人工知能における確率論の役割

――不確定要素を含む知識表現とその推論――

住田　潮

人工知能が計算機科学の花形分野として持てはやされるようになって久しい．その本質は，本来０－１演算で記述するには不向きであるような人間的判断を含む問題をいかにして電子計算機に処理させるか，という点にある．技術的には，対象となる知識を計算機に見あった形で表現し，それに対応する推論形式を開発することが問題となる．知識の内容が確定的であってもかなり難しいが，不確定要素を含む場合には困難さは飛躍的に増す．さまざまな方法論が提出されてきたが未だ統一されておらず，研究者は３つの大きな流れにわかれて激論を戦わせているのが現状である．本稿では，これら３つの異なる見解について概説を試みる．

1.　ベイズ理論による接近

不確定要素を含む知識の表現とその推論を，基本的に確率論の枠内で処理していこう，という立場をとる研究の流れが存在する．話を簡単にするため，病気診断システムを考えてみよう．今，Aを病名，Xをある検査の結果とする．Xが判明したとき，患者が病気Aを持つ条件つき確率 $P(A|X)$ を推定することが問題となる．ここで，ベイズ公式により，

(1)　$$P(A|X)=P(A,X)/P(X) =P(X|A)P(A)/P(X)$$

が成立する．ある病気を持つとわかっている患者に対する検査結果は予測され得る．(1)の最右辺に現われる確率は医者にとって推定しやすいものであり，ベイズ公式はそれらからより推定の難しい条件つき確率を導くことを可能にする．

もちろん，問題はこんなに単純ではない．考慮すべき病名は数多くあり，検査も複数と見なければならない．これらは互いに錯綜しており，ベイズ理論にもとづく推論体系を確立するためには，対象となる病名や検査結果の集合に関し，すべての部分集合の同時確率を知ることが必要になる．集合に含まれる事象の数に対し，知らなければならない同時確率の数は指数的にふくれ上がり，多くの場合そうした知識を完全に獲得することは不可能である．

対象となる事象間の確率的依存関係が木構造をなす時，その構造に沿った形で条件つき確率に関する独立性が保証され，上述した指数的膨張を線形的に抑えることが可能となる．図１で，A, B, C, D, Eを対象となる事象とし，矢印が確率的依存関係を表わすとしよう．Cが確定した時，A, B，あるいはEに関する新たな情報は，CからDへの推論に影響を与えない．この木構造に即した独立性が，同時確率の表現を容易にする．たとえば，

(2)　$$P(A,B,C,D,E) =P(E|C)P(D|C)P(C|A)P(B|A)P(A)$$

が成立する．

もっと一般的に，n個の確率事象 $X=(X_1,\cdots,X_n)$に対し，事象間の確率的依存関係がXの並び変え$Y=(Y_1,\cdots,Y_n)$によって表わされる木構造を持つとしよう．$F(Y_i)$でY_iの親を表わすと，

(3)　$$P(X_1,\cdots,X_n) =P(Y_1)\prod_{i=2}^{n} P(Y_i|F(Y_i))$$

となる．与えられた問題がこうした構造を備えていれば，ベイズ理論による接近は俄然現実味を帯びてくる．もちろん，木構造で表現される問題はきわめて特殊であり，普遍的に期待できるわけではない．

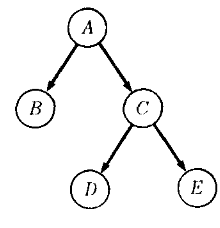

図 1　木構造を持つ確率的依存関係

しかし，一般的な問題に木構造を人工的に持ち込めるかと問うことは自然である．

事象$X=(X_1,\cdots,X_n)$に対し，人工的に確率事象$W=(W_1,\cdots,W_m)$を導入することにより(X,W)に木構造を与えることができる時，Xは木分割可能と呼ばれる．特

すみた　うしお
ロチェスター大学サイモン・スクール
国際大学大学院国際経営研究科

に$m=1$であれば，Xは星型分割可能と呼ばれる．木分割について一般的な方法を確立することは難しく，**図2**に見られる$n=3$の場合の星型分割の可能性でさえ，その判別は自明ではない．$X_i(1\leqq i\leqq 3)$が2つの値だけを取り得る確率変数である時は，Pearl [4] 星型分割可能性に関する必要十分条件を導いた．その内容は，「各変数が相互に正の相関を持ち，任意の2つの変数の相関係数は，残りの変数が与えられた時の条件付相関係数より大きくなければならない」というもので，単純な問題についても，星型分割の可能性の範囲はそんなに広くはないことを示唆する．

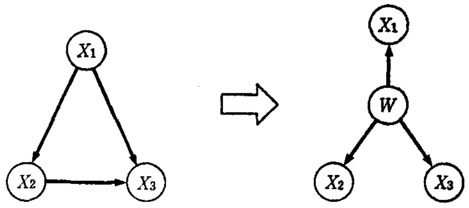

図 2 (X_1, X_2, X_3) の星型分割

Pearlは厳密に確率的意味づけが可能な範囲で星型分割を追及し，その結果，かなり強い条件を課すことを強いられた．Sarkar [6] は，同じ図2で表わされる問題についてラテン格子法における Lazarsfeld [3] の考え方を応用し，面白い方法を提唱している．

(X, W)の木構造について，$P(X)=\int P(X, W)\cdot P(dW)$が成立するように，7つの変数を含む7つの非線形方程式を解き，負の重みや1より大きな重み，あるいは複素数の重みさえ許すことで確率的解釈を放棄する一方，それらがあたかも確率的実体を持つかのごとくに機械的推論を行ない，最終結果が確率論の世界に帰ってくるように配慮している．常に可能であるわけではないが，従来の方法に比べて，より広い範囲の問題を取り扱えることを示した．

PearlやSarkarの方法は，与えられた問題が図2の形を基本単位として木構造でつながっている場合には，そのまま拡張できる．しかし，それでもなお，制限は強過ぎると言わねばならない．一般的には，確率的依存関係の強弱を判断し，星型分割が木構造をなすようにつながった形で実際の確率的依存関係を近似するが，そのさいの目的関数の取り方の難しさもあり，研究はまだこれから，というところである．

2. 可能性理論による接近

別の流れとして，不確定要素を含む知識表現とその推論を，Zadeh [9] のファジー集合論にもとづいて構築していこう，という立場をとる研究グループが存在する．$X=\{X_1, \cdots, X_n\}$なる全体集合が与えられた時，Xのファジー部分集合Aは，可能性関数$a: X\to[0,1]$によって確定される．これに対応して，さまざまな集合演算を，aが0か1だけをとる場合はブール代数に一致するような形で一般化する．具体的には，

(4) $A\subset B \Leftrightarrow a(x)\leq b(x) \quad \forall x\in X$

(5) $A\cap B \Leftrightarrow \min\{a(x), b(x)\} \quad \forall x\in X$

(6) $A\cup B \Leftrightarrow \max\{a(x), b(x)\} \quad \forall x\in X$

などである．数学的には，ブール変数を要素とするベクトル空間の極限として表現され得るから，確率論と全く無縁というわけではない．

対象を複数個の全体集合でまず表現し，それぞれのファジー部分集合を結びつけるような推論方法を構成することが，可能性理論の中核をなす．今，全体集合$X=\{X_1, \cdots, X_n\}$が$A_1, \cdots, A_N$なるファジー部分集合を持ち，同じように$Y=\{Y_1, \cdots, Y_m\}$と$B_1, \cdots, B_M$を考える．これらについての命題や推論規則は，すべて直積空間$X\times Y$のファジー部分集合として表現される．対象のXに関する状態を表わす変数をV，Yに関するものをUとすると，「VがA_iならばUはB_jである」という推論規則は，

(7) $\boldsymbol{R}(i,j)=[R_{qr}(i,j)];\ \begin{matrix}1\leq q\leq n\\ 1\leq r\leq m\end{matrix};$

$R_{qr}(i,j)=\max\{1-a_i(Xq), b_j(Yr)\}$

なる行列で与えられる．「VがA_iでUがB_jでない」場合に限り論理式が0の値をとるブール代数的表現を，ファジー集合論的に一般化したものである．同様に，「VはA_iである」とか「UはB_jである」といった命題は，$1\leq q\leq n, 1\leq r\leq m$，に対し，

(8) $\boldsymbol{F}(X,i)=[F_{qr}(X,i)]; F_{qr}(X,i)=a_i(X_q);$

$\boldsymbol{F}(Y,j)=[F_{qr}(Y,j)]; F_{qr}(Y,j)=b_j(Y_r)$

となる．命題や推論規則は，「かつ」であれば(5)から要素毎にminを，「あるいは」であれば(6)を用いてmaxをとることで結合され，最終的に得られる行列の行や列に関してmaxによる射影を行ない，それをさらに標準化することで結論が得られる．

簡単な例題として，ある産業ロボットに対し，$X=\{X_1, X_2, X_3\}$と$Y=\{Y_1, Y_2, Y_3, Y_4\}$なる全体集合を考えよう．ここでは$X_i=(l_i, l_{i+1})$ロボットの動きのある時間内における平均誤差範囲を表わし，$l_i<l_{i+1}$である．$Y_j=(t_j, t_{j+1})$は同じ時間内での製品の合格率の範囲を

表 1 Xに関するファジー部分集合

	X_1	X_2	X_3
KS	1.0	0.8	0.0
HS	0.4	1.0	0.6
FS	0.0	0.2	1.0

表 2 Yに関するファジー部分集合

	Y_1	Y_2	Y_3	Y_4
KR	0.0	0.0	0.8	1.0
HR	0.0	0.4	1.0	0.4
F	0.2	1.0	0.8	0.2
FT	1.0	0.6	0.0	0.0

表わし，やはり$t_j<t_{j+1}$である．Xに関しては，KS(きわめて正確)，HS(ほぼ正確)，FS(不正確) の3つ，Yについては，KR(きわめて良好)，HR(ほぼ良好)，F(普通)，FT(不調) の4つ，のファジー部分集合が**表1**，**表2**のごとくに与えられている．ここで，行和も列和も1ではないことに注意しよう．前と同じく，VとUを各々XとYの状態に関する変数とし，次のような規則がわかっているとする．

規則：VがKSならば，UはKRである．

あるいは，

規則2：VがKSかHSならば，UはKRかHRである．

さらに，

規則3：VがFSならば，UはFかFTである．

ロボットのXに関する現在の状態がHSかFSであるとして，製品の合格率を推測することが問題となる．

まず，規則1に対応する表を(7)から作成すると，**表3**のようになる．ついで，表1，表2と(6)から「KSかHS」と「KRかHR」に対応するファジー部分集合を構成し (**表4**，**表5**)，その結果に前述の手続を繰返して，規則2に対する表を作成する (**表6**)．「規則1あるいは規則2」に対応する**表7**は，表3と表6を要素毎にmaxで結合することで得られる．同じような手続をふんで，規則3に対応する**表8**を得る．さらに，表7と表8を要素毎にminで結合することで，規則全体を示す**表9**にたどりつく．「VはHSかFS」という命題は(8)によって**表10**で与えられ，これを表9と要素毎にminで結合することで最終的な**表11**を得る．列についてmaxで射影し，さらに標準化すると，Y_1からY_4に関する予測，28%，28%，22%，22%，が結論として導かれる．興味のある読者は，現在の状態を，「HSかFS」から「HS」へ，あるいは「FS」へ変えた時，常識に合致するように結論が変わるか否か，確かめてみていただきたい．

表 3 規則1

	Y_1	Y_2	Y_3	Y_4
X_1	0.0	0.0	0.8	1.0
X_2	0.2	0.2	0.8	1.0
X_3	1.0	1.0	1.0	1.0

表 4 $KS_\vee HS$

	X_1	X_2	X_3
$KS_\vee HS$	1.0	1.0	0.6

表 5 $KR_\vee HR$

	Y_1	Y_2	Y_3	Y_4
$KR_\vee HS$	0.0	0.4	1.0	1.0

表 6 規則2

	Y_1	Y_2	Y_3	Y_4
X_1	0.0	0.4	1.0	1.0
X_2	0.0	0.4	1.0	1.0
X_3	0.4	0.4	1.0	1.0

表 7 規則1あるいは規則2

	Y_1	Y_2	Y_3	Y_4
X_1	0.0	0.4	1.0	1.0
X_2	0.2	0.4	1.0	1.0
X_3	1.0	1.0	1.0	1.0

確率論的枠組みを離れた接近法としては，別に確実性理論にもとづくものがある．論理式の確かさに関して上限値と下限値を設け，これらの値を改変することで推論を進める2の方法は，医学的診断システム「MYCIN」を生み出すなど成果をあげているが，紙数の都合でここに詳述する余裕がない．

表 8 規則3

	Y_1	Y_2	Y_3	Y_4
X_1	1.0	1.0	1.0	1.0
X_2	1.0	1.0	0.8	0.8
X_3	1.0	1.0	0.8	0.2

表 9 規則全体

	Y_1	Y_2	Y_3	Y_4
X_1	0.0	0.4	1.0	1.0
X_2	0.2	0.4	0.8	0.8
X_3	1.0	1.0	0.8	0.2

表10 Vは$HS \vee FS$

	Y_1	Y_2	Y_3	Y_4
X_1	0.4	0.4	0.4	0.4
X_2	1.0	1.0	1.0	1.0
X_3	1.0	1.0	1.0	1.0

表11 結　論

	Y_1	Y_2	Y_3	Y_4
X_1	0.0	0.4	0.4	0.4
X_2	0.2	0.4	0.8	0.8
X_3	1.0	1.0	0.8	0.2
	1.0	1.0	0.8	0.8
	28%	28%	22%	22%

3. Dempster–Shafer 理論による接近

Dempster [1] が応用確率・統計で展開した考え方を，不確定要素を含む知識表現とその推論に持ち込んだのは，Shafer [7] である．今，$\Theta=\{\theta_1,\cdots,\theta_n\}$ を互い

に排反で全体をつくすような命題の集合とし，その巾集合を $B(\Theta)=\{A:A\subset\Theta\}$ と書く．便利のため，$B(\Theta)=\{A_i:0\leq i\leq M\}$ とし，$A_0=\phi, A_1=\Theta, |A_i|\geqq|A_{i+1}|, 1\leq i\leq M-1$ とする．$|\cdot|$は集合の要素数を表わす．ここで，確信度関数 $\mathrm{Bel}:B(\Theta)\rightarrow[0,1]$ を3つの公理によって次のように定義する．

（A 1） $\mathrm{Bel}(\phi)=0$

（A 2） $\mathrm{Bel}(\Theta)=1$

（A 3） $\mathrm{Bel}(\bigcup_{k=0}^{n} A_{i_k})\geq\sum_{\substack{I\subset\{0,1,\cdots,n\}\\ I\neq\phi}}(-1)^{|I|+1}\mathrm{Bel}(\bigcap_{k\in I} A_{i_k})$,

ここでnは$0\leq n\leq M$なる任意の整数．公理（A 3）は，不等号に置き換えることで，通常の確率公理系を一般化したものと考える．Θの部分集合Aに対し，$\mathrm{Bel}(A)$はAについての確信度を表わす．

巾集合上での離散的確率測度を $m:B(\Theta)\rightarrow[0,1]$ とすると，mとBelの間には，次の1対1対応関係が成立する．

(9) $\mathrm{Bel}(A)=\sum_{B\subset A} m(B)$;

$m(A)=\sum_{B\subset A}(-1)^{|A\setminus B|}\mathrm{Bel}(B)$.

したがって，mの方で推論に関する手順を構成しても，それをそのままBelへ変換できることになり，Dempsterの方法を取り入れる糸口がここに見出される．

考察の対象としてのΘに対し，今，2人の専門家が異なる判断を下したとしよう．具体的には，巾集合上での離散的確率測度m_1とm_2が与えられることになる．どうしたら2つの判断を統合できるか，という問題を解決するため，ShaferはDempsterの直和の概念を適用した．$C(m_1,m_2)$を

(10) $C(m_1,m_2)=1-\sum_{A_i\cap A_j=\phi} m_1(A_i)m_2(A_j)$

で定義し，$C(m_1,m_2)$が正である時m_1とm_2との直和$m=m_1\oplus m_2$を

(11) $m(A_k)=\sum_{i=0}^{M}\sum_{j=0}^{M}\delta_{(A_i\cap A_j=A_k)}m_1(A_i)m_2(A_j)/C(m_1,m_2)$

と定める．ここに$\delta_{(P)}$は命題Pが成立する時1の値をとり，それ以外は0の値をとる．このmがm_1とm_2とを統合した見解を表わす．$C(m_1,m_2)=0$ の場合，直和は定義されない．そのさい，m_1とm_2に関し，Θに対応する部分に微少量の確率を入れ込み，正の確率を持つ部分からそれに見合うだけ差し引く．すると，直和は可能となり，この微少量を零へ持っていった時の極限として，$C(m_1,m_2)=0$ の場合の直和を定義することができる．ただし，極限操作に関して連続性が成立せず，微少量の確率の入れ込み方，差し引き方に依存して，さまざまな極限値が現われてくる．詳しくは，[8] を参照されたい．

表12 $\Theta=\{\theta_1,\theta_2,\theta_3\}$ に関する m_1 と m_2

i	A_i	$m_1(A_i)$	$m_2(A_i)$
0	ϕ	0	0
1	Θ	α_1	α_2
2	$\{\theta_1,\theta_2\}$	0	0
3	$\{\theta_2,\theta_3\}$	0	$1-\alpha_2$
4	$\{\theta_1,\theta_3\}$	0	0
5	$\{\theta_1\}$	$1-\alpha_1$	0
6	$\{\theta_2\}$	0	0
7	$\{\theta_3\}$	0	0

直和の理解を深めるため，再び簡単な例題を考えよう．明日の天気に関し，θ_1＝“晴れ”，θ_2＝“曇り”，θ_3＝“雨”とし，$\Theta=\{\theta_1,\theta_2,\theta_3\}$と定義する．巾集合$B(\Theta)$は8つの要素を持ち，2つの判断$m_1$と$m_2$（**表12**）が与えられたとしよう．すると(10)から，

$C(m_1,m_2)=\alpha_1\alpha_2/(\alpha_1+\alpha_2-\alpha_1\alpha_2)$

となり，直和 $m=m_1\oplus m_2$ は次のようになるの．

$m(A_1)=\alpha_1\alpha_2/(\alpha_1+\alpha_2-\alpha_1\alpha_2)$

$m(A_3)=\alpha_1(1-\alpha_2)/(\alpha_1+\alpha_2-\alpha_1\alpha_2)$

$m(A_5)=(1-\alpha_1)\alpha_2/(\alpha_1+\alpha_2-\alpha_1\alpha_2)$

Shaferの考え方の特徴は，巾集合$B(\Theta)$上での離散的確率測度mをもって，不確定要素を含む事象に対する知識表現を与える点にある．常識的なΘ上での離散的確率測度を特別な場合として含むが，それより広く，$A\subset B$であっても$m(A)\leqq m(B)$である必要はない．特に，$m(\Theta)$は，全体集合に対する予測の確率配分である．全体集合の要素の1つが実現するのは確実だから，結果がΘのどれかであろう，という予測は何も言わないことと同じであり，結局，$m(\Theta)$は判断保留の度合いを示す．$m(\Theta)=1$であれば，いかなる判断をも下したくない，という態度の表明となる．この「判断保留の度合い」を陽に取り扱うことはShaferに独特のもので，他の方法では全く考慮されていない．

Dempster-Shafer理論の構造的性質を解明するため，住田・増田・舘・石川[8]は，直和を求める計算法に代数的表現を与え，それにもとづいてさまざまな結果を導き出した．特に，多数の意見を統合してゆく時，その極限がいかなる性質を持つかについての考察は興味深い．結果のいくつかを列記しよう．

(A) 専制君主が存在し得る.

他の意見がどうであっても，直和が可能でありさえすれば，統合された意見が自分の任意に指定する部分集合だけにしか正の確率を持たないようにすることができる.

(B) 統合された意見の不安定性

1人の意見によって，何千人もの意見が大きく変更され得る.

(C) 統合された意見の非代表性

統合された意見が，それを構成する個々の意見すべてとひどくかけ離れている，といった事態が起こり得る．もちろん，こうした事柄が常に生起するわけではないが，構造的な性質として理解しておくことは重要に思える.

4. ま と め

ある知識が不確定要素を持つ，という時，そこにはさまざまな「不確定性」が錯綜していると考えられる．まず，純粋に確率的に記述し得る不確定性がある．ついで，言語表現のあいまいさに付随する不確定性もあるかもしれない．機械の性能が「すばらしい」という実感を，厳密に物理的な量に対応させるのは難しい．さらに，情報不足による不確定性もあるに違いない．本稿で概説してきた3つの流れは，それぞれにこうした不確定性の一面を照射しているように思える.

不確定性が確率的に記述できる要素を強く持つものであれば，ベイズ理論による接近は自然であろう．言語表現の持つ曖昧さの把握が，知識表現を左右するほどに重要であれば，可能性理論や確実性理論は魅力あるものとして目に映るに違いない．情報不足，知識不足による判断保留を考慮することが大切であれば，Dempster-Shafer 理論は強力な援軍となる.

ベイズ理論をDempster-Shafer理論の立場から解釈し直す試み，あるいはベイズ理論やDempster-Shafer理論をファジー化する試み，など理論統一への動きは見られるが，いまだ根底的なものとはなっていない．論争は，ここ当分はつづくと思われる．重要なことは，各接近法の特質を理解し，自分の持つ問題がはらむ不確定性の内容を把握し，それらを照し合わせることを通して，その問題にふさわしい手法を取捨選択してゆくことであろう．3つの流れをより詳しく理解したいと思われる読者のために，比較的入手しやすい本[2],[5],[7]を参考文献として挙げておく.

参 考 文 献

[1] Dempster, A. P. (1967), "Upper and Lower Probability Induced by a Multivalued Mapping", *Annals of Mathematical Statistics*, Vol. 38, pp. 325-339.

[2] Dubios, D. and Prade, H. *Possibility Theory : An Approach to Computerized Processing of Uncertainty*, Plennum Press, New York.

[3] Lazarsfeld, P. F. (1966), "Latent Structure Analysis", in Measurement and Prediction, S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, J. A. Clausen (Eds), John Wiley, New Yovk.

[4] Pearl, J. (1986), "Fusion, Propagation, and Structuring in Belief Networks", *Artificial Intelligence*, Vol. 29, pp. 241-288.

[5] Pearl, J. *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*, Morgan Kaufmann Publishers, Inc., San Mateo. California.

[6] Sackar. S. (1991), "Probabilistic Modeling of Belief Networks in Expert Systems", Ph. D. Thesis in preparation, The Simon School, the University of Rochester, Rochester, New York.

[7] Shafer, G. (1976), *A Mathematical Model of Evidence*, Princeton University Press, Princeton, New Jergy.

[8] 住田・増田・舘・石川 (1989), "Structural Analysis of Dempster-Shafer Theory and Related Limit Theorems", *International Journal of Expert Systems*, Vol. 2, No. 2, pp. 163-198.

[9] Zadeh, L. A. (1965), "Fuzzy Sets", *Information and Control*, Vol. 8, pp. 338-353.