

半順序と情報処理*

鈴木 昇 一

Partial Ordering and Information—Processing

Shoichi Suzuki

* 本研究の一部は株式会社インターナショナルソフトウェア（〒112東京都文京区）によって支持された。

Abstract

We make it clear what common principle lies at sequences of the numerical calculations, the fuzzy operations, the neural-net operations, operations of a recursive program, the logical-resolution principle, the inductive inferences, the pattern recognitions and the extractions of any component from a system. It is reasonable that we consider the information-processing as sequences of operations of seeking after the least upper bound of a chain or a directed set \subseteq the data domain D , finding a partial ordering in D . An entropy associated with probabilistic events keeps on increasing as long as we leave events as it is. It is well known that the information-processing can play an active part in decreasing the entropy. The entropy-principle should be considered from a more general approach, that is, a principle of partial ordering presented here.

要 約

情報処理とは何かという疑問に答えたものである。データ領域の中にある半順序関係を導入し、その上限を求める操作が情報処理であると解釈すれば都合が良いことを数値計算、fuzzy 情報処理、ニューラルネット情報処理、プログラム計算、導出原理、帰納約推論、パターン情報処理、システムからの各要素システムの分離という諸例を介して、示している。

放置しておけば増大するエントロピーを減少させる働きが結局は、情報処理の働きであるとの結論を当然ながら得ている。

本研究によって各種情報処理に潜む共通な原理が浮彫になったと考えている。

1. まえがき

データ (data) とは、本来は立論の基礎になる論拠、資料ということである。(これに対し)、情と報とを結びつけた“情報”という言葉は、日本語の世界では、そう古いものではなく、森鷗外あたりの造語とも聞いたことがある。

情報の概念を基礎におき、これに関連する諸々の概念を導入し、情報の概念を全科学にわたる

基礎概念の一つとして設定する。他の基礎概念として、物質、エネルギーがある。これに情報を加えた3つをもって全科学の基礎概念とする。

物理学が生物科学に対して寄与して来た貢献を、われわれは基礎科学としての情報学 (informatics) が人文科学に対して果すべき役割と対比させてみる。そして、社会科学は人文科学、生物科学の両方面からつくりあげてゆく。

以上が、北川敏男理博の著書⁽²³⁾の各所から適当に抜粋した文章をつないだものである。

人間精神でさえ、脳システムの打ち出す情報であるという考え⁽¹⁵⁾があるのであるから，“情報”の概念を特定させるのは困難な仕事である。

物質 (matter) → エネルギー (energy) → 情報 (information)

という順に、得体が知れない存在に、ソフトウェア的になっているかも知れない。質量をエネルギーへと換算する比例係数としての“光の速さの自乗”の存在、秩序性の尺度である自由エネルギーを無秩序性の尺度としてのエントロピー (entropy)^{(28)~(31)}; (5)のIXへ換算する比例係数としての“絶対温度の逆数”の存在を思い起こすと、

質量、エネルギー、エントロピーの機能的等価性も知られている*¹ことになる。

得体が知れない“情報”というものを“処理”するとは数理的に何を意味しているのだろうか。明らかにデータ処理とは異なる。この種の疑問に答えるように努力した結果、本研究は生まれた。

著者は、情報処理の根底に、処理の効果を測定するための“半順序原理”があると考え、その解としての上限 (最小上界) に近づくような“近似の列”の各成分が“情報”であると定義したい。ある半順序関係⁽¹⁾の観点から並べることが可能な carrier, entity は情報を含有していると指摘したい。

著者は本論文で、情報学における“情報処理”とは、処理対象領域において適切に一種の半順序関係さえ設定できれば、入力を反映した初期値 (無知あるいは部分的に既知) から、入力が増すことによって出力の情報が増えることはない性質を持ったある種の単調作用素 (monotonic operator) を用いて、この半順序関係での上限 (未知, 解) に限りなく近づいていく“近似情報の列” (approximation information sequence) を求めていく過程であることを諸例を介し、明らかにしようと思う。

汎用コンピュータにおける情報処理動作は元来、非可逆的である⁽²⁾。例えば、3 プラス 2 を計算して結果 5 を得る項書き換え動作⁽³⁾ (term-rewriting operation)*²

$$3+2 \Rightarrow 5$$

においては、右辺の“5”は左辺の“3+2”より情報を圧縮した形になっており、“5”は“3+2”の情報を含むと考えられる。

何故ならば、

$$0+5 \Rightarrow 5, 1+4 \Rightarrow 5, 2+3 \Rightarrow 5, 4+1 \Rightarrow 5, 5+0 \Rightarrow 5$$

でもあるから、同一の結果“5”を得る左辺の情報は“3+2”以外にあるからである。この意味では

$$x \sqsubseteq y \quad (y \text{ は } x \text{ の情報を含む})$$

という半順序関係 (partial ordering) \sqsubseteq の下で

$$3+2 \sqsubseteq 5$$

と書ける。

目標 (goal concept), 訓練例 (training examples), 領域知識 (domain theory), 実行可能性基準 (operationality criterion) の4つが与えられた⁽⁴⁾場合*³の情報処理は解を求めていく方向 (半順序関係) からは, 情報を付加し, より詳しく定義されているように, データを変換する操作の列であり, 解とはこの半順序関係の極大要素に限りなく近づいたデータ (上限) である。求解過程は半順序関係の最小要素から出発せねばならない場合もあるが, この場合は全く事前知識を持っていない“無知”からの開始を意味する。情報処理に伴う知能の働きのある部分は, あるかも知れない解の一つに近づく方向へ, データ空間を探索することについて費やされるといえよう。

数値情報 (単値情報; single-valued information)

fuzzy 情報 (多値情報, あいまい性情報; many-valued information, fuzzy information)

ニューラル情報 (分散性情報; neural net information, distributed information)

言語・記号情報 (局所性情報; symbolic information, local information)

パターン情報 (変形可能な情報; pattern information, deformable information)

等の処理において, 底に流れている基本原理は何か?

本論文では

(i) 数値情報処理では

i-1 方程式 $f(x)=0$ の解法としての Newton-Raphson 法

i-2 連立1次方程式 $A\vec{a}=\vec{b}$ の解法

(ii) fuzzy 情報処理では

$[P \rightarrow Q] \wedge P$ ならば Q という推論法規則 (inference rule) としての三段論法 (modus ponens)

(iii) 一つ概念を一つの処理装置 PE (processing element) に割り当てる (情報の局所表現; local representation) のではなく, 多数の PE (ニューロン, neuron; 神経細胞) の結合における活性化度 (activation) の分布の違いが情報間の意味の相違とみて, 処理装置間の結合そのものが情報処理機能の中核とみるニューラル情報処理⁽³⁵⁾では

iii-1 Hopfield neural net によるエネルギーを最小にする多変数の値の求解法 (組み合わせ最適化法)

iii-2 error back propagation learning neural net による, 入力-出力関係の, 多数の事例からの同定法 (知識獲得法)

(iv) 一つの情報が分解すべき単位をもたない“記号”に基づく情報処理では

iv-1 プログラムによる処理 (プログラム計算)

iv-2 演繹的推論 (deductive reasoning) の一種である導出原理 (resolution principle)

iv-3 帰納的推論 (inductive reasoning) の一種であるモデル推論 (model inference)

(v) 一つの情報が分解すべき単位を持ってよい“パターン”の情報処理では

S. Suzuki の提唱した構造受精変換法

の各々の内に潜む基本原理を次の形で指摘する:

与えられた問題に応じ、対象の集合内に半順序関係 \sqsubseteq を設定し、この問題から定まる初期値 x_0 から出発し、

有向集合 (directed set) $d_n = \{x_0, x_1, \dots, x_n\}$ を作り、その上限

$$x = \sqcup_n d_n$$

を求めること。 □

claude E. Shannon が1948年に Bell system Technical Journal に発表した論文
通信の数学的理論 (A Mathematical Theory of Communication)

では、情報を運んでいるもの (carriers of information) がシンボル^{*4} (symbol), 記号 (string, letter) である場合の情報理論 (information theory) が構築されている。この理論は、生起確率分布をもつ記号の集合上の情報の尺度 (information measure) としての“情報の量” (amount of information) が、無記憶情報源のシンボル当りの平均情報量 (average amount of information), あるいはエントロピー (entropy) として、

$$-\sum_{k \in K} p_k \log_2 p_k,$$

ここに $0 \leq p_k \leq 1, \sum_{k \in K} p_k = 1$

という形で定義されている。この量は、ある属性が第 $k \in K$ 番目の値をとる確率が p_k である情報のもつ不確定さ (uncertainty) を計量化したものであり、確率分布

$$\vec{p} = \{p_k | k \in K\}$$

をもつ情報が確率分布

$$\vec{q} = \{q_k | k \in K\}$$

をもつ情報より、不確定さの程度が大きいとは、

$$-\sum_{k \in K} p_k \log_2 p_k \geq -\sum_{k \in K} q_k \log_2 q_k \quad (1.1)$$

が成り立つことであり、この事態を

$$\vec{p} \sqsubseteq \vec{q} \quad (1.2)$$

と書けば

$$\vec{p} \sim \vec{q} \Leftrightarrow -\sum_{k \in K} p_k \log_2 p_k = -\sum_{k \in K} q_k \log_2 q_k$$

の定義の下で、 \sqsubseteq はやはり \vec{p} の集合上の半順序関係である。この場合の情報処理とは

$$\vec{p} = \sqcup \{\vec{p}_1, \vec{p}_2, \dots\}$$

を求める変換

$$\vec{p}_{k+1} = f(\vec{p}_k)$$

をもたらす働きであり、上限 \vec{p} は

$$\exists k \in K, p_k = 1 \wedge (\forall j \in K - \{k\}, p_j = 0)$$

であることがわかる。上限 \vec{p} のもつ情報の量は 0 であり、不確定さは全く排除されていることがわかる。この種の情報処理の特異性は上限 \vec{p} をもたらず作用を得る過程で

$$-\sum_{k \in K} p_{nk} \log_2 p_{nk}$$

$$\text{ここに, } \vec{p}_n = \{p_{nk} | k \in K\}$$

$$0 \leq p_{nk} \leq 1, \sum_{k \in K} p_{nk} = 1$$

という情報の量を、情報処理システムが受け取るとしたことである。

f の働きで、 \vec{p}_n から

$$\vec{p}_{n+1} = f(\vec{p}_n)$$

という \vec{p}_{n+1} へ変換されたとき、この変換 f は

$$-\sum_{k \in K} p_{nk} \log_2 p_{nk} - [-\sum_{k \in K} p_{n+1k} \log_2 p_{n+1k}]$$

だけの“あいまい性”を減少させたことになり、正に情報処理といえるものになっていることであらう。

本研究で説明される“半順序原理”は実は、著者が現在研究続行中の「パターン認識の数学的理論⁽⁵⁾」での“情報処理原理”に気付いたとき、類推 (analogy) で思いついたものである。また、この観点から、大規模システム (large-scale system) からその要素システムを分離するというのは一種の情報処理であるとの感触を得たので、第 7 章でこの論を展開する。

カテゴリ (category) とは類概念のことであり、パターン (pattern) の意味とはその帰属しているカテゴリのことであるとされる。また、一般に、パターンとは、シンボルと対比したものであり、形態上、ある程度変形が許され、少し位の変形でもその意味が保有される媒体である。パターン認識情報処理学では、典型的なパターンとしてのプロトタイプ (prototype) (各カテゴリの代表パターン) に関する認識を使って、入力としてのパターンがどの一つのカテゴリに帰属するかの決定を問題とする。

S. Suzuki の

パターン認識の数学的理論⁽⁵⁾ (A Mathematical Theory of Recognizing Patterns)

では、パターン集合のつくる領域の構成 (domain construction) を基に、情報を運んでいるものがパターン (可分な一般抽象ヒルベルト空間の元) である場合の情報理論を構築しようとしている。本論文では、この S. Suzuki 理論が、従来の情報処理を如何に捨像して得られたかを念頭において、各種情報処理の働きが半順序関係を保存する写像としてとらえられることを解説しながら、情報の本質とは何かを浮彫りにしたい。

2. 誤差に基づく半順序関係——数値情報処理——

先ず、数値情報処理から、情報処理原理についての hint を得よう。

方程式

$$f(x) = 0, \text{ここに } x \text{ は実数値をとる変数}$$

を解く場合を考えよう。任意の x について

$$|f(x) - 0|$$

は x が方程式 $f(x) = 0$ の解でない程度 (誤差) を反映した量であることに注意して, 不等式

$$|f(x_1)| \geq |f(x_2)|$$

が成立する事態を

$$x_1 \sqsubseteq x_2$$

と書くと,

- (i) $x \sqsubseteq x$ (反射律; reflexive law)
- (ii) $x \sqsubseteq y$ かつ $y \sqsubseteq x$ ならば
 $x \sim y$ (反対称律; antisymmetric law)
- (iii) $x \sqsubseteq y$ かつ $y \sqsubseteq z$ ならば
 $x \sqsubseteq z$ (推移律; transitive law)

が満たされ, \sqsubseteq は半順序関係⁽¹⁾ (partial ordering) と呼ばれているものとなる。ただし,

$$|f(x_1)| = |f(x_2)|$$

という成立を $x_1 \sim x_2$ と書いている。

このとき, 逐次接近法 (successive approximation) で, 半順序関係の増大列

$$x_0 \sqsubseteq x_1 \sqsubseteq \cdots \sqsubseteq x_n \sqsubseteq x_{n+1} \sqsubseteq \cdots$$

をもたらす数列^{*5}

$$\{x_0, x_1, \cdots, x_n, x_{n+1}, \cdots\}$$

を求めると, 方程式 $f(x) = 0$ の近似解 x は, その上限

$$x = \sqcup \{x_0, x_1, \cdots, x_n, x_{n+1}, \cdots\}$$

で表現されることがわかる。よって, 関係

$$x_i \sqsubseteq x_{i+1}$$

を満たす様な x_i, x_{i+1} の関係を見つけることが上述の求解動作において必要なことの一つとなる。

例えば, 方程式 $f(x) = 0$ を不動点方程式の形に

$$x - g(x) = 0$$

に変換できる場合は, x_i, x_{i+1} の関係は

$$x_{i+1} - g(x_i) = 0$$

とおけるが,

$$x_i \sqsubseteq x_{i+1} \Leftrightarrow x_i \sqsubseteq g(x_i)$$

$$\Leftrightarrow |x_i - g(x_i)| \geq |g(x_i) - g(g(x_i))|$$

が保証されるためには、関数 g 、初期値 x_0 に関し諸条件が課されなければならない。

このことを例で考察してみよう。

例えば、Newton-Raphson 法⁽⁶⁾法では、この半順序関係 \sqsubseteq はある条件の下で保証されることは次の様にしてわかる。

方程式 $f(x) = 0$ の求解にあたって、 x - y 直交座標面における座標点 $(x_i, f(x_i))$ における関数 $f(x)$ の接線の方程式

$$y - f(x_i) = m(x - x_i)$$

$$\text{ここに、 } m_i \equiv \frac{d}{dx}f(x) \Big|_{x=x_i} = f'(x_i)$$

を導入し、この接線と x 軸との交点を x_{i+1} とするのが Newton-Raphson 法である。上式の接線の方程式において、

$$x = x_{i+1}, y = 0$$

とおいて、 x_{i+1} を求めれば、

$$x_{i+1} = x_i - f(x_i) / m_i$$

となり、

$$(a) f(x_i) \geq 0 \wedge f(x_{i+1}) \geq 0$$

$$(x_{i+1} - x_i) \cdot f'(x_i) \leq 0$$

あるいは

$$(b) f(x_i) \leq 0 \wedge f(x_{i+1}) \leq 0 \wedge$$

$$(x_{i+1} - x_i) \cdot f'(x_i) \geq 0$$

であれば、関数 $y = f(x)$ の Taylor 展開の近似式

$$f(x_{i+1}) = f(x_i) + (x_{i+1} - x_i) \cdot f'(x_i)$$

からわかるように、

$$|x_{i+1} - x_i| \text{ が十分小なる限り}$$

$$x_i \sqsubseteq x_{i+1}$$

が成立する。

もう一つ、例をあげておこう。

連立 1 次方程式

$$\sum_{j=1}^n a_{ij} \cdot x_j = b_i, \quad i = 1 \sim n$$

の解

$$\vec{x} = (x_1, x_2, \dots, x_n)$$

を求めることを考えよう*6。

$$f(\vec{x}) = \sum_{i=1}^n |R_i(\vec{x})|$$

$$\text{ここに, } R_i(\vec{x}) = \sum_{j=1}^n a_{ij} \cdot x_j - b_i$$

を導入すると, 方程式

$$f(\vec{x}) = 0$$

の解 \vec{x} が求めるものであることに注意する。

$$\Delta x_i = (0 \sim 0 \triangle x_i 0 \sim 0)$$

(第 i 番目の要素のみが $\triangle x_i$ の行ベクトル)

として, \vec{x} から

$$\vec{x} + \Delta \vec{x}_i = (x_1, x_2, \dots, x_{i-1}, x_i + \Delta x_i, x_{i+1}, \dots, x_n)$$

を作り,

$$R_i(\vec{x} + \Delta \vec{x}_i) - R_i(\vec{x})$$

を計算すると,

$$R_i(\vec{x} + \Delta \vec{x}_i) - R_i(\vec{x}) = a_{ii} \cdot \Delta x_i$$

となる。ここで,

$$R_i(\vec{x} + \Delta \vec{x}_i) = 0$$

を満たす Δx_i を求めると,

$$\Delta x_i = -R_i(\vec{x}) / a_{ii}$$

である。

\vec{x} が入力された第 k 段階では, $\vec{x}_k = \vec{x}$ であるが, ここで,

$$|R_i(\vec{x})| \geq \max_{j \neq i} |R_j(\vec{x})|$$

を満たす i を選び,

$$R_i(\vec{x} + \Delta \vec{x}_i) = 0$$

が満たされるように, \vec{x}_{k+1} を

$$\vec{x}_{k+1} = \vec{x}_k + (\Delta \vec{x}_k)$$

$$\text{ここに, } (\Delta \vec{x}_k) = \Delta \vec{x}_i$$

とおけば、各 k に対し、ある $m_k \geq k$ が存在し、

$$|f(\vec{x}_k)| \geq |f(\vec{x}_{m_k})|$$

が満たされることが期待され、

各 k に対し、ある $n_k (\geq 1)$ が存在して、

$$\vec{x}_k \sqsubseteq \vec{x}_{k+n_k}$$

つまり、

$$\vec{y}_k \sqsubseteq \vec{y}_{k+1}, \text{ ここに } \vec{y}_{k+1} = \vec{x}_{k+n_k}$$

の成立が期待される。

$$\{\vec{x}_0, \vec{x}_1, \vec{x}_2, \dots\}$$

はこの場合、鎖 (chain) ではなしに、条件を緩めた有向集合というものになっている。ちなみに、半順序関係 \sqsubseteq が導入された集合 D の部分集合 X が有向集合 (directed set) であるとは、 X の任意の 2 元 x, y (従って零個でない任意の有限個の元) について

$$x \sqsubseteq z \wedge y \sqsubseteq z$$

であるような $z \in X$ が存在することをいう。

よって、上述の連立 1 次方程式の解 \vec{x} が

$$\begin{aligned} \vec{x} &= \sqcup \{\vec{x}_0, \vec{x}_1, \dots, \vec{x}_k, \dots\} \\ &= \sqcup \{\vec{y}_1, \vec{y}_2, \dots\} \end{aligned}$$

と求められる。

以上を整理してみよう。

知能とは問題を解決する場面において必要とされる推論能力のことである、と考えよう。情報を捨象化した形式で、情報を知識として獲得したり (知覚・学習)、問題解決に便利のように知識を表現したり (構造的符号化、記憶)、知識を組み合わせ既知から未知を推定するため加工・利用したり (推論、探索、認知) する場面における問題解決能力を指しているといえよう。

反復法による方程式の求解過程

fuzzy 情報 (多値論理情報) における推論過程

ニューラル情報 (情報の分散表現; distributed representation) による “解くべき問題を反映している初期状態” から結果を表す目標状態への軌跡 (計算)

再帰プログラム (recursive program) の構造の決定法

2 値述語論理における演繹あるいは帰納推論

パターン情報処理におけるパターンの帰属すべきカテゴリの決定法

等における “情報処理基本原理” とは次のように指摘される:

取り扱うデータ領域 (data domain; パターン、シンボルなどの処理対象とするデータの中で、ある一定の性質を備えたものをまとめたもの) に、各データの持つ情報のあいまいさ (不確定さ) が減少する性質をもつ半順序関係を導入すれば、その半順序関係での有向集合 (あるいはそ

の特別なものとしての鎖) を作り出すことが問題解決における情報処理であり、解はこの有向集合列の上限 (あいまいさ、不確定さが最も少ないもの) として与えられる、というものである。

情報処理とは

データ構造 (data structure) 化したデータ領域に半順序関係 \sqsubseteq を発見し、初期値 x_0 から出発し、存在するであろう解 x に近づく鎖 (chain)

$$x_0 \sqsubseteq x_1 \sqsubseteq x_2 \sqsubseteq \dots \sqsubseteq x$$

(x へ至る近似の鎖; chain of approximations)

あるいは有向集合

$$\{x_0\}, \{x_0, x_1\}, \{x_0, x_1, x_2\}, \dots$$

を作ること

である。

備考 2. 1 データ領域を

$$N = \{0, 1, 2, \dots\}$$

(非負整数の集合)

とすれば、 N に関する一つのデータ構造の例として、

$\langle N; \text{SUC} \rangle$

があげられる。ここに、 N 上の後者関係 (successor relation) SUC は、

$$\text{SUC} = \{ \langle a, b \rangle \mid b = a + 1, a = 0, 1, 2, \dots \}$$

と定義される。数の大小関係 (半順序関係の一つ) \sqsubseteq が持ち込まれたデータ領域 N はデータ構造である。データ構造としての N の上では、加算・乗算なる 2 演算が可能であり、大小関係・等号関係が判定できる。このように、データ構造とはデータ領域にその上で可能な基本演算 (の意味する半順序関係) ら持ち込まれた集合と考えられる。

一般に、データ構造とは、定義域 (domain) と呼ばれる集合 D と、 D 上の関係の集合 R_1, R_2, \dots, R_k との作る $(k+1)$ 一組

$\langle D; R_1, R_2, \dots, R_k \rangle$

のことである⁽³³⁾。有用なデータ構造は、定義域が、有限個の初期要素から、データ構造に与えられた関係の中に埋め込まれている有限個の演算子を用いても生成 (generate) できる再帰的データ構造 (recursive data structure) である。

著者の考えによれば、データ領域 D に許される演算の集合を

$$\{A_1, A_2, \dots, A_m\}$$

とすると、面帰領域方程式 (reflective domain equation)

$$D = D_B \cup \{A_1 \vee A_2 \vee \dots \vee A_m\} \cdot D$$

を、再帰的データ構造 D を決定する方程式として採用できよう。この解 D は

$$D = \lim_{k \rightarrow \infty} \bigcup_{n=0}^k (A_1 \vee A_2 \vee \cdots \vee A_m)^n \cdot D_B$$

と与えられる。ここに、 D_B は基本領域 (basic domain) と呼ばれ、有限個の初期要素から成る集合である。 (備考 2. 1 終り)

$y \sqsubseteq z$ の意味は

“ z は y より詳しく定義されている” とか

“ z は y の情報を含む”

と解釈することで得られる。一般に、 $y \sqsubseteq z$ が成り立つとき、

z の持つ情報は y の持つ情報よりあいまいさ (不確定さ) が減少している

といえよう。このような解 x は、 $\sqcap E$ をデータ領域 D の部分集合 E の上限^{*7} とすれば

$$x = \sqcap \{x_0, x_1, x_2, \dots\}$$

と与えられる。 $x_i \sqsubseteq x$ が成り立つから、各 x_i の持つ情報は解 x に含まれるという意味で、 x_i は解 x を近似している。また、

$$i \leq j \text{ であれば } x_i \sqsubseteq x_j$$

であるから、 x_j は x_i よりも良い「解 x の近似」である。

今、 x_i から x_{i+1} を得る働きを関数 h として

$$x_{i+1} = h(x_i)$$

と書くと^{*8}、情報処理とは、無限の対象

$$x = \bigcup \{x_0, x_1, x_2, \dots\}$$

へ、有限の対象の列

$$\{x_0\}, \{x_0, x_1\}, \{x_0, x_1, x_2\}, \dots$$

によって近似しながら限りなく接近することであり、 h という定まった操作を繰り返して目標とする対象 (解 x) に近づく過程として考えられる。解 $x = \sqcap \{x_0, x_1, \dots\}$ は、 h の不動点方程式 (equation of fixed point)

$$x = h(x)$$

を満たし、 $x' = h(x')$ を満たす x と異なる解 (不動点解) x' があれば、

$$x \sqsubseteq x'$$

を満たすという意味で、解 x は最小不動点 (least fixed point) として確定させるのがよいと思われる。

備考 2. 2 連続関数の最小不動点定理⁽²⁾ (least fixed point theorem of a continuous function) によれば、

$$f: D \rightarrow D$$

を連続関数とすると,

$f(x) = x$ なる $x \in D$ で最小のものを α と表現すると, この α は

$$\begin{aligned} \alpha &= \sqcup \{ \perp, f(\perp), f^2(\perp), \dots \} \\ &= \sqcup_{n=0}^{\infty} f^n(\perp) \end{aligned}$$

で与えられる。ここに, ϕ を空集合とすると, \perp は $\perp = \sqcup \phi$ と定義される無定義要素 (undefined element) である。

半順序集合 D は, その任意の部分集合 (有限部分集合とは限らないことに注意) が上限をもつとき, 完備束 (complete lattice) であるという。ちなみに, 束 (lattice) とは, 上限が任意の有限な部分集合に対し存在するような半順序集合である。 D_1, D_2 を完備束とすると, $f: D_1 \rightarrow D_2$ が連続であるとは,

任意の有向集合 $x \subset D_1$ に対し,

$$f(\sqcup x) = \sqcup f(x)$$

が成り立つことである。 $f: D_1 \rightarrow D_2$ が連続であれば, 単調であることが証明される⁽²⁾が, 次の(a), (b)より, 関数 f の単調性と連続性との差異は有向集合 x が無限集合のときのみ存在すること⁽²⁾がわかる:

(a) $\sqcup X \in X$ であれば, $f(\sqcup X) \sqsubseteq \sqcup f(X)$ は任意の f について成立する

(b) 有向集合 X に対し, $\sqcup X \notin X$ ならば X は有限集合ではあり得ない。 (備考 2. 2 終)

当然ながら, x_i から x_{i+1} を求める関数

$$h: D \rightarrow D$$

の働きを実行可能なように組み立てれば,

解 x を求める構成的な方法 (解 x へ具体的に接近する実行可能な手段)

が得られることに注意しておこう。

上述のごとく, 情報処理の働きをとらえれば, **知能**とは正に, 有向集合あるいは鎖を発現する働きといえよう。

3. fuzzy 情報処理

その意味を受け取っても一意的に確定しなくて, あいまい性 (不確定さ, fuzziness) が残るとき情報処理を考えよう。一般に, 決定しようとしても決定しようとしても一意的に決定すること自体がそんなに意味がなくて, その内のいずれかの性質, 値をもつかが漠然なまま処理する場合の情報 が fuzzy information である。一意的に決定されない“非決定性” (non-deterministic) の一種があいまい性であり, 強制的に決定しても決定することの効果 があまり認められない場合 残るあいまい性をもった情報が fuzzy information である。

ファジィ情報 A は, A を集合 (ファジィ集合; fuzzy set) とみて, 要素 x が A に帰属する程度をあらわす関数 (帰属度関数; membership function)

$$\mu_A : X \rightarrow \{0, 1\}$$

よって、要素 x とそのグレード (grade) $\mu_A(x)$ との対の集合

$$A = \{(x, \mu_A(x)) \mid x \in X\}$$

として表現される。あるいは、この等価な表現法として、

$$A = \sum_{i=1}^n \mu_A(x_i) / x_i$$

… X が離散集合 $\{x_1, x_2, \dots, x_n\}$ の場合

$$A = \int_x \mu_A(x) / x$$

… X が連続集合の場合

を採用することがある。 $\mu_A(x)$ は x が A に帰属する程度 (degree ; grade of membership) である。

真の解 x がデータ領域 D の、広がりをもったある部分に存在しているけれどもそのどこの位置に正しく存在しているかが不鮮明になっているまま処理することを fuzzy 情報処理ということにする。勿論、二つのしきい値 (threshold) α, β を

$$0 \leq \alpha < \beta \leq 1$$

と設け、

- (イ) $\mu_A(x) \geq \beta$ ならば、 x は A に帰属する
- (ロ) $\mu_A(x) \leq \alpha$ ならば x は A に帰属しない
- (ハ) $\alpha < \mu_A(x) < \beta$ ならば、 x は A に帰属するとも帰属しないともいえない中間的な状態にある

という解釈において、 $\alpha = \beta$ とすると、形式的にはあいまい性が排除される。

ファジイ関係 (fuzzy relation)

ファジイ事象 (fuzzy event)

ファジイエントロピー (fuzzy entropy)

ファジイ位相空間 (fuzzy topological space)

ファジイ論理 (fuzzy logic)

ファジイ補間 (fuzzy interpolation)

ファジイ行列 (fuzzy matrix)

ファジイシステム (fuzzy system)

ファジイアルゴリズム (fuzzy algorithm)

ファジイ多段決定 (fuzzy multi-stage decision)

ファジイ言語 (fuzzy language)

タイプ2ファジイ集合

L-ファジイ集合

などの諸概念が登場するが⁽³²⁾、本研究ではファジイ論理と推論規則に的を絞って論じよう。

データ領域 D の中に一種の半順序 (partial ordering) \sqsubseteq を導入できると、 D の2要素であるデータ x, y を比較できる。もし、データ x, y の間に、

$$x \sqsubseteq y$$

が成立していると判明した際には

y は x より詳しく定義されている

y は x の情報を含む

とか,

x の情報は y の情報の一部分である

x は y を近似する (approximate)

との解釈が可能になる。

ファジィ情報 x, y については, 不等式 $0 \leq x, y \leq 1$ を満たす実数とし,

$$y \leq x \leq 2^{-1}, 2^{-1} \leq x \leq y \quad (3.1)$$

のいずれか一方が成立する場合を

$$x \sqsubseteq y \quad (3.2)$$

と書くと, 2 元関数 \sqsubseteq は 2 章のイ, ロ, ハを満たし,

$$D = \{x \mid 0 \leq x \leq 1\} \quad (3.3)$$

の上の半順序であることが知れる。この場合は確かに

x のあいまいさは y のもつあいまいさより大である

あるいは

y のもつあいまいさは x のもつあいまいさより小である

との解釈が可能になる。この解釈が fuzzy 推論の基礎である。0 を偽 (false), 1 を真 (true) とする 2 値記号論理と異なり, 命頭 p が真である程度

T(p)

を, 閉区間 $[0, 1] = \{x \mid 0 \leq x \leq 1\}$ 上の値で表現する fuzzy 推論では, ファジィ集合, ファジィ概念を想定することにより, 記号推論よりも少量ではあるが情報意味内容に立ち入った推論規則が考えられていることである。

3.1 あいまいさを保存あるいは単調変換する演算

$$0 \leq x \leq 1, 0 \leq y \leq 1$$

として, 2 元関係

$$x \sqsubseteq y$$

とは, 式 (3.1) の代りに,

$0 \leq h \leq 1$ なるしきい値 (a threshold of the reliability of the information) h を固定した条件の下で

$$y \leq x \leq h \vee h \leq x \leq y \quad (3.4)$$

が成り立つことを定義する。

命題 3. 1 式 (3. 4) で定義された 2 元関係 \sqsubseteq は 2 章のイ, ロ, ハを満たし, 半順序である。

x, y の内, 大きい方, 小さい方を各々,

$$\max(x, y), \min(x, y)$$

と書き, また, \bar{x} を

$$\bar{x} \equiv 1 - x$$

と定義する。 x から \bar{x} を得る演算は補演算ということにする。

次の命題 3. 2 は, 補演算がしきい値 $h = 2^{-1}$ のときに限り, あいまいさを保存することを指摘している。

命題 3. 2 *9

$$x \sqsubseteq y \text{ であれば } \bar{x} \sqsubseteq \bar{y}$$

が成り立つしきい値 h は $h = 2^{-1}$ のみである。

次の命題 3. 3 の i, ii は各々, \max 演算, \min 演算があいまいさを保存することを指摘している。

命題 3. 3 *9

$$x_1 \sqsubseteq x_2 \wedge y_1 \sqsubseteq y_2$$

であれば,

$$(i) \max(x_1, y_1) \sqsubseteq \max(x_2, y_2)$$

$$(ii) \min(x_1, y_1) \sqsubseteq \min(x_2, y_2)$$

次の命題 3. 4 の i, ii は, \max 演算, \min 演算があいまいさを減少させる働きを備えていることも示している。

命題 3. 4

$$(i) x \sqsubseteq y \text{ であれば } x \sqsubseteq \max(x, y).$$

$$(ii) x \sqsubseteq y \text{ であれば } x \sqsubseteq \min(x, y).$$

$$(iii) h = 2^{-1} \text{ の条件の下で,}$$

$$h \leq \max(x, \bar{x}) \leq 1 \text{ かつ}$$

$$0 \leq \min(x, \bar{x}) \leq h$$

が成立する。

(命題 3. 2 の証明)

$x \sqsubseteq y$ が成り立っているとしよう。このとき,

$$h \leq x \leq y \vee y \leq x \leq h$$

が成り立っている。

(i) $h \leq x \leq y$ の場合

$$1 - h \geq \bar{x} \geq \bar{y}$$

を得るが,

$$h \geq 1 - h \tag{3.5}$$

であれば, $\bar{x} \sqsubseteq \bar{y}$ が成り立つ。

(ロ) $y \leq x \leq h$ の場合

$$\bar{y} \geq \bar{x} \geq 1 - h$$

を得るが,

$$1 - h \geq h \tag{3.6}$$

であれば, $\bar{x} \sqsubseteq \bar{y}$ が成り立つ。

イ, ロのいずれが起こってもよいから, 2式(3.5), (3.6)から

$$h = 1 - h$$

を得, これを解くと, $h = 2^{-1}$ がいえた。 □

(命題3.3の証明)

$x_1 \sqsubseteq x_2$ より

$$1a \quad x_2 \leq x_1 \leq h$$

$$1b \quad h \leq x_1 \leq x_2$$

のいずれか一方が成り立っており, また, $y_1 \sqsubseteq y_2$ より

$$2a \quad y_2 \leq y_1 \leq h$$

$$2b \quad h \leq y_1 \leq y_2$$

のいずれか一方が成り立っている。

i の証明

A. $x_2 \leq x_1 \leq y \wedge y_2 \leq y_1 \leq h$ の場合

A.1 $\max(x_1, y_1) = x_1$ の場合

$$x_2 \leq x_1 = \max(x_1, y_1) \leq h$$

$$y_2 \leq y_1 \leq \max(x_1, y_1) = x_1 \leq h$$

がいえる。よって,

$$\max(x_2, y_2) \leq \max(x_1, y_1) \leq h$$

が成り立ち,

$$\max(x_1, y_1) \sqsubseteq \max(x_2, y_2)$$

A.2 $\max(x_1, y_1) = y_1$ の場合

$$x_2 \leq x_1 = \max(x_1, y_1) = y_1 \leq h$$

$$y_2 \leq y_1 = \max(x_1, y_1) \leq h$$

かといえ、

$$\max(x_2, y_2) \leq \max(x_1, y_1) \leq h$$

$$\therefore \max(x_1, y_1) \sqsubseteq \max(x_2, y_2)$$

B. $x_2 \leq x_1 \leq h \wedge h \leq y_1 \leq y_2$ の場合

$$\max(x_1, y_1) = y_1$$

$$\max(x_2, y_2) = y_2$$

であるから、

$$h \leq y_1 = \max(x_1, y_1) \leq y_2 = \max(x_2, y_2)$$

$$\therefore \max(x_1, y_1) \sqsubseteq \max(x_2, y_2)$$

C. $h \leq x_1 \leq x_2 \wedge y_2 \leq y_1 \leq h$ の場合

$$\max(x_1, y_1) = x_1$$

$$\max(x_2, y_2) = x_2$$

であるから、

$$h \leq x_1 = \max(x_1, y_1) \leq x_2 = \max(x_2, y_2)$$

$$\therefore \max(x_1, y_1) \sqsubseteq \max(x_2, y_2)$$

D. $h \leq x_1 \leq x_2 \wedge h \leq y_1 \leq y_2$ の場合

D.1 $\max(x_1, y_1) = x_1$ の場合

$$h \leq x_1 = \max(x_1, y_1) \leq x_2 \leq \max(x_2, y_2)$$

を得て、

$$\max(x_1, y_1) \sqsubseteq \max(x_2, y_2)$$

D.2 $\max(x_1, y_1) = y_1$ の場合

$$h \leq y_1 = \max(x_1, y_1) \leq y_2 \leq \max(x_2, y_2)$$

を得て、

$$\max(x_1, y_1) \sqsubseteq \max(x_2, y_2)$$

以上の A, B, C, D より

$$\max(x_1, y_1) \sqsubseteq \max(x_2, y_2)$$

が示された。

ii の証明

E. $x_2 \leq x_1 \leq h \wedge y_2 \leq y_1 \leq h$ の場合

E.1 $\min(x_1, y_1) = x_1$ の場合

$$x_2 \leq x_1 = \min(x_1, y_1) \leq h$$

$$\min(x_2, y_2) \leq x_2$$

を得て,

$$\min(x_2, y_2) \leq x_2 \leq \min(x_1, y_1) \leq h$$

$$\therefore \min(x_1, y_1) \sqsubseteq \min(x_2, y_2)$$

E.2 $\min(x_1, y_1) = y_1$

$$\min(x_2, y_2) \leq y_2 \leq y_1 = \min(x_1, y_1) \leq h$$

$$\therefore \min(x_1, y_1) \sqsubseteq \min(x_2, y_2)$$

F. $x_2 \leq x_1 \leq h \wedge h \leq y_1 \leq y_2$ の場合

$$\min(x_1, y_1) = x_1$$

$$\min(x_2, y_2) = x_2$$

を得て,

$$\min(x_2, y_2) = x_2 = x_1 = \min(x_1, y_1) \leq h$$

$$\therefore \min(x_1, y_1) \sqsubseteq \min(x_2, y_2)$$

G. $h \leq x_1 \leq x_2 \wedge y_2 \leq y_1 \leq h$ の場合

$$\min(x_1, y_1) = y_1$$

$$\min(x_2, y_2) = y_2$$

を得て,

$$\min(x_2, y_2) = y_2 \leq y_1 = \min(x_1, y_1) \leq h$$

$$\therefore \min(x_1, y_1) \sqsubseteq \min(x_2, y_2)$$

H. $h \leq x_1 \leq x_2 \wedge h \leq y_1 \leq y_2$ の場合

H.1 $\min(x_1, y_1) = x_1$ の場合

$$h \leq x_1 = \min(x_1, y_1) \leq x_2$$

$$h \leq x_1 = \min(x_1, y_1) \leq y_1 \leq y_2$$

を得て,

$$h \leq \min(x_1, y_1) \leq \min(x_2, y_2)$$

$$\therefore \min(x_1, y_1) \sqsubseteq \min(x_2, y_2)$$

H.2 $\min(x_1, y_1) = y_1$ の場合

$$h \leq \min(x_1, y_1) \leq x_1 \leq x_2$$

$$h \leq y_1 = \min(x_1, y_1) \leq y_2$$

を得て,

$$h \leq \min(x_1, y_1) \leq \min(x_2, y_2)$$

$$\therefore \min(x_1, y_1) \sqsubseteq \min(x_2, y_2)$$

以上の E, F, G, H より

$$\min(x_1, y_1) \sqsubseteq \min(x_2, y_2)$$

が示された。 □

(命題 3. 4 の証明)

i の証明

$x \sqsubseteq x$ であるから, $x \sqsubseteq y$ であれば, 命題 3. 3 の i より

$$x = \max(x, x) \sqsubseteq \max(x, y)$$

が得られた。

ii の証明

$x \sqsubseteq x$ が成り立っているから, $x \sqsubseteq y$ であれば, 命題 3. 3 の ii を適用して,

$$x = \min(x, x) \sqsubseteq \min(x, y)$$

を得て, 示された。

iii の証明

$0 \leq x \leq 2^{-1}$ であれば $2^{-1} \leq 1-x \leq 1$ を得,

$$0 \leq x \leq 2^{-1} \leq 1-x \leq 1$$

を得る。また,

$2^{-1} \leq x \leq 1$ であれば $0 \leq 1-x \leq 2^{-1}$ を得,

$$0 \leq 1-x \leq 2^{-1} \leq x \leq 1$$

を得る。よって, $0 \leq x \leq 1$ について

$$2^{-1} \leq \max(x, 1-x) \leq 1$$

$$0 \leq \min(x, 1-x) \leq 2^{-1}$$

がいえ, 証明された。 □

さて,

$$x \sqcup y \equiv \max(x, y)$$

$$x \sqcap y \equiv \min(x, y)$$

$$\bar{x} \equiv \text{not}(x) \equiv 1-x$$

とおけば、 \sqcup , \sqcap , $-$ は各々、
 fuzzy 選言 (fuzzy disjunction)
 fuzzy 連言 (fuzzy conjunction)
 fuzzy 否定 (fuzzy negation)

といわれる演算となっている。

$0 \leq x \leq 2^{-1}$ であれば、 x に対する情報は近似的に偽 (false)

$2^{-1} \leq x \leq 1$ であれば、 x に対する情報は近似的に真 (true)

ということにすれば、命題 3. 4 の iii は

$x \sqcup \bar{x}$ は近似的に真であり、

$x \sqcap \bar{x}$ は似的に偽である。

ことを指摘している。この両事実は、通常の 2 値命題論理⁽¹⁾ (two-valued logic) では、 $x = 0$, $x = 1$ に対応する情報は各々近似的に偽、真を表わし、 $x \sqcup \bar{x} = 1$, $x \sqcap \bar{x} = 0$ であることに照応している。なお、 $x \sqsubseteq y$ であれば、式 (3.4) からわかるように、条件 $h=2^{-1}$ の下で、 x なる情報が近似的に偽あるいは真であれば、 y についても各々そうである、ということになる。

4 命題 3. 1 ~ 3. 4 の意味するところによれば、fuzzy 情報処理とは、

$$0 \leq x \leq 1, 0 \leq y \leq 1$$

として、

$$x \sqcup y, x \sqcap y, \bar{x}$$

という 3 種類の演算を繰り返して、初期値な x_0 から出発して、あいまいさを減少させる鎖

$$x_0 \sqsubseteq x_1 \sqsubseteq x_2 \sqsubseteq \dots$$

を得る過程であるといえよう^{*10}。この際、 x_i はある命題 A_i の真理値 (the truth-value of a formula A_i) を表わしているとして、精確には、 x_i は、 $T(A_i)$ と書く方がよい。

3. 2 fuzzy 導出原理

J. A Robinson (1965) の導出原理 (resolution principle) によれば、「最小単位となる命題である」基本命題^{*11} (primitive proposition) p , リテラル (literal)

$$q_1, q_2, \dots, q_m, r_1, r_2, \dots, r_n$$

に対し、二つの節 (clause) つまり親節 (parent clause)

$$C_1 = p \vee q_1 \vee q_2 \vee \dots \vee q_m$$

$$C_j = \sim p \vee r_1 \vee r_2 \vee \dots \vee r_n$$

に関し、その導出節 (resolvent clause)

$$C_{ij} = q_1 \vee q_2 \vee \dots \vee q_m \vee r_1 \vee r_2 \vee \dots \vee r_n$$

を考えると,

$$C_i = \text{真}, \forall C_j = \text{真ならば } C_{ij} = \text{真}$$

が成り立つ⁽¹¹⁾。fuzzy logic では, この導出原理は次の命題 3. 5 のごとく指摘される^{(10), (12)}。

命題 3. 5 (ファジィ導出原理, fuzzy resolution principle)

$$T(C_i) \sqcap T(C_j) > 2^{-1}$$

であれば,

$$T(C_i) \sqcap T(C_j) \leq T(C_{ij}) \leq T(C_i) \sqcup T(C_j)$$

命題 p が真でる割合 (真理値) $T(p)$ は不等式 $0 \leq T(p) \leq 1$ を満たすが, このとき,

$1 - T(p)$ が命題 p の fuzziness

である。ならば, 命題 3. 5 は次の事実を指摘している: fuzzy 導出原理においては, 二つの親節の連言の fuzziness が 2^{-1} より小であれば, その推論結果である導出節の fuzziness は更に減少する。□

3. 3 三種の推論規則

導出原理を推論規則 (inference rule) として用いる利点は公理系 (axiom system) を必要としないことである。このことを以下に三段論法⁽¹⁴⁾で示してみよう。

本節では, しきい値 h を常に

$$h = 2^{-1}$$

と選んで固定しておく。

modus ponens (分離法則) は

$$P \rightarrow Q = \sim P \vee Q$$

と定義され, これを

if P then Q

と書く⁽¹²⁾。2 命題 3. 2, 3. 3 から,

$$T(P_1) \sqsupseteq T(P_2) \text{ かつ } T(Q_1) \sqsupseteq T(Q_2)$$

であれば,

$$T(\text{if } P_1 \text{ then } Q_1) \sqsupseteq T(\text{if } P_2 \text{ then } Q_2)$$

が成り立ち, 分離法則を推論規則として用いて得る推論はあいまいを減少させることが知れるという事実にも注意しておこう。

if $(P \wedge (\text{if } P \text{ then } Q))$

then Q

(3. 7)

が肯定式 (modus ponendo ponens) といわれるものであり、三段論法 (syllogism) の特別なものである。

同様に、2 命題 3. 2, 3. 3 を適用して、

$$T(P_1) \sqsubseteq T(P_2) \text{ かつ } T(Q_1) \sqsubseteq T(Q_2)$$

であれば、

$$T(\text{if } (P_1 \wedge (\text{if } P_1 \text{ then } Q_1)) \text{ then } Q_1)$$

$$\sqsubseteq T(\text{if } (P_2 \wedge (\text{if } P_2 \text{ then } Q_2)) \text{ then } Q_2)$$

が成立し、

あいまいの少ない、命題を持つ情報を用いる方があいまいさの少ない推論が可能であることを示している。

一般に、2 命題 3. 2, 3. 3 より主張可能な事実は次の通りである：

正しい一般的な 2 値論理命題から正しい特殊な 2 値論理命題を導く演繹的推論 (deductive reasoning) を fuzzy 推論で考えれば、必ず、半順序 \sqsubseteq の増大列が得られる様に、つまり、あいまいさが減少する形で情報処理がなされている。 \square

命題 3. 5 を肯定式 (3. 7) に適用すれば、

$$C_i = P$$

$$C_j = \sim P \vee Q$$

$$C_{ij} = Q$$

である場合の導出原理に対応するから、

$$\min(T(P), \max(1 - T(P), T(Q))) > 2^{-1}$$

であれば、

$$\min(T(P), \max(1 - T(P), T(Q))) \leq T(Q) \leq \max(T(P), \max(1 - T(P), T(Q)))$$

が成り立つことが知れる。

上述の肯定式の他に、前提 (premise) となる事実や仮定からある結論 (conclusion) を導き出す基本的な推論形式として、

(a) 否定式 (modus tollend tollens)

$$\text{if } ((\text{if } p \text{ then } Q) \wedge \sim Q) \text{ then } \sim P$$

(b) 三段論法 (syllogism)

$$\text{if } (\text{if } P \text{ then } Q) \wedge (\text{if } Q \text{ then } R) \text{ then } (\text{if } P \text{ then } R)$$

があるが、同様であるから割愛する。

3.4 fuzzy 集合

全体集合 U が

$$U = \{u_1, u_2, \dots, u_n\}$$

の場合, fuzzy 集合 F は

$$T_F(u_1), T_F(u_2), \dots, T_F(u_n)$$

を指定することで表現される。ここに,

$T_F(u_i)$ は $u_i \in V$ が F に帰属している割合 (帰属度) を表現している。

とし,

$$0 \leq T_F(u_i) \leq 1, \quad i = 1 \sim n$$

とする。勿論, 2 値論理体系とは異なり,

$$\sum_{i=1}^n T_F(u_i) = 1$$

である必要はない。

(i) $T_F(u_i) > 2^{-1}$ のとき, U 内の要素 u_i は fuzzy 的に F に帰属している

(ii) $T_F(u_i) < 2^{-1}$ のとき, U 内の要素 u_i は fuzzy 的に F に帰属していない

という。このとき, fuzzy 集合 F を

$$\begin{aligned} F &= T(u_1)/u_1 + T(u_2)/u_2 + \dots + T(u_n)/u_n \\ &= \sum_{i=1}^n T(u_i)/u_i \end{aligned}$$

と表記することがある。ここに, $/$ はセパレータ (分離記号, separator), $+$ は離接 (選言) の意である。例えば,

$$U = \{1, 2, \dots, n\}$$

として,

F is a small integer

という fuzzy 集合 (fuzzy 概念) F は

$$T_F(u_1 = 1) = 1, \quad T_F(u_2 = 2) = 1, \quad T_F(u_3 = 3) = 0.8, \quad \dots$$

と指定することで規定され,

F is a very small integer

という fuzzy 集合は

$$T_F(u_1 = 1) = 1, \quad T_F(u_2 = 2) = 1, \quad T_F(u_3 = 3) = 0.64, \quad \dots$$

と規定されてよい。

Fuzzy 推論とは, 与えられた fuzzy 集合

$$F = \sum_{i=1}^n T_F(u_i)/u_i$$

から、演繹的推論を繰り返さない、
 $F_0 = F$ として、あいまいさの減少例（鎖）

$$F_0 \sqsubseteq F_1 \sqsubseteq \dots \sqsubseteq F_i \sqsubseteq F_{i+1} \sqsubseteq \dots$$

を得、

$$x = \sqcup \{F_0, F_1, \dots, F_i, F_{i+1}, \dots\}$$

を解とすることであるといえよう。ここに、

$$G = \sum_{i=1}^n T_G(u_i) / u_i$$

として、

$$F \sqsubseteq G \text{ とは}$$

$$\forall i (= 1 \sim n), T_F(u_i) \sqsubseteq T_G(u_i), \text{ つまり}$$

$$T_G(u_i) \leq T_F(u_i) \leq 2^{-1} \vee 2^{-1} \leq T_F(u_i) \leq T_G(u_i)$$

と定義される。

3. 5 半順序での上限、下限による表現

2値論理系では、命題 p の意味が判明しなければ、通常その真偽が確定しないから、 p の真理値が p の意味に相当する。ただし、この際、 p を真にする状況がすべて思い浮かぶことが必要とされる。この立場を推し進めれば、fuzzy 論理系では、

$$T_F(u_i) > 2^{-1} \text{ を満足する } u_i \text{ の集合}$$

が fuzzy 概念 F の意味を規定しているといえよう。fuzzy 論理が2値論理よりすぐれているのは、概念 F を真にする状況 u_i の集まりを陽に指定していることだと、著者には思える。

本節では前節と同様、しきい値 h を $h = 2^{-1}$ と選んでおくものとする。

$$2^{-1} \leq x \sqcup \bar{x} \leq 1$$

が成立し、また、

$$0 \leq x \sqcap \bar{x} \leq 2^{-1}$$

が成立する（命題3. 4のiiiを参照）。更に、

$$x_1 \sqsubseteq x_2 \text{ かつ } y_1 \sqsubseteq y_2$$

であれば、

$$(イ) \quad x_1 \sqcup y_1 \sqsubseteq x_2 \sqcup y_2$$

$$(ロ) \quad x_1 \sqcap y_1 \sqsubseteq x_2 \sqcap y_2$$

$$(ハ) \quad \bar{x}_1 \sqsubseteq \bar{x}_2 \text{ かつ } y_1 \sqsubseteq y_2$$

が成り立ち (2 命題 3. 2, 3. 3 を参照), 誠に都合が良い。

$$x \sqcup y$$

を二つの元 x, y から成る集合 $\{x, y\}$ の上限 (supremum), または最小上界 (least upper bound) といい,

$$x \sqcap y$$

を集合 $\{x, y\}$ の下限 (infimum), または最大下界 (greatest lower bound) ということにする^{*12},

$x \sqcup y$: x と y とをあわせた情報をもったデータ

$x \sqcap y$: x と y とに共通な情報をもったデータ

という解釈が可能になる。

集合 A の上限, 下限は各々, $\sqcup A, \sqcap A$ とも書かれる。特に, A が $A = \{x, y\}$ の場合

$$\sqcup \{x, y\} = x \sqcup y$$

$$\sqcap \{x, y\} = x \sqcap y$$

と書いていた訳である。半順序 \sqsubseteq が導入されたデータ集合 D はその任意の部分集合が上限をもつとき, 完備束と称されるが (備考 2. 2 を参照), 完備束 D について, \emptyset (空集合), D (全体集合) は D の部分集合であるから

$$\perp = \sqcup \emptyset$$

$$\top = \sqcup D$$

が定義でき,

\perp : 無定義要素 (undefined element)

\top : 過剰定義要素 (overdefined element)

と呼ぶことができよう。 \perp, \top については

\perp : 全く情報をもっていないデータ

\top : 情報をもち過ぎて矛盾した情報をもつデータ

$x \sqcup y = \top$: データ x とデータ y とは互いに両立し得ない情報をもつ

という解釈⁽²⁾が可能になる。例えば, しきい値 h を $h = 2^{-1}$ と選んでいる fuzzy 推論では, 式 (3.4) で定義される fuzzy 半順序関係 \sqsubseteq について,

$$\perp \sqsubseteq \frac{1}{2} \sqsubseteq x \sqsubseteq 0 \sqsubseteq \top \text{ if } 0 \leq x \leq \frac{1}{2}$$

$$\perp \sqsubseteq \frac{1}{2} \sqsubseteq x \sqsubseteq 1 \sqsubseteq \top \text{ if } \frac{1}{2} \leq x \leq 1$$

が成り立つ。

4. ニューラルネット情報処理

約50兆の細胞から成る人間の精神は数百億の脳神経細胞で作られた階層構造をもつシステム (大脳, 小脳, 脳幹, 脊髄) がうちだす「情報」である^{*13}といわれている⁽¹⁵⁾。

脳は動物が動くために必要なるものとして、進化して来たという。その証拠に植物には脳がないという。最初に神経という電線（軸索，アクソン）が作られ、電線と化した神経は体内のところどころに数万と集まって、「神経節」（ガングリオン）という小型の脳を作る。タコやイカ、昆虫などでは、脳は体のあちこちに分散しているという。そしてさらに進化し、魚類からはじまる脊椎動物では、分散していた脳は背面から頭部に一極集中し、数億の神経が集まって、動物を運動・行動させる脳コンピュータとして動き出すという⁽¹⁵⁾。

例えば、大脳辺縁系のなかには、記憶・学習の担い手である海馬、本能的な「認知能力」を発揮する根底の脳としての扁桃核、行動力を出す側坐核などがある。

「知」は大脳新皮質の前頭葉と側頭葉によって醸成されたものが前頭連合野から、人間の「知能」として創出されるものであるという⁽¹⁵⁾。

被覆のない裸伝線の神経（無髄神経）と、絶縁被覆が巻きついているため、前者に比べ伝達効率が約100倍である神経（有髄神経）との二種類がある。人間の脳はもともと無髄神経を主役とするアナログ型コンピュータであり、その一部が有髄神経に進化して、デジタル型コンピュータをつけ加えた構造になっているという⁽¹⁵⁾。

アナログ型の脳である脳幹の上位にデジタル型の小脳と大脳がかぶさっており、人間の脳では進化した部分ほどデジタル化が進んでいる⁽¹⁵⁾。神経の接続部はシナプス（synapse）と呼ばれるが、神経の情報は神経伝達物質のもたらず電気的な強弱の信号として軸索の表面を伝わる。

脳にデジタル型化した部分があることが、人間が言語をもてるようになった最大の理由であるという。また、大脳新皮質の前頭連合野とその近傍の部分だけ、自らが放出した神経伝達物質を自分で再回収するオートレセプター（自己受容体）が欠落しており、情報が一方向だけに流れ、決して逆もどりはしない。ここに、人間精神の創造力の鍵があるという。つまり、受け手がないから、創造できるという⁽¹⁵⁾。

以上の如くその一部が解明されている脳の構造や機能を参考にして、多数の計算要素（神経細胞，ニューロン；neuron）の相互作用によって情報処理を行うシステムがニューラルネット（neural network）あるいはニューロ・コンピュータ（neuro computer）である⁽³⁵⁾。

4. 1 相互結合形ネットの2値モデル

n 個のニューロンが互いに結合しているネット、いわゆる相互結合形ネット（highly-interconnected networks, mutual connected networks）を考え、

$x_i(t)$ ：時刻 t での第 i ($= 1 \sim n$) ニューロンの出力

$u_i(t)$ ：時刻 t での第 i ニューロンの平均膜電位

w_{ij} ：第 i ニューロンから第 j ニューロンへのシナプス結合（synaptic connection）の実数値
重み

h_i ：第 i ニューロンのしきい値（実数）

を導入する。

$$\text{psn}(u) = 1 \text{ if } u \geq 0, = 0 \text{ if } u < 0$$

という2値の正符号関数（the positive sign function）を導入し、各ニューロンは2値情報処理素子（two-state processing element）を想定すると、このネットは binary model と称せられ、その動作方程式は、

$$u_i(t + t) = \sum_{j=1}^n w_{ij} \cdot x_j(t) - h_i, \quad i = 1 \sim n$$

ここに,

$$x_j(t) = \text{psn}(u_j(t))$$

となる。

$$\mathbf{x} = (x_i)_{1 \leq i \leq n}$$

とおき、時刻 t での、このネットのエネルギー $E(t) = E(t/\mathbf{x})$ を

$$E(t) = E(t/\mathbf{x}) = -2^{-1} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \cdot x_i(t) \cdot x_j(t) + \sum_{i=1}^n h_i \cdot x_i(t)$$

と定義しよう。そうすると、以下のエネルギーの非増加定理が成り立ち、 n 個のニューロンから成るニューロン集団のエネルギー $E(t)$ が時刻 t の進展と共に減少していくことがわかる。

[定理 4. 1] ⁽¹⁶⁾ (エネルギーの非増加定理) 条件

$$w_{ij} = w_{ji}, \quad i, j = 1 \sim n \quad (\text{結合の対称性})$$

$$w_{ij} = 0 \quad (\text{自己回帰結合なし}), \quad i = 1 \sim n$$

の下で、ある一つのニューロン出力 x_k のみ注目し、遷移

$$x_k(t) \in \{0, 1\} \rightarrow x_k(t + \Delta t) \in \{0, 1\}$$

を考える。ただし、

$$\forall j \in \{1, 2, \dots, n\} - \{k\}, \quad x_j(t + \Delta t) = x_j(t)$$

とする。このとき、

$$E(t) \geq E(t + \Delta t) \quad \square$$

さて、二つの出力 $\mathbf{x}(t)$, $\mathbf{x}(s)$ 間の半順序関係 \sqsubseteq を

$$\mathbf{x}(t) \sqsubseteq \mathbf{x}(s)$$

$$\Leftrightarrow E(t) \geq E(s)$$

と定義でき、超大規模集積回路の配置問題や巡回セールスマンの問題などの組合せ最適化問題の解を得たい場合、目的関数をエネルギー E で等価的に表現できるように、

$$\text{重み } W = (w_{ij})_{1 \leq i, j \leq n}$$

$$\text{しきい値 } h = (h_i)_{1 \leq i \leq n}$$

を適切に設定しておけば、この種の問題の解は、

$$\sqcup_m \{ \mathbf{x}(k \cdot \Delta t) \mid k = 1 \sim m \}$$

と与えられることになる。

4. 2 相互結合形ネットの連続モデル

前節 4. 1 とは異なり、各ニューロンが 0 から 1 までの連続値をとる相互結合形の連続モデル (continuous model) を考え、そのシステム方程式が

$$\tau \cdot \frac{d}{dt} u_i(t) = -u_i(t) + \sum_{j=1}^n w_{ij} \cdot x_j(t) - h_i, \quad i = 1 \sim n$$

ここに、

$$x_i = f(u_i(t))$$

$$\tau: \text{時定数 } (> 0)$$

であるとしよう。ただし、結合の対称性

$$w_{ij} = w_{ji}, \quad i, j = 1 \sim n$$

を仮定しておく。そして、時刻 t でのエネルギー $E(t) = E(t/x)$ を

$$E(t) = E(t/x) = -2^{-1} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \cdot x_i(t) \cdot x_j(t) + \sum_{i=1}^n h_i \cdot x_i(t) + \sum_{i=1}^n g(x_i(t)) \quad (4. 1)$$

と定義しよう。

このとき、次の定理 4. 2 が成立し、エネルギー $E(t)$ はやはり、時刻 t の非増加関数となり、前節と同様に、

ニューラルネットの、半順序による情報処理の働きを説明できる。

[定理 4. 2] (エネルギー E の減少定理)

$f(u)$ は一変数 u の単調増加関数

とし、

$$g(x) = \int^x dy f^{-1}(y)$$

とおくと、

$$(i) (\partial/\partial x_k) E = -\tau (d/dt) u_k$$

$$(ii) \tau \cdot \frac{d}{dt} E = -\sum_{k=1}^n [\tau \cdot \frac{d}{dt} u_k]^2 \frac{d}{du} f(u)|_{u=u_k} \leq 0 \quad \square$$

関数 $f(u)$ の例を揚げておこう⁽¹⁷⁾。

細胞発火関数と呼ばれる一実変数 u の関数 $f(u)$ をシグモイド関数

$$f(u) = \frac{1}{1 + e^{-cu}}, \quad \text{where } c > 0$$

を採ると、

$$(d/du) f(u) = C \cdot f(u) [1 - f(u)] \geq 0$$

$$f^{-1}(y) = -C^{-1} \log[(1 - y)/y]$$

が成り立つ。また、 $g(x)$ を

$$g(x) = \int_{\frac{1}{2}}^x dy f^{-1}(y)$$

とおくと、エントロピー関数

$$H(x) = -x \log x - (1-x) \log(1-x)$$

を用いて、

$$g(x) = C^{-1} \cdot [\log 2 - H(x)] \quad (4.2)$$

と表現できることに注意する。また、

正定数 C が大きい、小さいことは各々、ニューロン間の結合の強さが大きい、小さいことに対応する。

$$0 \leq x_i(t) \leq 1$$

であるから、第 i ニューロン出力 $x_i(t)$ はある第 i 命題 Q_i が真である程度 $T(Q_i)$ と考えることができ (第3章を参照)、ニューラルネットは fuzzy 情報処理 をしていると想定できなくはない。このとき、上述の $g(x)$ の具体形である式 (4.2) を見てもわかるように、式 (4.1) のエネルギー $E(t)$ の中に、エントロピー

$$\begin{aligned} & \sum_{i=1}^n g(x_i(t)) \\ &= \sum_{i=1}^n C^{-1} \cdot [\log 2 - H(x_i)] \end{aligned}$$

ここに、

$$0 \leq H(x_i) \leq \log 2$$

$$x_i(t) = 2^{-1} \Leftrightarrow H(x_i) = \log 2$$

$$x_i(t) \in \{0, 1\} \Leftrightarrow H(x_i) = 0$$

が含まれることになったことをも考えあわせれば興味あることである。

4.3 前進形の高層ネットと学習の働き

feedback 結合のない、 m 個の層から成るニューラルネット、いいかえれば、

$k = 1$ の層が入力層 (input layer)

$k = 2 \sim (m-1)$ の層が隠れ層 (hidden layers)

$k = m$ の層が出力層 (output layer)

であるような m 個の層から成る

各層内ではニューロン間の結合がないような前進形の高層ネット (multi-layer feedforward network)

を考える。細胞発火関数 $f(u)$ を導入し、システム方程式

$$x_j^k = f(u_j^k), \quad j = 1 \sim n(k)$$

$$u_j^k = \sum_{i=1}^{n(k-1)} W_{ij}^{k-1} \cdot x_i^{k-1}$$

で表現されるニューラルネットを考えよう。

ここに、

u_j^k : 第 k ($= 1 \sim m$) 層内第 j ($= 1 \sim n(k)$) ニューロンへの実数入力 of 総和

x_j^k : 第 k 層内第 j ニューロンの出力

であり、第 k 層は $n(k)$ 個のニューロンから成っているとしている。

一般に、学習 (learning) とは構造的な知識の獲得を意味する⁽²³⁾。本節では、学習過程 (learning process) とは、ニューラルネット構造上の知識を得るため、つまり、システム内のパラメータ集合

$$W = \{W_{ij}^{k+1} \mid k = 1 \sim m - 1, i = 1 \sim n(k), j = 1 \sim n(k+1)\}$$

を決定するため、入力とそれに対応する希望出力の対から成る訓練例の集合 (a set of training examples) を用い、時刻 t での重み $W_{ij}^{k+1}(t)$ を

$$W_{ij}^{k+1}(t + \Delta t) = W_{ij}^{k+1}(t) + \Delta W_{ij}^{k+1}(t)$$

へと変更する過程とみなそう。

$$s = \{s_i \mid i = 1 \sim n(1)\}$$

が第 1 層に入力されたとき、つまり、

$$u_j^1 = s_j, \quad j = 1 \sim n(1)$$

としたとき、第 m 層の現実出力 (actual output) は、 W, s の関数として、

$$x_j^m(W, s), \quad j = 1 \sim n(m)$$

と書かれるが、この入力 s に対応する第 m 層の希望出力 (desired output) が

$$y = \{y_j \mid j = 1 \sim n(m)\}$$

であるとき、自乗誤差 (square error between the actual outputs and the desired outputs) E は、 W, s, y の関数であり、

$$\begin{aligned} E &= E(W) = E(W, \langle s, y \rangle) \\ &= \sum_{j=1}^{n(m)} (x_j^m(W, s) - y_j)^2 \end{aligned}$$

と表わされる。二つの重みに W, W' に関し、

$$\begin{aligned} W &\sqsubseteq W' \\ \Leftrightarrow E(W) &\geq E(W') \end{aligned}$$

と定義すると、2 元関数 \sqsubseteq は重み W の集合上の半順序関係である。

適当な初期値 $W^{(0)}$ から出発し、上限

$$W = \sqcup \{W^{(0)}, W^{(1)}, W^{(2)}, \dots\}$$

を求める過程が、自乗誤差 E を極小にする重み W を求めるという意味で、入力・希望出力の対

の訓練例 $\langle s, y \rangle$ の集合から知識を得る学習過程である。

現時点 t で得られている $W(t)$ から、学習の成立条件

$$W(t) \subseteq W(t + \Delta t)$$

を満たす $W(t + \Delta t)$ を求める方法は誤差逆伝播学習アルゴリズム⁽¹⁸⁾ (error back propagation learning algorithm) として知られている*14。

最急降下法 (gradient descent method) によれば、正数 ε を十分小さく選んで、

$$\Delta W_{k-1}^k(t) = -\varepsilon \cdot \frac{\partial E(w(t))}{\partial w_{k-1}^k(t)}$$

とすればよいことがわかっている。計算結果は次の通りである：

$$\Delta W_{k-1}^k(t) = -\varepsilon \cdot d_k^k \cdot x_{k-1}^{k-1}$$

ここに、

$$d_j^m = 2 \cdot (x_j^m(W(t), s) - y_j) \cdot \left. \frac{dg(u)}{du} \right|_{u=u_j^m}$$

であり、

$k = 1 \sim m - 1$ に対し、

$$d_k^k = \left(\sum_{j=1}^{n_{k+1}} W_{jk}^{k+1}(t) \cdot d_j^{k+1} \right) \cdot \left. \frac{dg(u)}{du} \right|_{u=u_k^k}$$

□

一般化された誤差 d_k^{k+1} が出力層→隠れ層→入力層という様に、情報処理出力を得る方向と逆の方向に伝播され、重み $W_{k-1}^k(t)$ が更新されて行くことが上述の公式から理解できよう。誤差逆伝播という名の所以である。

なお、相互結合形のネットは、エネルギー関数の各極小点を与えるニューロン出力の組を記憶内容とすると、通常の digital computer での、正確なアドレスを与えて記憶内容を引き出す方式と異なり、

不正確であってもよい「情報」(内容番地) を与えて記憶内容を引き出す「内容番地記憶 (content adressable memory)」

としての一種の連想記憶^{(20), (34)} (associative memory) を行なうニューロ・コンピュータとして利用されることが多いが、上述の誤差逆伝播学習形ニューラルネットはこの様な連想計算 (associative computing) を行なう連想子 (associator) として利用されること以外に、ある写像の近似実現として利用されることが多い。

5. 記号情報処理

計算機による情報処理はその意味をもたせた記号の組合せの形式的表現であるプログラムの働きでなされ、特に、人間の行なっている情報処理プロセスを表現しようとする人工知能プログラムは記号による推論を可能にしている。プログラムの働きとは記号の書き換えであり、記号のこ

のような書き換えによる処理はデータ処理の典型的なものである。また、情報はデータの上ののっており、データの中にかくされているという指摘⁽²⁴⁾は情報とデータとの関係を明らかにしているともいえよう。

5.1 プログラムによる情報処理

λ 記法 (lambda notation) では、 x を A 上を働く変数、 $E(x)$ を変数 x を含むかも知れない集合 B の要素を表現する式とすると、

$$\lambda x \in A. E(x)$$

は

$$f(a) = E(a) \in B \quad (a \in A)$$

であるような関数

$$f: A \rightarrow B$$

を表現する。例えば、

$$f(a) = \sqrt{a}, \quad a \in [0, \infty) = \{x \mid 0 \leq x < \infty\}$$

という平方根関数は

$$\lambda x \in [0, \infty). \sqrt{x}$$

と書かれる。プログラム言語としての λ 論理での λ 言語⁽²⁾ (λ language) について説明する余裕はないが、 λ 言語で書かれた次のプログラム f は、 x を入力変数とし、

$x = 0$ の場合は値 0

$x \neq 0$ の場合は値 $(2x - 1) + f(x - 1)$

を出力する関数型の再帰プログラム (functional recursive program) である。なお、通常の流れ図プログラム (flowchart program) を再帰プログラムに翻訳できることがプログラム理論により知られている⁽²⁾。:

$$N = \{0, 1, 2, \dots\}$$

とし、条件関数 (conditional function)

$$\text{cond}(b, x, y) = \begin{cases} x & \text{if } b = \text{true} \\ y & \text{if } b = \text{false} \end{cases}$$

を

if b then x else y fi

と書いて

$$f(x) = \lambda x \in \mathbb{N}. \text{ if } x = 0 \text{ then } 0 \text{ else } (2x - 1) + f(x - 1) \text{ fi} \quad (5. 1)$$

□

このプログラム $f(x)$ が入力変数 x に対し、何を計算しているのか、その具体的な解を求めることを考えよう。

$$x = \perp \text{ (無定義要素)}$$

は

x はいつも値をもたなくて、 x は定義されない
を意味するとして、2 元関係

$$x \sqsubseteq y$$

を

$$x = \perp \text{ または } x = y$$

と定義すれば、2 章の反射律イ、反対称律ロ、推移律ハを満たし、 \sqsubseteq は半順序関係である。また、関係 f, g に対し、

$$f \cong g \Leftrightarrow \forall x, f(x) \sqsubseteq g(x)$$

と定義すると、この 2 元関係 \cong も反射律、反対称律、推移律を満たし、半順序関係である。

$F(g)$

$$= \lambda g. \lambda x.$$

$$\text{if } x = 0 \text{ then } 0 \\ \text{else } (2x - 1) + g(x - 1) \\ \text{fi}$$

という汎関数 (functional) を導入して、

$$f_0 = \perp$$

として、

$$f_{k+1} = F(f_k) \\ = \lambda x. \\ \text{if } x = 0 \text{ then } 0 \\ \text{else } (2x - 1) + f_k(x - 1) \\ \text{fi}$$

を求めていけば、

$$f_0 \cong f_1 \cong f_2 \cong \dots$$

が成り立ち、式 (5. 1) のプログラム f は再帰的方程式 (recursive equation)

$$F(f) = f$$

を満たすから、もし

$$F(h) = h$$

を満たす他の解 h があれば、

$$f \sqsubseteq h \wedge F(f) = f$$

を満たす最小の f (最小不動点, least fixed point) を解とみなせば、備考 2. 2 の最小不動点定理より、解 f は

$$f = \sqcup \{f_0, f_1, f_2, \dots\}$$

と上限を求める形で与えられることがわかる。実際に求めてみると、

$$f_0 = \perp$$

$$f_1 =$$

$$\lambda x.$$

$$\text{if } x = 0 \text{ then } 0$$

$$\text{else } (2x - 1) + f_0(x - 1)$$

$$\text{fi}$$

$$= \lambda x.$$

$$\text{if } x = 0 \text{ then } 0 \text{ else } \perp \text{ fi}$$

$$f_2 = \lambda x.$$

$$\text{if } x = 0 \text{ then } 0$$

$$\text{else } (2x - 1) + f_1(x - 1) \text{ fi}$$

$$= \lambda x.$$

$$\text{if } x = 0 \text{ then } 0$$

$$\text{else}$$

$$\text{if } x = 1 \text{ then } (2 \cdot 1 - x) + f_1(1 - 1)$$

$$\text{else } \perp \text{ fi}$$

$$\text{fi}$$

$$= \lambda x.$$

$$\text{if } x = 0 \text{ then } 0$$

$$\text{else}$$

$$\text{if } x = 1 \text{ then } 1$$

$$\text{else } \perp$$

$$\text{fi}$$

$$\text{fi}$$

$$f_3 = \lambda x.$$

```

if x = 0 then 0
else
  if x = 1 then 1
  else
    if x = 2 then 4
    else ⊥
  fi
fi
...

```

であることが示され、よって

$$f = \sqcup \{f_0, f_1, f_2, \dots\}$$

は

$$f(x) = \lambda x \cdot x^2$$

であることがわかる。これは

$$\begin{aligned}
& (2x - 1) + f(x - 1) \\
&= (2x - 1) + (x - 1)^2 \\
&= x^2 \\
&= f(x)
\end{aligned}$$

であることから、その正当性が了解できよう。

f の第 n 近似プログラム f_n は

$$x \in \{0, 1, 2, \dots, n - 1\} \text{ なる入力 } x \text{ に対しては } f_n(x) = x^2$$

という正確な出力をもたらすが、

$$x \in \{n, n + 1, n + 2, \dots\} \text{ なる入力 } x \text{ に対しては停止しない}$$

プログラムとなっていることに注意しておこう。

5. 2 演繹的推論の一種である導出原理による情報処理

推論にあたっては、諸概念、性質 (property)、カテゴリ (category)、関係 (relation) などを導入し、人間が取扱う種々の情報とその構造をもっとも基本的要素 (プリミティブな要素; primitive information) に分解することが必要とされる。人間が明確な概念として把握できる表現力をもつ構造⁽²⁴⁾を情報の型式 (information type) とそに上の操作クラスター (operation cluster) とで表現する抽象情報構造、いいかえれば、情報そのもの (単体; object) と、その利用手段としてのアクセス関数 (access function) の組はプリミティブな情報を意識せずには構築できないが、プリミティブ要素を組み合わせると、

2項関係モデル, 多項関係モデル, 2項/多項関係モデル, 階層モデル, ネットワークモデルなどの情報空間モデルが得られる。このようにして, 実世界を記述する概念スキーマ (conceptual schema) の枠組の中で推論がなされることになる。

以下, 公理系を必要としない記号的演繹推論の代表的なものとして, 導出原理を説明しよう (3.2節の内容と一部重複する)。この場合は, 情報の型式, 操作の形式として, 各々節形式, 導出原理を採用していることになる。

ある前提 (premise) からある結論 (conclusion) を導き出すことを一般に推論 (reasoning) という。推論には大きく分類して二種類ある。真なる一般的な言明 (statement) から真となる特殊な言明 (事実) を導き出すのが演繹的推論であり, 逆に真となる特殊な言明 (事実) から真となる可能性を秘めている一般的な言明を導き出すのが帰納的推論 (inductive reasoning) である。本節では演繹的推論を, 次節 5.3 では帰納的推論を半順序の立場から説明しよう。

第1階述語論理 (first order predicate logic) での任意の述語論理形式 (formula of predicate logic) からスコーレム標準型 (skolem standard form)

$$\forall x_1 \forall x_2 \cdots \forall x_m [C_1 \wedge C_2 \wedge \cdots \wedge C_m] \quad (5.2)$$

という節形式 (clause form) を得ることができる。節形式は連言標準形 (conjunctive canonical form) ともいわれる。ここに, x_1, x_2, \dots, x_m は $C_1 \wedge C_2 \wedge \cdots \wedge C_m$ に登場する相異なるすべての変数であり, C_j はリテラル (literal) といわれる「素式 (基本命題) あるいはその否定形」の選言といわれる節。

式 (5.2) を

$$C = \{C_1, C_2, \dots, C_n\}$$

と書いて, 節集合 (clause set) という。節集合 C が真であるとは式 (5.2) からわかるように, その要素である節 C_1, C_2, \dots, C_n がすべて真であることであり, C が偽であるとは C_1, C_2, \dots, C_n の中の少なくとも一つが偽であることに注意しておく。

論理式の集合

$$\{P_1, P_2, \dots, P_n\}$$

と論理式 Q を考える。 P_1, P_2, \dots, P_n をすべて真とするような任意の解決において Q もまた真であるとき, 論理式の形式で

$$(P_1 \wedge P_2 \wedge \cdots \wedge P_n) \rightarrow Q \quad (5.3)$$

と書き, Q は P_1, P_2, \dots, P_n からの論理的帰結 (logical consequence) であるという。

論理式が充足不能 (unsatisfiable) であるとは, この論理式を真にするような解釈 (interpretation) は一つも存在しないことをいう。式 (5.3) が充足不能でないとき, つまり, 式 (5.3) を真にする解釈が少なくとも一つ存在するとき, 式 (5.3) の否定形は充足不能であり, 逆も成り立つ。

式 (5.3) の否定形

$$= \sim[\sim(P_1 \wedge P_2 \wedge \cdots \wedge P_n) \vee Q]$$

$$= P_1 \wedge P_2 \wedge \dots \wedge P_n \wedge \sim Q$$

であるから、 P_1, P_2, \dots, P_n, Q が節であれば、式 (5.3) の否定形は節形式である。この例からわかるように、節形式が充足不能であることが決定できれば、任意の論理式を真にする解釈が存在するか否かを決定できる^{*15}。いいかえれば、節形式の充足不能性を判定する手続きで、記号論理 (symbolic logic) での演繹的推論が可能になることを考慮し、以後、節形式の充足不能性を判定する手続きを考えよう。この様な手続きとして、導出原理がある。

空節は真にするどんな解釈も存在しないという意味で、恒偽式 (contradiction) であるとみなす。また、議論の対象となる個体 (individual) の集合を対象領域 (universe of discourse) というが、対象領域のすべての要素に対し、節集合 C が真となる対象領域と解釈が一つも存在しないとき^{*17}、この節集合 C は充足不能であるという。

導出原理

節集合

$$C = \{C_1, C_2, \dots, C_n\}$$

より適当な節 C_i と C_j を選び、これらを親節とする C_{ij} を求めて、新たな節集合

$$C \sqcup \{C_{ij}\}$$

を作る。 C_{ij} は

$$C_i = P \vee Q$$

$$C_j = \sim P \vee R$$

であれば、

$$C_{ij} = Q \vee R$$

と定義される^{*16}。

以下、この操作を繰り返していき、いずれかの時点で、空節 (リテラルを1個も含まない節) が導かれれば、節集合 C は充足不能であり、逆も成り立つ。□

上述の導出原理は命題論理で説明されたものであるが、変数 (個体変数) の集合の各要素に各々ある値を代入するという“代入操作” (単一化; unification) を導入すれば、述語論理での導出原理は結局は命題論理の導出原理に帰着する。

理解しやすくするためもあるが、以下、これ迄通り、命題論理で説明しよう。

T, F を各々、真 (true)、偽 (false) として、

$$T \sqsubseteq T \wedge T \sqsubseteq F \wedge F \not\sqsubseteq F \wedge F \not\sqsubseteq T$$

という半順序 \sqsubseteq を考える。 $\not\sqsubseteq$ は \sqsubseteq が成立したいことを表現している。このとき、

$$C_0 = C$$

として、上記の導出原理を次々と適用して、節集合の例

$$C_0, C_1, C_2, \dots, C_i, C_{i+1}, \dots, C_n$$

を求める。そうすれば、

$$C_0 \sqsubseteq C_1 \sqsubseteq C_2 \sqsubseteq \dots \sqsubseteq C_i \sqsubseteq C_{i+1} \sqsubseteq \dots \sqsubseteq C_n$$

が成立し、

$$C_n = F \wedge C_n \not\sqsubseteq C_{n+1}$$

という n が有限の値で存在すれば、

$$C_0 = C_1 = \dots = C_{n-1} = T \text{ であり、しかも}$$

C_n には空節が含まれており、

$$\sqcup \{C_0, C_1, \dots, C_n\} = C_n$$

が成立し、もとの節集合 $C = C_0$ は充足不能である。また、このような n が存在しなければ、もとの節集合は充足不能でない。

5. 3 帰納的推論の一種であるモデル推論

観測された事実から正しい知識を形成する段階で活用される重要な推論として帰納的推論があるが、E. Y. Shapiro による

「Inductive Inference of Theories from Facts」

、いわゆるモデル推論が代表的である。一般に、述語論理式は解釈によって真になったり偽になったりするが、与えられた論理式を真とするような解釈をその論理式のモデル (model) という。このようなモデルを推論するのがモデル推論である。

$$A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_m$$

をアトム (素式) とすると、節は

$$\sim A_1 \vee \sim A_2 \vee \dots \vee \sim A_n \vee B_1 \vee B_2 \vee \dots \vee B_m$$

$$\text{つまり } (B_1 \vee B_2 \vee \dots \vee B_m) \leftarrow A_1 \wedge A_2 \wedge \dots \wedge A_n$$

と表わされるが (これらに含まれている変数はすべて全称記号 \forall によって束縛されているものとする)、 m が高々 1 であるようなこのような節を、すなわち、正のリテラル (アトム) を高々一つしか含まない節をホーン節という。ちなみに負のリテラルとはアトムの否定形である。ホーン節は

$$B_1 \leftarrow A_1 \wedge A_2 \wedge \dots \wedge A_n \text{ (規則節)}$$

$$B_1 \leftarrow \text{ (事実節)}$$

$$\leftarrow A_1 \wedge A_2 \wedge \dots \wedge A_n \text{ (疑問節)}$$

のいずれかの形式を持っている。Prolog 言語で書かれたプログラムはホーン節の有限な集合である。

さて、shapiro の提唱したモデル推論手法を最汎アトムが使用されない形式で適用して、観測言語が一階言語 L のグランドアトム (個体変数を含まないアトム) であり、仮説言語が L のホーン又は (ホーン節) である場合、多数の入力・出力の対 (グランドアトムとその真偽とから

成る対) の例によって, Prolog 言語で書かれたプログラムを合成するシステムとその言語 C による実現については, 文献 (21), (22) に説明されているが, このモデル推論によれば, モデル M において true, false のいずれかであること*¹⁸が判明している観測文 α とその真偽値 V との対 (例文) である “事実” $\langle \alpha, V \rangle$ を次々と与え, 更に,

観測文 α が与えられると, もし α が M において真ならば “true”, そうでなければ “false” を返す様な “M に対する神託 (oracle)”

の存在を仮定すると,

真の観測文をすべて導出する “仮説言語によって記述される有限個の文の集合” を発見できる。

論理式の集合を真とする解釈を論理式のこの集合のモデル M と称するが,

M で真となる観測文の集合を L_M とし, M で真となる, 仮説言語で表現された論理式 (ホーン節) の集合 Σ があって,

$$\Sigma \vdash L_M \quad (\Sigma \text{ から } L_M \text{ が導き出される})$$

が成り立つ場合を考えよう。事実, $\langle \alpha_i, V_i \rangle$ の例

$$\langle \alpha_i, V_i \rangle, \quad i = 1, 2, \dots$$

が与えられた条件の下で,

$$\Sigma \vdash L_M$$

となる最小の集合 Σ を決定することを問題としよう。ここに, α_i はグランドアトムといわれるもので, V_i は α_i の真理値である。

$C_0 = \square$ (空節) として, 事実の有限部分列

$$\langle \alpha_j, V_j \rangle, \quad j = 1 \sim i + 1$$

について,

$$V_j = \text{真ならば}, \quad C_{i+1} \vdash \alpha_j \quad (\text{導出可能})$$

$$V_j = \text{偽ならば}, \quad C_{i+1} \not\vdash \alpha_j \quad (\text{導出不能})$$

が満たされるように, 精密化操作と矛盾点追跡操作^{(21), (22)}とを適用し, C_i から C_{i+1} を求める。この過程を,

$$C_{i+1} = f(C_i)$$

と書く。5. 2 節と異なり,

$$F \sqsubseteq F \wedge F \sqsubseteq T \wedge T \not\sqsubseteq T \wedge T \not\sqsubseteq F$$

という半順序関係 \sqsubseteq を導入する。

$$C_0 \sqsubseteq C_1 \sqsubseteq C_2 \sqsubseteq \dots \sqsubseteq C_n \not\sqsubseteq C_{n+1}$$

なる有限の n が存在すれば,

$$\sqcup \{C_0, C_1, \dots, C_n\} = C_n$$

が成立し、 C_n が求める \mathfrak{E} である。

6. パターン情報処理

一般に、人工知能学の応用としての知識工学では、共通な性質を有する対象の集合をクラス (class) と呼び、その一つ一つの対象をインスタンス (instance) と呼ぶ。パターン認識情報処理工学⁽²⁷⁾では、クラス、インスタンスを各々カテゴリ (category)、パターン (pattern) と呼び、パターンがどのカテゴリに帰属するかという決定を問題とする。パターンは印刷文字、手書き文字、人の発生した音声波形、言語音声などの総称である。

さて、鈴木昇一の「パターン認識の数学的理論⁽⁵⁾」では、パターン $\varphi \in \Phi$ が $j \in \gamma$ を添字にもつカテゴリ \mathfrak{C}_j のいずれかに帰属しているという知識を φ, γ の対である

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

で表わす。全カテゴリ集合を

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

とする。カテゴリ番号の集合 J の部分集合から成る集合を 2^J と表記すると、 $\gamma \in 2^J$ は候補カテゴリ番号をリスト形式で表現したものである。

パターン認識の数学的理論⁽⁵⁾では*19, 与えられた入力パターン $\varphi \in \Phi$ に対し、

$$\varphi_0 = \varphi, \gamma_0 = J \tag{6. 1}$$

として、鎖

$$\begin{aligned} \langle \varphi_0, \gamma_0 \rangle &\sqsubseteq \langle \varphi_1, \gamma_1 \rangle \sqsubseteq \dots \\ &\sqsubseteq \langle \varphi_n, \gamma_n \rangle \sqsubseteq \langle \varphi_{n+1}, \gamma_{n+1} \rangle \sqsubseteq \dots \end{aligned} \tag{6. 2}$$

を作り、

$$\sqcup \{ \langle \varphi_0, \gamma_0 \rangle, \langle \varphi_1, \gamma_1 \rangle, \dots, \langle \varphi_n, \gamma_n \rangle, \dots \} \tag{6. 3}$$

を、入力パターン φ に関する認識情報処理結果とする。ここで、 $\langle \varphi_n, \gamma_n \rangle \in \langle \Phi, 2^J \rangle$ から $\langle \varphi_{n+1}, \gamma_{n+1} \rangle$ への変換を構造受精変換⁽⁵⁾

$$TA(\gamma_n)T$$

にとると、

$$\langle \varphi_{n+1}, \gamma_{n+1} \rangle = TA(\gamma_n)T \langle \varphi_n, \gamma_n \rangle$$

ここに、

$$\begin{aligned} \varphi_{n+1} &= TA(\gamma_n)T \varphi_n \\ \gamma_{n+1} &= \hat{\gamma}_n(\varphi_{n+1}) \end{aligned}$$

であり、ある一般的な条件の下で、

$$\begin{aligned} & \exists j \in J, \\ & \sqcup \{ \langle \varphi_0, \gamma_0 \rangle, \langle \varphi_1, \gamma_1 \rangle, \dots, \langle \varphi_n, \gamma_n \rangle, \dots \} \\ & = \langle \omega_j, [j] \rangle \end{aligned} \tag{6.4}$$

の成立が示される。 $\omega_j \in \Omega \subset \Phi$ は、第 $j \in J$ 番目のカテゴリ \mathcal{C}_j の特色を最もよく表わしているという意味で、 \mathcal{C}_j を代表している“代表パターン” (typical pattern) である。この ω_j は \mathcal{C}_j のもつ典型的な事柄に関する知識を備えているプロトタイプ (prototype) としてのパターンであり、また、 $[j]$ は唯一つのカテゴリ番号 $j \in J$ のみから成るリストである。式 (6.4) は

入力パターン φ はカテゴリ \mathcal{C}_j の代表パターンと呼ばれる ω_j に最終的に変形され、 φ はカテゴリ \mathcal{C}_j に帰属する

という具合に、認識推断されたことを示している。

説明不足を少し補っておこう。

$$\begin{aligned} & \langle \varphi, \gamma \rangle, \langle \eta, \mu \rangle \in \langle \Phi, 2^J \rangle \text{ についての等号の成立} \\ & \langle \varphi, \gamma \rangle = \langle \eta, \mu \rangle \end{aligned}$$

とは、次の3事柄 a, b, c が成立することをいう：

- (a) $T\varphi = T\eta$
- (b) $\hat{\gamma}(\varphi) = \hat{\mu}(\eta)$
- (c) $\forall j \in \hat{\gamma}(\varphi), SM(\varphi, \omega_j) = SM(\eta, \omega_j)$ □

また、 $\langle \varphi, \gamma \rangle, \langle \eta, \mu \rangle \in \langle \Phi, 2^J \rangle$ についての半順序関係の成立

$$\langle \varphi, \gamma \rangle \sqsubseteq \langle \eta, \mu \rangle$$

とは、次の3事柄イ、ロ、ハが成立することをいう：不等式

$$0 < s_j < 1$$

を満たす一つの $s_j (j \in J)$ を選んで固定する。

(イ) 次の $イ_1, イ_2$ のいずれか一つが成立する。

$$(イ_1) \quad T\varphi = T\gamma$$

(イ₂) $\gamma_1, \gamma_2, \dots, \gamma_n \in 2^J$ が存在して

$$T\eta = TA(\gamma_n)T \cdots \cdot TA(\gamma_2)T \cdot TA(\gamma_1)T \cdot T\varphi$$

$$(ロ) \quad \hat{\gamma}(\varphi) \supseteq \hat{\mu}(\eta)$$

(ハ) $\forall j \in \hat{\mu}(\eta)$ について、次の $ハ_1, ハ_2$ のいずれか一つが成立する。

$$(ハ_1) \quad SM(\eta, \omega_j) \leq SM(\varphi, \omega_j) \leq s_j$$

$$(ハ_2) \quad s_j \leq SM(\varphi, \omega_j) \leq SM(\eta, \omega_j)$$
 □

ここに、

収縮写像 $T: \Phi \rightarrow \Phi$

類似度関数 $SM: \Phi \times \Omega \rightarrow \{s | 0 \leq s \leq 1\}$, ここに, $\Omega = \{\omega_j | j \in J\}$

大分類関数 $BSC: \Phi \times J \rightarrow \{0, 1\}$

を導入して,

$$\gamma \in 2^J, \varphi \in \Phi$$

として, 構造受精作用素

$$A(\gamma): \Phi \rightarrow \Phi$$

をパターン $\varphi \in \Phi$ に作用させて得られるパターン

$$A(\gamma)\varphi$$

は次の様に定義される:

$$A(\gamma)\varphi = \begin{cases} 0 & \text{if } \gamma = \phi \text{ (empty set)} \\ \sum_{k \in \gamma} SM(\varphi, \omega_k) \cdot T\omega_k \| T\omega_k^{-1} & \text{if } \gamma \neq \phi \wedge \sum_{k \in \gamma} BSC(\varphi, k) = 0 \\ \sum_{k \in \gamma} SM(\varphi, \omega_k) \cdot BSC(\varphi, k) \cdot T\omega_k \| T\omega_k^{-1} & \text{if } \gamma \neq \phi \wedge \sum_{k \in \gamma} BSC(\varphi, k) > 0 \end{cases} \quad \square$$

また, パターン φ の, $\gamma \in 2^J$ の部分リスト (部分集合) である有効な候補カテゴリ番号のリスト $\hat{\gamma}$ は次の様に定義される:

$$\hat{\gamma}(\varphi) = \begin{cases} \phi & \text{if } \gamma = \phi \\ \{k \in \gamma | SM(\varphi, \omega_k) > 0\} & \text{if } \gamma \neq \phi \wedge \sum_{k \in \gamma} BSC(\varphi, k) = 0 \\ \{k \in \gamma | SM(\varphi, \omega_k) > 0 \wedge BSC(\varphi, k) = 1\} & \text{if } \gamma \neq \phi \wedge \sum_{k \in \gamma} BSC(\varphi, k) > 0 \end{cases} \quad \square$$

これ迄の他の研究者達の提唱している手法, 理論と異なる新しい諸点として, 次の i, ii, iii 等があげられる。特に, i のアイディアは全く鈴木によるもので, この結果, 立論内容はこれ迄の他の研究者の諸研究と全く異質なものになり, 深い情報処理論理を備えたものになった。

(i) パターン $\varphi \in \Phi$ そのものではなく, パターン φ とその候補カテゴリ番号リスト $\gamma \in 2^J$ の対リスト, $\langle \varphi, \gamma \rangle$ を導入し, φ の代わりに, $\langle \varphi, \gamma \rangle \in \langle \Phi, 2^J \rangle$ を処理対象としたこと。

(ii) 具体的な半順序関係 \cong を

$$\langle \Phi, 2^J \rangle = \{\langle \varphi, \gamma \rangle | \varphi \in \Phi, \gamma \in 2^J\}$$

の上に導入したこと。

(iii) 式 (6.2) から理解できるように, 候補カテゴリ番号リスト $\gamma_0, \gamma_1, \dots, \gamma_n, \gamma_{n+1}$ の間に非増加的包含関係

$$\gamma_0 \supseteq \gamma_1 \supseteq \dots \supseteq \gamma_n \supseteq \gamma_{n+1} \supseteq \dots$$

(カテゴリ候補の絞り上げ)

が成立することが望ましいが、この包含関係を $\langle \varphi_n, \gamma_n \rangle$ から $\langle \varphi_{n+1}, \gamma_{n+1} \rangle$ への構造受精変換 $TA(\gamma_n)T$ で定着したこと。 □

要するに、これ迄にも他の研究者諸氏によって多段認識情報処理過程の導入は考えられてなくもなかったが、S. Suzuki 以外の研究においては、認識結果に直接結びつく何んらかの半順序関係 \sqsubseteq は全く提案されていないこと、並びに、カテゴリ候補を絞り上げるパターン変換として、

S. Suzuki は構造受精変換

$$\langle \varphi, \gamma \rangle \rightarrow TA(\gamma)T\langle \varphi, \gamma \rangle$$

を考案したが、この様なパターン変換はこれ迄の諸研究においては全く提案されていないのである。

7. エルゴード定理 (情報エントロピー増大の法則) と情報処理

本章では、大規模システム (Large-Scale System) からその要素システムを分離・抽出することがエントロピーを減少させ、正に情報処理操作であることを示す。

v 個の事象

$$S_1, S_2, \dots, S_v$$

は互いに排反する確率事象とし、これらの内少なくとも一つの事象が生起することは各時刻 $t (= 0, 1, 2, \dots)$ において常に確実であるとする。

$p_{ij}(t, t+1)$: 時刻 t に事象 S_i が生起したという条件の下に、時刻 $t+1$ につき単位時間の後に、事象 S_j が生起する条件付確率 (conditional probability)

を導入する。 $p_{ij}(t, t+1)$ は、

時刻 t に発生した事象 S_i が単位時間の後に、つまり時刻 $t+1$ に状態 S_j に移る遷移確率 (transition probability)

とも呼ばれることがある。このとき、次の2条件 i, ii を課する:

条件(i) $p_{ij}(t, t+1)$ は $t' < t$ を満たすすべての時刻 t' の状況 (過去) の歴史に関係がない (単純 Markov 性)。

条件(ii) $\forall t' (= 0, 1, 2, \dots), p_{ij}(t, t+1) = p_{ij}(t+t', t+t'+1)$

が成立する (定常 (stationary) 性)。□

条件 ii より、 $p_{ij}(t, t+1)$ は p_{ij} と書いてよいから、以後、この記法を採用する。

$$0 \leq p_{ij} \leq 1, \sum_{j=1}^v p_{ij} = 1 (i, j = 1 \sim v)$$

が成立している。このとき、

$p_{ij}^{(n)}$: S_i が生起したときに、 n 単位時間の後、 S_j が生起する遷移確率

を持ち出すと、当然

$$p_{ij}^{(1)} = p_{ij}, i, j = 1 \sim v$$

が成立しているが、上述の2条件 i, ii の下で、確率の合成法則

$$p_{ij}^{(n)} = \sum_{k=1}^v p_{ik} p_{kj}^{(n-1)} = \sum_{k=1}^v p_{ik}^{(n-1)} \cdot p_{kj}, \quad n = 2, 3, \dots$$

が成立している。極限の遷移確率

$$p_{ij}^{(\infty)} \equiv \lim_{n \rightarrow \infty} p_{ij}^{(n)}$$

は、j を固定した連立1次方程式

$$\sum_{k=1}^v p_{ik} x_{kj} = x_{ij}, \quad i = 1 \sim v$$

の解 x_{ij} として求まり、実は

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n p_{ij}^{(m)} = p_{ij}^{(\infty)}$$

が成立している⁽¹⁹⁾。このとき、あらたに、

$$\text{条件(iii)} \quad \sum_{k=1}^v p_{kj} = 1, \quad j = 1 \sim v$$

を導入すると、この条件iiiの下で、エントロピー増大の法則

$$-\sum_{j=1}^v p_{ij}^{(n)} \log_2 p_{ij}^{(n)} \leq -\sum_{j=1}^v p_{ij}^{(n+1)} \log_2 p_{ij}^{(n+1)}, \quad n = 1, 2, 3, \dots$$

が成り立つ⁽¹⁹⁾。すなわち、時刻 $t = 0$ において、状態 S_i から出発すると、n 単位時間後には“標本平均 (sample mean)”の意味で、各状態

$$S_1, S_2, \dots, S_v$$

が各々、

$$p_{i1}^{(n)}, p_{i2}^{(n)}, \dots, p_{iv}^{(n)}$$

の割合で確率論的に混合している状態を得る。この確率的混合の割合は、情報エントロピー (平均情報量)

$$-\sum_{j=1}^v p_{ij}^{(n)} \log_2 p_{ij}^{(n)}$$

が時間と共に増大するようものになっている。つまり、1章の2式 (1.1), (1.2) での半順序 Ξ で表現すれば、

$$\begin{aligned} & \{p_{ij}^{(n)} \mid j = 1 \sim v\} \\ \Xi & \{p_{ij}^{(n+1)} \mid j = 1 \sim v\} \end{aligned}$$

を得、上限を求める情報処理の向きは時間の向きと逆である。

以上が Boltzmann の H 定理 (ergodic theorem) の一つの数学的定式化である。

$$q_j = \frac{1}{v}, \quad j = 1 \sim v$$

$$\Leftrightarrow \max\{q_j \mid j = 1 \sim n\} \quad \{-\sum_{j=1}^v q_j \log_2 q_j\} = \log_2 v, \quad \text{where } 0 \leq q_j \leq 1 \wedge \sum_{j=1}^v q_j = 1$$

に注意すると、

$$\lim_{n \rightarrow \infty} - \sum_{j=1}^v p_{ij} \log_2 p_{ij}^{(n)} = \log_2 v$$

$$\text{すなわち, } \lim_{n \rightarrow \infty} p_{ij}^{(n)} = 1/v$$

となるためには、条件iiiに加えて、

条件(iv) n_0 が存在して、すべての i, j について、

$$p_{ij}^{(n_0)} > 0 \text{ (不可分割性条件)}$$

を要請すれば十分であることがわかっている⁽¹⁹⁾。

この場合でも、情報処理の働きとはそのままにしていたら平均情報量 (情報エントロピー) が増大していく時間的发展*²⁰ (現象の経過方向) に対し、平均情報量を減少させていくように、現象を記述する情報構造を変換することであると定義できる。

さて、本論に入ろう。

一般に、一変数 $z (\geq 0)$ の関数 $h(z)$ を

$$h(z) = \begin{cases} -z \log_e z & \text{if } z > 0 \\ 0 & \text{if } z = 0 \end{cases}$$

とすると、

$$\frac{d^2}{dz^2} [-z \log_e z] = -\frac{1}{z} < 0$$

であるから、 $h(z)$ のグラフは上に凸であり、この事実を使えば、次の命題 7.1 を得る。

[命題 7.1] (混合システムの H 定理)

$$0 \leq p_{kl}, \sum_{l \in L} p_{kl} = 1$$

として、パラメータ $k \in K$ に依存する確率分布 $\{p_{kl} | l \in L\}$ に対し、

$$\begin{aligned} & \sum_{k \in K} q_k [-p_{kl} \log_2 p_{kl}] \\ & \leq -[\sum_{k \in K} q_k p_{kl}] \log_2 [\sum_{k \in K} q_k p_{kl}] \end{aligned} \quad (7.1)$$

ここに、 $0 \leq q_k, \sum_{k \in K} q_k = 1$

が成立し、両辺を $l \in L$ にわたり総和すれば

$$\begin{aligned} & \min_{k \in K} - \sum_{l \in L} p_{kl} \log_2 p_{kl} \\ & \leq \sum_{k \in K} q_k [- \sum_{l \in L} p_{kl} \log_2 p_{kl}] \\ & \leq - \sum_{l \in L} [\sum_{k \in K} q_k p_{kl}] \log_2 [\sum_{k \in K} q_k p_{kl}] \end{aligned} \quad (7.2)$$

(証明) $x - y$ 平面上の # K (集合 K 内に含まれる要素の総数) 個の点

$$Q_k = (p_{kl}, h(p_{kl})), k \in K$$

に各々、質量 q_k を与えたときの質量中心 (center of mass)

$$C = (\sum_{k \in K} q_k p_{kl}, \sum_{k \in K} q_k \cdot h(p_{kl}))$$

はこれらを含む最小の凸集合 (convex set) の内部または周囲に落ちる。

ところで、 $h(z)$ は上に凸な関数であるから、点

$$D = (\sum_{k \in K} q_k p_{kl}, h(\sum_{k \in K} q_k \cdot p_{kl}))$$

は点 C より y 座標が小さくない。

$$\therefore \sum_{k \in K} q_k \cdot h(p_{kl}) \leq h(\sum_{k \in K} p_{kl} q_k \cdot p_{kl}) \quad (7.3)$$

を得、この両辺の $l \in L$ にわたる総和をとれば、

$$\sum_{k \in K} q_k \sum_{l \in L} h(p_{kl}) \leq \sum_{l \in L} h(\sum_{k \in K} q_k \cdot p_{kl}) \quad (7.4)$$

の成立が知れる。式 (7.3)、式 (7.4) の両辺を $\log_e 2$ で割ったものが各々、所要の式 (7.1)、式 (7.2) である。□

上述の命題 7.1 から主張可能な事実は次の通りである：

第 $k \in K$ 番目のシステムの、第 $l \in L$ 番目の性質の存在確率 p_{kl} を、確率分布

$$\{q_k | k \in K\}$$

で混合して得られる総合システムにおいて、第 $l \in L$ 番目の性質の混合状態の存在確率

$$\sum_{k \in K} q_k \cdot p_{kl}$$

によるエントロピー

$$- \sum_{l \in L} [\sum_{k \in K} q_k \cdot p_{kl}] \log_2 [\sum_{k \in K} q_k \cdot p_{kl}]$$

は、各要素システムのエントロピー

$$- \sum_{l \in L} p_{kl} \log_2 p_{kl}$$

の、 $\{q_k | k \in K\}$ による一次結合

$$\sum_{k \in K} q_k [- \sum_{l \in L} p_{kl} \log_2 p_{kl}]$$

より必ず小さくない。□

$$0 \leq \sum_{k \in K} q_k \cdot p_{kl}, \quad \sum_{l \in L} \sum_{k \in K} q_k p_{kl} = 1$$

が成立していることに留意すると、情報処理とは、混合確率分布

$$\{\sum_{k \in K} q_k \cdot p_{kl} | l \in L\}$$

で表現される“混合状態 (mixture)”から

$$\min_{k \in L} - \sum_{l \in L} p_{kl} \log_2 p_{kl} = - \sum_{l \in L} p_{ll} \log_2 p_{ll}$$

を満たすエントロピー最小の確率分布

$$\{p_{ll} | l \in L\}$$

に対応する状態を分離することであるとすれば、1章の2式(1.1), (1.2)での半順序関係 \sqsubseteq での表現

$$\{\sum_{k \in K} q_k p_{kl} \mid l \in L\} \sqsubseteq \{p_{jl} \mid l \in L\}$$

が成立し、# K 個の要素システムから成る総合システムからこの様な第 $j \in J$ 番目の要素システムを分離することになり、このような分離を行なう情報処理の働きは確かに情報エントロピーを減少させている。

8. むすび

パターンを量子化 (quantizing, quantization; 連続量を離散的な値で代表し近似すること) して得られるもの (ベクトル) にシンボルを割り当てれば (パターンの符号化), パターンはシンボルの系列となる故, パターンはいわば, シンボルの, ある種の構造体とみられなくはない。然しながら, このようなパターンのシンボル化 (符号化) が問題となるのは無論, パターンの少量の変形あるいは規則的な座標変換に対しても, 変形前あるいは座標変換前と全く異なるシンボルの集合体を得られることが多いので, また, シンボルの集合体からパターンへの復元 (復号) が多意的であるので, シンボル情報とパターン情報とが同一に扱えないことにある。

情報は処理しようとしている対象によって運ばれているものである。情報を運んでいるもの (carriers of information) には, 主として, シンボルとパターンとがあるが, このような集まりの中で, 更に, 処理の対象となるものを考え, この内ある一定の性質を満たすものの集まりを

データ領域 (data domin)

と称する。かくの如きデータ領域へ, 一種の半順序関係を導入することの意義を介し,

連続量としての数値, 2 値記号, シンボル, 多値記号 (あいまい量をもつ記号), 構造をもつ
パターン

等に対する処理原理がこの半順序関係の 上限操作 であることを明らかにした*²⁰。

通常データ領域 D には, 半順序関係 \sqsubseteq が導入できる場合が多いことは諸例で示してきた。見掛け上異なっている情報処理の働きには共通な構造 (半順序関係 \sqsubseteq に関する上限 \sqcup を決定する構造) が潜んでいることが明らかになった。入力 $x \in D$ が与えられた場合の, 与えられた問題の求解過程としての, データ x の情報処理とは, ある写像 f にこの問題の諸性質を反映させ, $x_0 (= x)$ を初期値として, x_i に対し,

$$x_{i+1} = f(x_i)$$

を, 鎖の性質

$$x_0 \sqsubseteq x_1 \sqsubseteq x_2 \sqsubseteq \dots \sqsubseteq x_i \sqsubseteq x_{i+1} \sqsubseteq \dots$$

あるいは, 有向性

$$\exists k, x_i \sqsubseteq f^k(x_i)$$

が満たされるように, データ構造列

$$\{x_0\}, \{x_0, x_1\}, \{x_0, x_1, x_2\}, \dots, \{x_0, x_1, \dots, x_i\}, \dots$$

を求め (これが求解過程である), その上限

$$y = \sqcup_i \{x_0, x_1, \dots, x_i\}$$

を解とすることである。この上限 $\sqcup_i \{x_0, x_1, \dots, x_i\}$ は f が連続であれば, 不動点方程式

$$f(\sqcup_i \{x_0, x_1, \dots, x_i\}) = \sqcup_i \{x_0, x_1, \dots, x_i\}$$

を満たすから, 不動点方程式

$$f(\{x_0, x_1, \dots, x_i\}) = \{x_0, x_1, \dots, x_i\}$$

が成立する番号 i が見つかったら直ちに停止すれば, この不動点 $\{x_0, x_1, \dots, x_i\}$ が通常解となる訳である。これが本研究で得た, 入力 x に対する不動点意味論 (fixed point semantics) 的解釈である。

$$\exists k (\geq 1), x_j \sqsubseteq x_{j+k}$$

となっているから, x_{j+k} は x_j よりも解 y のより良い近似となっている訳だし, 事実, 計算機プログラムの不動点意味論⁽²⁾では, 写像 f がプログラムに相当し, 上限 $y = \sqcup_i \{x_0, x_1, \dots, x_i\}$ がその解としてのプログラム出力である (5.1 節を参照)。なお, 鎖を取り扱ってれば, この上限 $y = \sqcup_i \{x_0, x_1, \dots, x_i\}$ は

$$y = \sqcup_i x_i$$

と書けることに注意しておく。

入力 $x \in D$ に対し, x から定まる列

$$x_0, x_1, x_2, \dots$$

を作り,

$$f(x_i) = x_i$$

となる対 $\langle x, x_i \rangle$ (ここに, x は任意入力, x_i は x に対応する希望出力) の集合を有限個の x に対し決定し, 可能な限り, 半順序関係

$$a \sqsubseteq b \text{ ならば}$$

$$a \sqsubseteq f(a) \sqsubseteq b \sqsubseteq f(b)$$

$$\vee a \sqsubseteq b \sqsubseteq f(a) \sqsubseteq f(b)$$

を保存するように, 写像 f に含まれる諸パラメータを改良する働きは一種の学習である (4.3 節を参照)。

S. Suzuki の提案しているパターン認識の数学的理論⁽⁵⁾も正に,

$\langle \varphi, \gamma \rangle$ はパターン φ がカテゴリ \mathcal{C}_j , $j \in \gamma$ のいずれかに帰属しているという知識であるという約束の下で, ある半順序 \sqsubseteq を導入し, $\langle \varphi, \gamma \rangle$ から, 半順序関係

$$\langle \varphi, \gamma \rangle \sqsubseteq \langle \eta, \mu \rangle$$

を満たす知識 $\langle \eta, \mu \rangle$ を得るのがパターン情報処理であり、不動点方程式

$$f(\langle \eta, \mu \rangle) = \langle \eta, \mu \rangle$$

を解くのが認識の働きとしている。

このように、データ領域 D 上での情報処理とは D の中に半順序関係 \sqsubseteq を導入し、与えられた問題に応じ、適切な鎖

$$x_0 \sqsubseteq x_1 \sqsubseteq x_2 \sqsubseteq \cdots \sqsubseteq x_i \sqsubseteq x_{i+1} \sqsubseteq \cdots (x_i \in D)$$

を作ることであり、その上限 (least upper bound)

$$\sqcup \{x_0, x_1, x_2, \dots, x_i, x_{i+1}, \dots\}$$

をその情報処理結果とすれば良いことを明らかにして来た。データ領域上で半順序関係さえ適切に設定できれば、プログラム検証理論⁽¹³⁾と同様に、人工知能理論では、上限を求める具体的手続きが既知から未知を推し測る際の推論規則 (inference rule) といえよう。

この際、基本的に肝要な4点は次 $i \sim iv$ であることを指摘し、本論文を終えよう：

- (i) 半順序関係 \sqsubseteq の適切な考案。
- (ii) 各情報処理段階でのデータ x_i を、何らかの直観的描像が付与可能な様に具体的に構成すること。
- (iii) 第 i 情報処理段階から第 $(i + 1)$ 段階への変換

$$x_i \rightarrow x_{i+1}$$

を実行可能なように設定すること。

- (iv) 不動点方程式 $f(x_i) = x_i$ の成立と同値な簡便な条件の発見。 □

文献

- (1) 青木利夫, 高橋渉: 集合・位相空間要論, 培風館, 1979—09
- (2) 中島玲二: 数理情報学入門, 朝倉書店, 1982—03
- (3) 二木厚吉, 外山芳人: 項書き換え型計算モデルとその応用, 情報処理, Vol. 24, No. 2, pp. 133—146, 1983—02
- (4) 松原仁, 半田剣一: 説明に基づく一般 ≠ 部分計算, 人工知能学会誌, Vol. 6, No. 2, pp. 276—279, 1991—03
- (5) 鈴木昇一: パターン認識の数学的理論,
 - 第I部 (考え方, PRL84—6, pp. 1—10, 1984—05)
 - 第II部 (認識抽象と公理系, 定理系, PRL84—30, pp. 65—74, 1984—09)
 - 第III部 (認識抽象と不動点諸定理, PRL84—38, pp. 65—73, 1984—09)
 - 第IV部 (パターンの素領域, PRL85—27, pp. 1—10, 1985—09)
 - 第V部 (認識停止と認識同値, PRU86—8, pp. 65—74, 1986—05)
 - 第VI部 (類似度関数の三構成法, PRU86—35, pp. 51—60, 1986—07)

- 第Ⅶ部 (類似度関数の実現と解析, PRU86—69, pp. 1—8, 1986—12)
- 第Ⅷ部 (大分類関数の自己組織化, PRU87—1, pp. 1—8, 1987—05)
- 第Ⅸ部 (帰属関係あいまい度と認識情報量, PRU87—28, pp. 1—10, 1987—07)
- 第Ⅹ部 (mixture 条件の研究, PRU88—30, pp. 1—8, 1988—07)
- 第Ⅺ部 (認識プログラム FERT の近似の鎖, PRU89—1, pp. 1—8, 1989—05)
- 第Ⅻ部 (ポテンシャル関数による認識過程の評価, PRU89—27, pp. 1—8, 1989—07)
- 第Ⅼ部 (認識プログラム FERT_D の不動点認識定理, PRU89—40, pp. 1—8, 1989—09)
- 第Ⅽ部 (線形帰属係数法と諸基本定理, PRU89—66, pp. 1—8, 1989—11)
- 第Ⅾ部 (パターンの構造的類似性をもたらす 4 種類の収縮写像, PRU89—77, pp. 1—8, 1989—12)
- 第Ⅿ部 (コネクションリスト・モデルと収縮写像, PRU89—136, pp. 9—16, 1990—03)
- 第ⅰ部 (ホップフィールドネットワーク 2 値モデルと収縮写像 (1), PRU90—5, pp. 1—8, 1990—05)
- 第ⅱ部 (ホップフィールドネットワーク 2 値モデルと収縮写像 (2), PRU90—15, pp. 1—8, 1990—06)
- 第ⅲ部 (ホップフィールドネットワークの連続モデルと 2 種類の収縮写像 (1), PRU90—29, pp. 9—16, 1990—07)
- 第ⅳ部 (ホップフィールドネットワークの連続モデルと 2 種類の収縮写像 (2), PRU90—125, pp. 1—8, 1991—02)
- 第ⅴ部 (誤差逆伝播ニューラルネットモデルと特徴抽出 (1), PRU91—1, pp. 1—8, 1991—05)
- 第ⅵ部 (誤差逆伝播ニューラルネットモデルと特徴抽出 (2), PRU91—29, pp. 23—28, 1991—06)
- 第ⅶ部 (誤差逆伝播ニューラルネットモデルと特徴抽出 (3), PRU91—42, pp. 1—8, 1991—07)
- 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]
- (6) 長嶋秀世, 長嶋裕二: マイコンによる数値計算法, 昭晃堂, 1987—05
- (7) 鈴木昇一: Rosenfeld 型の確率的弛緩ラベリング法の基本的諸性質, 情報研究 (文教大学情報学部), Vol. 11, pp. 163—181, 1990—12
- (8) 向殿政男: Fuzzy 論理における 2, 3 の性質について, 電子通信学会論文誌, Vol. 58—D, No. 3, pp. 150—157, 1975—03
- (9) 山下利之, 山下清美: 合成命題におけるファジィ論理, 心理学研究, Vol. 60, No. 5, pp. 312—315, 1989—05
- (10) RICHARD CT. LEE: Fuzzy Logic and the Repolution Principle, Journal of the Association for Computing Machinery, Vol. 19, No. 1, pp. 109—119, 1972—01
- (11) 太原育夫: 人工知能の基礎知識, 近代科学社, 1988—09
- (12) 丁 亜理, 向殿政男: 修正原理および「固定値」法に基づく新しい近似推論の提案, 電子情報通信学会論文誌 D—I, Vol. J74—D—I, No. 2, pp. 117—123, 1991—02
- (13) 小川瑞夫, 小野諭: 最小不動点計算に基づくプログラムの帰納的性質の導出, 情報処理学会論文誌, Vol. 32, No. 7, pp. 914—923, 1991—07
- (14) 森田憲一, 西原典孝, 江村恵里子: Syllogism (三段論法的推論) の効率化と知識処理への応用, 電子情報通信学会論文誌, Vol. J70—D, No. 2, pp. 405—414, 1987—02
- (15) 大木幸介: 脳がここまでわかって来た (分子生理学による心の解剖), 光文社, 1990—12
- (16) Hopfield, J. J. and Tank, D. W.: Neural Computation of Decisions in Optimization Problems, *Bilo. Cybern.*, Vol. 52, pp. 141—152, 1985
- (17) Hopfield J. J.: Neurons with graded response have collective computational properties like those of two-state

- neurons, Proc., Natl. Acad. Sci. USA. 81, pp. 3088—3092, 1984—05
- (18) Rumelhart, D. E., Hinton, G. E. and Williams, R. J.: Learning Internal Representations by Error Propagation, In Parallel Distributed Processing, Vol.1., McClelland, J. L., Rumelhart, D. E. and The PDP Research Group, Cambridge, MIT Press, MA, 1986
- (19) 吉田耕作：物理数学概論，産業図書，1974—10
- (20) J. コホネン：システム論的連想記憶（情報工学・心理学のために），中谷和夫訳，サイエンス社，1980—10
- (21) 鈴木昇一，中村三郎：知識情報処理における帰納的推論，情報研究（文教大学情報学部），No. 9，pp. 173—196，1988—12
- (22) 鈴木昇一，中村三郎：最汎アトムを用いない精密化方法による prolog プログラムの帰納的自動合成システムの，C 言語による実現，情報研究（文教大学情報学部），No. 10，pp. 151—167，1989—12
- (23) 北川敏男：情報と認識（NHK 大学講座），日本放送出版協会，1979—10
- (24) 長尾真：ソフトウェア工学に対する思想，情報処理，Vol. 19，No. 9，pp. 980—891，1978—09
- (25) データベース・モデル研究委員会（小林功武他）：データ・モデル調査報告，情報処理，Vol. 19，No. 9，pp. 883—889，1978—09
- (26) J. W. ロイド：論理プログラミングの基礎，佐藤雅彦・森下真一訳，産業図書，1987—06
- (27) 鈴木昇一：認識工学（上），柏書房，1975
- (28) 鈴木昇一：測度的不変量検出形認識系の構成理論，電子通信学会論文誌 D，Vol. 55—D，No. 8，pp. 531—538，1972—08
- (29) 鈴木昇一：画像情報量とその手書き漢字への応用，画像電子学会誌，Vol. 4，No. 1，pp. 4—12，1975—04
- (30) 鈴木昇一，柴山秀雄，福永一保，古田晋吾：認識の量子論と画像の微分エントロピー，芝浦工業大学研究報告理工系編，Vol. 23，No. 1，pp. 117—125，1979—03
- (31) 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—68，1990—12
- (32) 水本雅晴：ファジィ理論とその応用，サイエンス社，1989—06
- (33) Melvin Fitting：計算論理と論理プログラミング，山崎・守谷・田中共訳，丸善株式会社，1989
- (34) 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—29，1986—12
- (35) 鈴木昇一：分析的／全体的処理と STOCHASTIC NEURO-COMPUTER（1）誤差逆伝播モデル，電子通信学会技術研究報告 [ニューロコンピューティング]，Vol. 90，No. 483，NC90—68，pp. 1—6，1991—03

補 注

* 1 相対論的力学では，エネルギーと質量の同等則（low of equivalence of energy and mass）の一つの形は「質量 m の質点の静止エネルギー（rest energy） E は $E = mc^2$ ，ここに， c は真空中の光速」である。また，熱力学では，作業物質が，変化の途中で絶対温度（absolute temperature） T の相手の熱源から熱量 Q を受けるとき $Q > 0$ ，与えるとき $Q < 0$ とすると， Q/T を作業物質がその熱源から得た換算熱量（reduced heat quantity）という。微小な可逆変化（reversible process）に対しては，微分形式

$$dS = dQ/T$$

で定義される S を系のもつエントロピー（entropy）と称する。断熱系（adiabatic system）の中で考えている相全

体の自由エネルギー (free energy) の増加 δF は、断熱系全体のエントロピーの増加 δS と

$$T \cdot \delta S = - \delta F$$

という関係にある。

* 2 これは、

SS...S0 (S が n 個ある) を数 n の表現

として、

$$SSS0 + SS0 \Rightarrow SSSSS0$$

の略記法である。

* 3 知識工学では、人間の頭の中にある知識を形式的に表現することを知識表現 (knowledge representation) といっている。この種の知識表現が適切になされていることを前提としていることに注意する。

* 4 一般に、シンボルとは形態上全く変形が許されないもので、少しでも変形があると意味が保存されない表現である。

* 5 X を関係 \sqsubseteq に関する半順序集合とする。X の要素の列

$$x_0, x_1, \dots$$

がすべての n に対して $x_n \sqsubseteq x_{n+1}$ であるとき、この要素の列を鎖 (chain) と呼ぶ。

* 6 近似解から出発し、途中で得られる近似解を修正する計算を反復し真の解に収束させる手法を反復法、緩和法あるいは弛緩法 (relaxation method) というが⁽⁷⁾、以下の求解法はこの一種である。

* 7 D を半順序集合 (partially ordered set) とするとき、D の部分集合 E の上限 (supremum ; least upper bound) e とは、次の 2 条件 i, ii を満たす D の要素であり (E の要素とは限らない)、

$$e = \sqcup E$$

と書く⁽²⁾。

(i) (上界性 ; $E \sqsubseteq e$)

$$\forall x \in E, x \sqsubseteq e$$

(ii) (最小上界性 ; $E \sqsubseteq e'$ ならば $e \sqsubseteq e'$) $\exists y \in D, \forall x \in E, x \sqsubseteq y$ ならば $e \sqsubseteq y$

* 8 D_1, D_2 を半順序集合するとき、

$$f : D_1 \rightarrow D_2$$

が単調関数 (monotonic function) であるとは、任意の $x, y \in D_1$ に対し

$$x \sqsubseteq y \text{ ならば } f(x) \sqsubseteq f(y)$$

が成り立つことをいう⁽²⁾。関数 f が単調であれば、

任意の $X \subset D_1$ に対し、

$$\sqcup f(x) \sqsubseteq f(\sqcup X)$$

が成り立つが、

$$\sqcup f(x) \sqsupseteq f(\sqcup X)$$

は必ずしも成り立たない

ということである。ここに、

$$\sqcup f(x) = \sqcup \{f(x) \mid x \in D_1\}$$

本論文でいう $h : D \rightarrow D$ は少なくとも

$$x_i \sqsubseteq x_{i+k}, \text{ つまり}$$

$$x_i \sqsubseteq h_k(x_i) \text{ (} x_i \text{ に対し } h \text{ を } k \text{ 回作用)}$$

を満たす k が十分大きい i に対し存在するものであらねばならない。

*9 文献(8)では、2命題3, 2, 3. 3の別証明が与えられていると考えることができる。

*10 $\overline{x \sqcap y} = 1 - \min(1 - x, 1 - y)$

を計算すると、

$$\overline{x \sqcap y} = \max(x, y)$$

が成立しているから、ド・モルガンの法則が成立している。よって、 $x \sqcap y$, \overline{x} という2演算を定義すると、ド・モルガンの法則を適用して、 $x \sqcup y$ を

$$x \sqcup y = \overline{\overline{x} \sqcap \overline{y}}$$

と定義できる。ところで、fuzzy 連言 $x \sqcap y$ を

$$x \sqcap y = x \cdot y$$

と定義すると、

$$\begin{aligned} \overline{x \sqcap y} &= 1 - (1 - x) \cdot (1 - y) \\ &= x + y - x \cdot y \end{aligned}$$

となるから、

$$x \sqcup y = x + y - x \cdot y$$

と定義することも考えられるが、本論文ではこの定義を採用しない。

なお、 $x \sqcap y$ を

$$x \sqcap y = \min(x, y) \cdots \cdots \text{ミニマム規則 (minimum rule)}$$

$$x \sqcap y = x \cdot y \cdots \cdots \text{積規則 (multiplying rule)}$$

のいずれで定義するのが良いのかの議論は文献(9)でなされている。

*11 基本命題とは5種類の論理記号(結合子: connective)

~ (否定; negation)

∧ (連言; conjunction)

∨ (選言; disjunction)

→ (含意; implication)

↔ (同値; equivalence)

を含まない命題をいい、リテラルとは

基本命題または基本命題の否定

をいい、節とは、

リテラルの選言

をいう。

*12 \sqcap は半順序集合 D で定義された下限、あるいは最大下界を求める演算である。半順序集合 D の部分集合 E の上限 $\sqcup E$ は必ず存在する訳ではないが、存在するとすれば一つだけある。 D の部分集合 E の下限 $\sqcap E$ は上限演算 \sqcup を用いて、

$$\sqcap E = \sqcup \{x \in D \mid \text{任意の } y \in E \text{ について,}$$

$$x \sqsubseteq y\}$$

$$= \sqcup \{x \in D \mid x \sqsubseteq E\}$$

と定義してもよい。下限もあるとは限らないが、存在すれば一つだけである⁽¹²⁾。

*13 電圧、電流などと違って、エネルギーはその効果でしか測れない。この意味で、電圧、電深等をハードウェア的とすると、エネルギーはソフトウェア的である。データはハードウェア的であるが、情報はそれに対応してソフトウェア的である。情報は、それを知って思考・行動に、また感情・意欲に移したときの効果でしか測れ

なくて、音声・文字・データをハードウェア的だとすると、ソフトウェア的であろう。これは著者の考えであり、情報というものを定義しにくい理由の一つと考えられる。

*14 この学習アルゴリズムを特別な場合として含む“可分な一般抽象 Hilbert 空間上のアルゴリズム”については著者の研究（文献（5）の第XXI部～第XXII部）を見よ。

*15 以下に説明される導出原理では、節集合 C が充足不能であれば、もとの論理式も充足不能すなわち恒偽式である。

*16 T を常に真 (true) なる式とすると、

$$C_i = T \wedge C_j = T$$

であれば、 p の真偽にかかわらず、

$$C_{ij} = T \text{ である}$$

ことに注意する。 T は素式の真偽にかかわらず常に真となる論理式つまり恒真式 (tautology) の代表的なものである。

*17 これは、どのような“対象領域”のどのような“解釈”に対しても対象領域のある要素については節集合 C が偽となることである。

*18 変数を含まない素式 (アトム) をグランドアトム (ground atomic formula) というが、言語 L の各グランドアトムに対し、true または false を割り当てる写像 M のことをモデルといてよい。

*19 本理論の研究は続行中であり、本章の記述はその意味で便宜的であることを断っておきたい。

*20 時刻 t, s での確率分布を各々、

$$\{p_i(t) \mid i \in L\}, \{q_i(s) \mid i \in L\}$$

ここに、

$$0 \leq p_i(t), q_i(s), \sum_{i \in L} p_i(t) = \sum_{i \in L} q_i(s) = 1$$

とすると、時間の向き (過去から未来の向き) とは、平均情報量が増大していく方向、つまり、不等式

$$-\sum_{i \in L} p_i(t) \log_2 p_i(t) \leq -\sum_{i \in L} q_i(s) \log_2 q_i(s)$$

が成立するならば、

時刻 s は時刻 t の未来である

と定義することがある。

*20 多くの可能性の中から特定の解を発見する探索戦略法⁽¹¹⁾も結局は、ある半順序関係の下での上限を求め半順序原理に帰着できるが、紙面の都合上割愛された。

(鈴木昇一, 半順序と情報処理, 第12号投稿論文)

文教大学情報研究