えることができる。

前章で述べた新聞のスクラップのような平坦なデータ構造を仮定して、ある利用者にとって今対象としているオブジェクト(記事)が興味の対象であるか否かという決定属性を推定する。これにより前章までに述べてきた手法を、

- 計算機環境との対話により、利用者が自分自身の興味を同定する(自分自身を知る)ことができる。
- 利用者の興味を刺激するデータを提示することにより、創造的な活動に役立てる(利用者の創造的活動の材料を提供することによりアイデアを生み出す過程を支援する)。

といったような環境の実現のために利用することが考えられる。

7 おわりに

数量的な規則性に関する定式化を行なうひとつの手法として、定性推論の枠組を利用したシステム同定法について概要を述べた。また、データ分散するエージェント間のデータ共有のために必要な要素技術課題を明らかにした。

参考文献

- [Abe 91a] 安部伸治、名倉正計, “ブラックボックスの入出力間の定性的因果関係モデルを自動獲得する手法の検討”, 1991年電子情報通信学会秋季全国大会(1991).
- [Abe 91b] 安部伸治、名倉正計, “ブラックボックスの入出力間の定性的因果関係モデルを自動獲得する手法の検討”, 情報処理学会第81回人工知能研究会, 情報処理学会研究報告 92-AI-81-2(1991.3.5).
- [Abe 92a] 安部伸治, “力学系の定性的同定問題”, 情報処理学会第45回全国大会.
- [Abe 92b] 安部伸治, “定性的システム同定法の提案”, 人工知能学会合同研究発表会 AI シンポジウム 92 資料 SIGF/H/K/S/I-9201, pp.81-88(1992).
- [Abe 93a] Shinji ABE, “A Qualitative System Identification Method”, The Seventh International Workshop on Qualitative Reasoning about Physical Systems(1993).
- [Abe 93b] 安部伸治, “定性的システム同定法の提案 - (1) 定性的な状態の推定”, 人工知能学会誌 Vol.8 No.5, pp.649-656(1993).
- [Abe 93c] 安部伸治, “定性的システム同定法の提案 - (2) 制約規則の推定”, 人工知能学会誌 Vol.8 No.5, pp.657-664(1993).
- [Cellier 91] Fransis E. Cellier, *General System Problem Solving Paradigm for Qualitative Modeling*, in Paul A. Fishwick and Paul A. Luker editors, *Qualitative Simulation Modeling and Analysis*, Advances in Simulation 5, Springer-Verlag, pp.51-71(1991).
- [de Kleer 84] Johan de Kleer and John Seely Brown, *Qualitative Physics Based on Confluences*, Artificial Intelligence, Vol.24, pp.7-83(1984).

- [Forbus 84] **Kenneth D.Forbus**, *Qualitative Process Theory*, Artificial Intelligence, Vol.24, pp.85-112(1984).
- [Kiryama 91] **Takashi Kiriyama, Masahiko Nakamura, Tetsuo Tomiyama and Hiroyuki Yoshikawa**, *Model Generation and Management Using Qualitative Reasoning*, Proceedings of 5th Annual Conference of JSAI, pp.269-272(1991).
- [Kuipers 86] **B.J.Kuipers**, *Qualitative Simulation*, Artificial Intelligence, Vol.29, pp.289-338(1986).
- [Nishida 87] **Toyoaki Nishida and Shuji Doshita**, *Reasoning with Model Lattices*, in M.Tokoro, Y.Anzai and A.Yonezawa editors, *Concepts and Characteristics of Knowledge-Based Systems*, pp.325-347, Elsevier Science Publishers B.V.(North-Holand),(1989). Selected and Review Papers from the IFIP TC 10/WG 10.1 Workshop Mount Fuji, Japan, 9-12 November(1987).
- [Nishida 89] **Toyoaki Nishida**, *Technique of Qualitative Reasoning*, in Kazuhiro Fuchi, Fumio Mizoguchi, Yasukazu Furukawa, Yuichiro Anzai editors, *Qualitative Reasoning*, chapter 3, pp.77-128, Kyoritsu-Print(1989).
- [Weld 89] **Daniel S.Weld and Johan deKleer** editors. *Readings in Qualitative Reasoning about Physical Systems*, Morgan-Kaufman (1989).
- [Lipski, JR 79] **Witold Lipski, JR.** *On Semantic Issues Connected with Incomplete Information Databases*, ACM Trans. on Database Systems, Vol.4, No.3, pp.262-296 (1979).
- [Rough Set 94] 中村 昭, 横森 貴, 小林 聡, 谷田 則, 米村 崇, 津本 周作, 田中 博, 数理科学リレー連載“ラフ集合”, 数理科学 No.373, pp.78-83, No.374, pp.86-91, No.375, pp.76-83, No.376, pp.78-85, No.377, pp.78-83, No.378, pp.76-83 (1994).