情報学基礎 10－1
(1988. 9. 16)

データ知識表現システムの論理学的基礎について*)

中村　昭，　高　建明
広島大学　工学部 応用数学教室

観測，測定，記述などからでは，対象が一意的に決まらない場合が多い。このような正確でないデータ知識システムを計算機で処理可能にするためには，その表現を論理的にしなくてはならない。区別できない（indiscernibility）とか，類似（similarity）といった概念が，この目的に沿って導入される。これらの新しい概念は論理学の様相演算に対応させることができる。また，正確でないデータ知識システムの表現能力も，これらの概念を用いて比較することができる。
本稿では，このような問題の論理学的基礎について，ポーランド学派の動向とわれわれの結果を報告する。

ON LOGICAL FOUNDATIONS OF DATA-BASED KNOWLEDGE REPRESENTATION SYSTEMS

Akira NAKAMURA and Jian-Ming GAO
Department of Applied Mathematics, Hiroshima University

Higashi-Hiroshima, 724 Japan

Usually, data-based knowledge systems contain vague and imprecise information of objects. To discuss these vagueness and imprecision from a viewpoint of knowledge representation, it is necessary to introduce new notions. As one of these concepts, Polish logicians Z. Pawlak and E. Orlowska have proposed " indiscernibility " and " similarity ". These concepts correspond to modal operators of the classical modal logics.

In this paper, we report some of our basic results on these modal operators together with activities of Polish research works on this field.

---

*) この研究のいくつかの部分は，筆者中村のポーランド科学アカデミー（PAS）滞在中になされたものである。同アカデミー（PAS）の招へいに筆者は感謝したい。

1.　はじめに

人工知能の研究にその起源をもつ知識工学(knowledge engineering)は，人間の知的活動を解明する情報処理技術として近年急速に脚光をあびるようになった。人間の持つ柔軟な知的情報処理能力を計算機に賦与しようとするこの技術は，基礎研究とともその重要性が益々認識されている。

知識に関する研究は，もともと哲学の問題であった（例えば［1］）。あるいはその生成発達についての問題として心理学とも無関係ではなかった［2］。また，最近は認識の問題を取り扱う認知科学(cognitive science)の中心テーマでもある。しかしながら，工学的見地から知識の問題を考察するならば，その表現(representation)と利用(utilization)という両面が強調して議論されなくてはならない。

ここでは，データ知識の表現を中心として「あいまい・不正確な」ものを含む複雑多様な知識・情報への数理論理学からのアプローチを試みたい。特に，筆者（中村）のポーランド訪問（63.7.25--8.8　ポーランド科学アカデミー滞在）を通してポーランド学派が用いている手法を解説してみたい。

2.　あいまい・不正確な知識と rough sets

1965年に L. Zadeh は fuzzy sets という概念を提案しあいまい（fuzzy）な知識・情報への数理的接近を試みた［3］。以来この提案は工学的技術への応用もあわせ各分野で注目されていることは周知の通りである。

これに対して，1982年ポーランドの数理論理学者 Z. Pawlak は，あいまいな（vague）不正確な（imprecise）データを解析するための数学的道具として rough sets という概念を導入した［4］。前者が帰属度関数（membership function）を基礎とするのに対して，後者は具体的データの基盤の上で考えられるものである。その基本思想は区別できない（indiscernibility）という概念にある。現実の世界から得られる測定，観察，記述等はもともと不確実性を含むものである。それらのデータを解析するには実用的道具が有効である。rough sets はこのような背景のもとで提案されたもので，近年各国で注目を集め始めている。

さて，indiscernibility の定義からはじめよう。

<u>定義2.1</u>　KR（knowledge representation）システム S は，つぎの 4組である。

$$S = (OB,\ AT,\ \{VAL_a\}_{a \in AT},\ f)$$

ただし，

OB は objects と呼ばれる要素の非空の集合

AT は attributes と呼ばれる要素の非空の集合

VAL_a は attribute a の値の非空の集合

$$f:\ OB \times AT \rightarrow \bigcup_{a \in AT} VAL_a$$

<u>例2.1</u>　Table 1.1 による KRシステムにおいて objects は次の attributes をもつ人間である：姓，出身地，生年，婚姻状態。

object	姓	出身地	生年	婚姻状態
1	Smith	London	1958	独身
2	Jones	Glasgow	1930	寡夫
3	Brown	Gerald	1910	既婚
4	Howard	Edinburgh	1926	離婚
5	Wells	Liverpool	1970	独身

Table 1.1

<u>例2.2</u>　Table 2.2 は刑事データファイルの記載である。

Name	H.C.	E.W.	E.O.	E.S.	E.C.
John	Full	Thin	Narr.	Medium	Blue
Claude	Rec.	Bushy	Narr.	Wide	Brown
Robert	Bald	Medium	Medium	Close	Green
Nancy	Full	Thin	Wide	Medium	Hazel
Ingmar	Rec.	Bushy	Wide	Wide	Blue
Roberto	Full	Thin	Narr.	Close	Blue
Marcel	Rec.	Bushy	Narr.	Medium	Black
Johanna	Full	Thin	Medium	Medium	Green
Jurgen	Bald	Thin	Wide	Close	Green
Tom	Full	Medium	Narr.	Medium	Hazel

Table 2.2

<u>例2.3</u>　Table 3.3 は心臓病の観察記録である。

患者	Gas.	Dys.	Cy.	Pulm.	H.R.	H.	Ed.	Deg
p1	37	1	1	1	62	0	0	1
p2	43	2	3	4	76	8	3	3
p3	42	1	2	1	71	1	0	1
p4	43	0	3	2	80	5	1	1
p5	48	1	3	3	92	6	3	3
p6	38	1	3	2	87	5	1	2
p7	54	0	0	0	95	1	0	2
p8	40	3	0	0	128	1	0	0
p9	40	1	0	0	111	1	0	1
p10	50	0	1	0	68	2	1	1

Table 3.3

<u>定義2.2</u>　$S = (OB, AT, \{VAL_a\}_{a \in AT}, f)$ をKRシステムとする。P（⊆ AT）に対し indiscernibility relation ind(P) は，次式で定義される OB の上の二項関係である。

$(x, y) \in ind(P) \iff$ すべての a（∈ P）に対して $f(x, a) = f(y, a)$。

<u>例2.4</u>　Table 2.4 を例にとって説明する。

	a	b	c
o1	0	0	1
o2	0	0	1
o3	0	1	0
o4	1	2	0
o5	1	0	1
o6	1	2	0
o7	0	1	0
o8	0	2	1
o9	1	0	1
o10	0	1	0
o11	0	2	1

Table 2.4

ind(P) の同値類は次のとおりである。

ind({a}): {o1, o2, o3, o7, o8, o10, o11}　{o4, o5, o6, o9}

ind({b}): {o1, o2, o5, o9}　{o3, o7, o10}　{o4, o6, o8, o11}

ind({c}): {o3, o4, o6, o7, o10}　{o1, o2, o5, o8, o9, o11}

定義 2.2 において特に P = φ の場合を ind(φ) = OB×OB とする。

このとき，つぎが容易に得られる。

<u>Fact 2.1</u> 任意の集合P, Q（⊆ AT）に対して

(1) ind(P) は反射的，対称的，推移的である。

(2) ind(P ∪ Q) = ind(P) ∩ ind(Q)。

(3) P ⊆ Q ならば ind(Q) ⊆ ind(P)。

したがって (1)からind(P)は同値関係であることがわかる。

さて，ここで ind(P) を用いてデータ知識の近似（approximation）概念を導入しよう。

<u>定義2.3</u> X ⊆ OB, P ⊆ AT とする。

(1) $\underline{ind}(P)X = \{o \in OB \mid$ すべての o'（∈ OB）に対して $(o, o') \in ind(P)$ ならば $o' \in X\}$

(2) $\overline{ind}(P)X = \{o \in OB \mid$ あるo'（∈ OB）に対して $(o, o') \in ind(P)$ かつ $o' \in X\}$

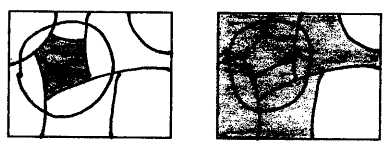
(3) $bn(P)X = \overline{ind}(P)X - \underline{ind}(P)X$

$\underline{ind}(P)X$, $\overline{ind}(P)X$, bn(P)X をそれぞれ X の P-下近似（P-lower approxiomation），P-上近似（P-upper approximation），P-境界（P-boundary）という。

<u>定義2.4</u>　X ⊆ OB, P ⊆ AT とする。次式が成立するとき，X は P-定義可能（P-definable）という。

$\underline{ind}(P)X = \overline{ind}(P)X$ か X = φ 。

定義2.3を図示すれば，つぎのようになる。



集合 X が P-定義可能であれば，それは性質 P によって決まる。そうでなければ，X は性質 P によっては一般に規定されない。P によって大体は決まるけれど正確には定まらない。すな

わち X は P において rough set である。一般に universe（世界）についてのわれわれの知識は十分正確でない。したがって，一般にある集合を区別する（discern）ことはできない。このように考えれば，データ知識への近似概念（rough concept）の導入は極めて自然である。

定義2.5 P-rough set とは $\overline{\mathrm{ind}}(P)X \neq \underline{\mathrm{ind}}(P)X$ を満足する OB の部分集合 X のことである。

定義2.6 X（⊆ OB）に対し $\in_p$，$\overline{\in}_p$ が次式で定義される。

(1) $x \in_p X \iff x \in \underline{\mathrm{ind}}(P)X$

(2) $x \overline{\in}_p X \iff x \in \overline{\mathrm{ind}}(P)X$。

(1) を "x は確かに X に属する"，(2) を "x は可能的に X に属する" と読む。

定義2.6 から $\underline{\mathrm{ind}}(P)$，$\overline{\mathrm{ind}}(P)$ が様相論理と密接な関係を持つことがわかるが，これについては後で述べよう。

3. 類似性について

indiscernibility に強い関係のある類似性（similarity）の概念についてここで述べよう。

多くの実際的問題では，一つの対象の一つの属性の値は決定的（a single value）ではない。このような KRシステムを多値 KRシステムという。

定義3.1 多値 KRシステム S とは，次で定義される4組である。

$$S = (OB, AT, \{VAL_a\}_{a \in AT}, g)$$

ただし，OB, AT, VAL は定義2.1 と同様であるが，$g \subseteq OB \times AT \times VAL$ で，$o \in OB$，$a \in AT$ に対して $v \in VAL_a$ がつねに存在する。

例3.1

	Language	Degree
p1	F, D	BS, MS, Ph.D
p2	H, R	BS
p3	F, D, S	BS, MS
p4	F	BS, MS
p5	F, D	BS
p6	R	BS

Table 3.1

多値 KRシステムは，また $f: OB \times AT \to 2^{VAL_a}$ で定義される KRシステムと考えてもよい。

さて，P（⊆AT）に対し，in(P)，sim(P)がつぎのように定義される。

定義3.2 P（⊆AT）に対し

(1) $(o, o') \in in(P) \iff$ すべての $a \in P$ に対して $f(o,a) \subseteq f(o',a)$。

(2) $(o, o') \in sim(P) \iff$ すべての $a \in P$ に対して $f(o,a) \cap f(o',a) \neq \phi$。

この in(P)，sim(P) を P-包含，P-類似という。

Fact 3.1

(1) in(P) は反射的で推移的である。

(2) sim(P) は反射的で対称的である。

例3.2

	Indication	Contra-indication	Side effect
M_1	i_1, i_2, i_3	c_1, c_2	c_1, c_2, c_3
M_2	i_2, i_4	c_2, c_3	c_1, c_2, c_4
M_3	i_4	c_2	c_2, c_3
M_4	i_2, i_3	c_2	c_1, c_3
M_5	i_3	c_1, c_3	c_1, c_2
M_6	i_4	c_2	c_2, c_4

Table 3.2

in(AT): (M_4, M_1) (M_6, M_2) (M_i, M_i)
$i = 1, \ldots, 6$

sim(AT): (M_1, M_2) (M_1, M_4) (M_2, M_3) (M_2, M_4)
(M_2, M_6) (M_3, M_4) (M_i, M_i)
$i = 1, \ldots, 6$

上の (M_i, M_j) に対して (M_j, M_i)。

ind(P) の場合と同様に，$\underline{\mathrm{in}}(P)$，$\overline{\mathrm{in}}(P)$ 及び $\underline{\mathrm{sim}}(P)$，$\overline{\mathrm{sim}}(P)$ が次のように定義される。

定義3.3

(1) $\underline{\mathrm{in}}(P)X = \{\ o \in OB \mid \forall o'\ (o,o') \in \mathrm{in}(P) \Rightarrow o' \in X\ \}$

(2) $\overline{\mathrm{in}}(P)X = \{\ o \in OB \mid \exists o'\ (o,o') \in \mathrm{in}(P)$ かつ $o' \in X\ \}$

(3) $\underline{\mathrm{sim}}(P)X = \{\ o \in OB \mid \forall o'\ (o,o') \in \mathrm{sim}(P) \Rightarrow o' \in X\ \}$

(4) $\overline{\mathrm{sim}}(P)X = \{\ o \in OB \mid \exists o'\ (o,o') \in \mathrm{sim}(P)$ かつ $o' \in X\ \}$

上の定義から $\underline{\mathrm{in}}(P)$，$\overline{\mathrm{in}}(P)$，$\underline{\mathrm{sim}}(P)$，$\overline{\mathrm{sim}}(P)$ もそれぞれ様相論理の必然性，可能性に対応することがわかる。

4. KRシステムの表現能力

いくつかの KRシステムが与えられたとき，その表現能力を比較することは興味ある。

いま，$S_i = (\ OB\ ,\ AT_i,\ VAL_i\ ,\ f_i\ )$ とし，$ST = \{S_i\}_{i \in I}$ とする。

定義4.1 二つの KRシステム S_i，S_j ($\in ST$) が次の関係を満足するとき，S_iは S_jより強い表現能力があるといい，$S_i \leqq S_j$ とかく。すなわち

$$S_i \leqq S_j \Leftrightarrow \mathrm{ind}(AT_i) \subseteq \mathrm{ind}(AT_j)$$

これは，KR システム S_iの方が ind(AT) によって OB を同値類に分類した際 S_jによるのより細かいことを表している。つまり，データ知識が豊富なことを意味している。

Fact 4.1 つぎの関係は同値である。

(1) $S_1 \leqq S_2$

(2) すべての $X \subseteq OB$ に対して $\overline{\mathrm{ind}}(AT_1)X \subseteq \overline{\mathrm{ind}}(AT_2)X$

(3) すべての $X \subseteq OB$ に対して $\underline{\mathrm{ind}}(AT_2)X \subseteq \underline{\mathrm{ind}}(AT_1)X$。

さて，ここで KRシステム $S = (\ OB,\ AT,\ \{VAL_a\}_{a \in AT}\ ,\ f\ )$ の attributes の従属性について考えよう。

定義 4.2 P，Q ($\subseteq AT$) について，$P \rightarrow Q$ が次式で定義される。

$P \rightarrow Q \Leftrightarrow \mathrm{ind}(P) \subseteq \mathrm{ind}(Q)$。そして $\rightarrow$ を従属性の関係という。

$P \rightarrow Q$ は，P に属する attribute で二つの対象が indiscernible であれば，Q に属する attribute でも indiscernible であることを意味している。

例4.1

	a1	a2	a3	a4
o1	0	0	0	0
o2	0	1	0	2
o3	1	1	0	1
o4	1	1	0	1
o5	0	1	1	2

Table 4.1

このとき，

ind({a1}): {o1, o2, o5}, {o3, o4}

ind({a2}): {o1}, {o2, o3, o4, o5}

ind({a3}): {o1, o2, o3, o4}, {o5}

ind({a4}): {o1}, {o2, o5}, {o3, o4}

したがって，$\{a4\} \rightarrow \{a2\}$，$\{a4\} \rightarrow \{a1\}$ である。

5. Rough sets にもとづく論理系について

知識表現システムの解析に，最近様相論理が用いられていることは良く知られている。われわれの場合も2節と4節でこのことにふれた。問題を計算機で実行可能にするためには表現言語を正確に記述しなくてはならない。それには記号論理学の 論理式wff の概念が十分適合する。ここでこれまで述べてきた KRシステムにもとづく論理学について考えよう。まず，2節におけると同じ記号を用いて成立するいくつかの式を与えておく。

Fact 5.1

(1) $\underline{ind}(P)(X \cap Y) = \underline{ind}(P)X \cap \underline{ind}(P)Y$

(2) $\underline{ind}(P)X \subseteq X$

(3) $\underline{ind}(P)\underline{ind}(P)X = \underline{ind}(P)X$

(4) $\underline{ind}(P)OB = OB$

(5) $\overline{ind}(P)(X \cup Y) = \overline{ind}(P)X \cup \overline{ind}(P)Y$

(6) $X \subseteq \overline{ind}(P)X$

(7) $\overline{ind}(P)\overline{ind}(P)X = \overline{ind}(P)X$

(8) $\overline{ind}(P)\phi = \phi$

(9) $\overline{ind}(P)X = -\underline{ind}(P)-X$

(10) $\underline{ind}(P)X = -\overline{ind}(P)-X$

(11) $X \subseteq Y \Longrightarrow \overline{ind}(P)X \subseteq \overline{ind}(P)Y$ かつ $\underline{ind}(P)X \subseteq \underline{ind}(P)Y$

これらの事実を用いて，2節の KRシステムの公理系を与えることができるが，まずそのシンタックスとセマンティックスの定義からはじめる。

定義5.1 論理系 INDL* の記号は次からなる。

(1) VAROB: 対象を表す変数に集合

(2) $\overline{ind}(P)$, $\underline{ind}(P)$, ただし P は性質の集合を表す変数

(3) 論理記号: $\neg$, $\vee$, $\wedge$, $\supset$, $\equiv$

(4) カッコ: (,)。

定義5.2 INDL* の WFF（論理式の集合）は次で定義される。

(1) VAROB $\in$ WFF

(2) F, G $\in$ WFF $\Longrightarrow$ $\neg F$, $F \vee G$, $F \wedge G$, $F \supset G$, $F \equiv G \in$ WFF

(3) F $\in$ WFF $\Longrightarrow$ $\underline{ind}(P)F$, $\overline{ind}(P)F \in$ WFF

つぎで INDL* のセマンティックスは与えられる。

定義5.3 モデルとは，つぎのシステムのことである。

$$M = (OB, AT, \{ind(P)\}_{P \subseteq AT}, v)$$

ただし，

OB, AT, ind(P) は KRシステムのそれと同じ
v は付値関数で，すべての F $\in$ VAROB に対して $v(F) \subseteq OB$。

定義5.4 対象 o がモデル M で wff F を充たす (satisfy)（M, o sat F とかく）とは つぎのことである。

(1) すべての F $\in$ VAROB に対して
M, o sat F $\Longleftrightarrow$ $o \in v(F)$

(2) M, o sat $\neg F$ $\Longleftrightarrow$ M, o sat F でない

(3) M, o sat $F \vee G$ $\Longleftrightarrow$ M, o sat F または M, o sat G

(4) M, o sat $F \wedge G$ $\Longleftrightarrow$ M, o sat F かつ M, o sat G

(5) M, o sat $F \supset G$ $\Longleftrightarrow$ M, o sat $\neg F \vee G$

(6) M, o sat $F \equiv G$ $\Longleftrightarrow$ M, o sat $(F \supset G) \wedge (G \supset F)$

(7) M, o sat $\underline{ind}(P)F$ $\Longleftrightarrow$ すべての o' に対し $(o, o') \in ind(v(P))$ ならば M, o' sat F

(8) M, o sat $\overline{ind}(P)F$ $\Longleftrightarrow$ ある o' に対し $(o, o') \in ind(v(P))$ かつ M, o' sat F

wff F は M, o sat F である o が存在するとき充足可能といわれる。また，すべての $o \in$ OB に対し M, o sat F のときモデル M で真であるといわれる。更に任意のモデル M で wff F が真のとき F は恒真 (valid) といわれ $\models F$ とかかれる。この INDL* に対して，健全な ($\vdash F \Longrightarrow \models F$) そして完全な ($\models F \Longrightarrow \vdash F$) 公理系がつぎで与えられている（$\vdash$は証明可能を表す記号）[5]。

A1. 古典命題論理のすべての恒真式

A2. $\underline{ind}(P)(F \supset G) \supset (\underline{ind}(P)F \supset \underline{ind}(P)G)$

A3. $\underline{ind}(P)F \supset F$

A4. $F \supset \underline{ind}(P)\neg\underline{ind}(P)\neg F$

A5. $\underline{ind}(P)F \supset \underline{ind}(P)\underline{ind}(P)F$

R1. $\dfrac{F,\ F \supset G}{G}$　　R2. $\dfrac{F}{\underline{ind}(P)F}$

この INDL* システムは $\underline{ind}(P)$, $\underline{ind}(Q)$ 等原理的には複数個の様相演算子を含むものである。したがってある意味でマルチ様相論理系とも考えられる。

さらに，この INDL* の決定問題も解決されている [6]。この決定問題を利用すれば，二つの表現言語が同値であるかどうかも判定されるから，KRシステムへの質問言語の様式を簡単にすることも可能となり実際にも有効と思われる。これらの証明はかなり数学的技巧を要するのでここでは省略する。4節の多値 KRシステムに対しても同じ問題がすでに議論され，いくつかが

解決されていることを注意しておこう［5］，［7］，［8］。

6. 将来の課題と展望

KRシステムから出発して，indiscernibility，rough set，similarity 等を導入し，あいまい・不正確なデータ知識システムへのアプローチをポーランド学派の流れにそって解説してきた。本稿では述べなかったが，KRシステムに時間のパラメーターを導入すれば temporal KRシステムが得られる［5］，［9］。事実これについてもいくつの研究結果が得られている。また具体的 KRシステムから帰納による類推の議論も試みられている［10］。これらについては別の機会に我々の結果と共に報告したい。更にINDL* は性質を示すパラメーターが P，Q 等であったが，その間に集合論演算子∪，∩，－を導入したらどうなるか（この公理系は未解決）等の問題がある［11］。そして更に rough relation を定義し［12］，rough quantifier［13］を導入し rough predicate logic が提案できるか？（すでポーランド学派は考えている。）そしてその実際の応用はどうかなどの問題がある。また，rough sets とfuzzy sets とはどのような関係にあるのかも興味ある問題である［14］。

Rough sets の提案者 Dr. Z. Pawlak が述べているように，この理論は AI の分野，例えば machine learning，帰納的推論，パターン認識，あいまいなデータの解析，decision support system 等々に強いインパクトを与えるように思われる。

文献

[1] J. G. Fichte: Grundlage der gesamten Wissenschaftslehre, 1794.

[2] U. Neisser: Cognitive Psychology, Appleton, 1967

[3] L. A. Zadeh: Fuzzy sets, Infor. and Contr. 8, 338-353, 1965

[4] Z. Pawlak: Rough sets, Intern. J. of Computer and Information Sciences, 11, 341-356, 1982

[5] E. Orlowska and Z. Pawlak: Logical foundations of knowledge representation, ICS PAS Reports 537, 1984

[6] N. Kadota and A. Nakamura: A decidability result on a logic for data analysis, to appear in Bulletin of the Polish Academy of Sciences (Theoretical Computer Science)

[7] E. Orlowska and Z. Pawlak: Representation of nondeterministic information, Theoretical Computer Science 29, 27-39 1984

[8] A. Nakamura and J. M. Gao: A decidability result on the logic of nondeterministic information, to appear in Bulletin of the Polish Academy of Sciences

[9] W. Penczek: The temporal logic for event structures, ICS PAS Reports 616, 1987

[10] E. Orlowska: Semantical analysis of inductive reasoning, ICS PAS Reports 547, 1984

[11] E. Orlowska: A logic of indiscernibility relations, Lecture Notes in Computer Science 208, 177-186, 1985

[12] Z. Pawlak: Rough relations, ICS PAS Reports 435, 1981

[13] L. Szczerba: Rough quantifiers, to appear in Bulletin of the Polish Academy of Sciences.

[14] A. Nakamura and J. M. Gao: Logic for fuzzy data analysis, Proceedings of International Workshop on Fuzzy System Applications, Aug. 20-24, 1988, Iizuka Japan, 119-120