9D－7

証拠の蓄積が形成する信念論理に基づく知識獲得

村井哲也[1)]　宮腰政明[2)]　新保　勝[2)]

[1)]北海道教育大学　[2)]北海道大学

1．まえがき

人工知能では近年，信念の取り扱いが一つの焦点である．中でも，論理に基づく形式的方法(知識と信念の様相論理など)と測度に基づく数値的方法(ベイズ理論やDempster－Shaferの証拠理論[6])が研究されており，それぞれ一長一短がある．前者は既成の知識や信念の集合からの推論を扱うことができる．しかし，知識や信念自体の形成過程を説明できない．一方，後者はベイズ規則やDempsterの結合規則などで信念形成を説明する．しかし，推論との関係はあまり明白でない．

最近，我々はDempster-Shaferの証拠理論（以下，D-S理論と略す）に基づく様相論理[2]の有限モデルを構成し，いくつかの様相論理体系との健全性と完全性を証明した[4,5]．この結果は証拠の蓄積が知識や信念の論理と密接に関係することを示し，論理的枠組の中での信念形成と知識獲得の実現可能性を示唆する．

本研究では，我々は日常得られる様々な証拠に基づいて信念を形成し，確実性のある信念を知識として我々の認識体系に組み込んで行くという観点に立ち，証拠の蓄積による信念形成および知識獲得を定式化する．第２章では本研究で定式化する知識獲得の考え方を述べ，第３章でその定式化に必要な可能世界限定モデルを説明する．第４，５章ではそれぞれ，可能世界限定モデルにおける信念形成と知識獲得について論じる．

2．基本的考え方

我々は真偽が分からない命題に対して，様々な証拠を集め，その真偽を確実なものとしてから，自身の認識体系に組み込んで行くと考えられる．これは証拠の蓄積に基づいて知識を獲得する過程とみなせる．

例えば，ある犯罪に対して，A，B２名の容疑者が浮かび上がったとする．A，Bがこの犯罪の犯人であるという命題をそれぞれp，qで表わすとき，現実世界を部分的に記述する候補として

α_1：　$p \wedge q$　$(=\{p, q\})$
α_2：　$p \wedge \neg q$　$(=\{p\})$
α_3：$\neg p \wedge q$　$(=\{q\})$
α_4：$\neg p \wedge \neg q$　$(=\phi)$

の４つが論理的に可能である．これらはCarnap[1]のいわゆる状態記述であり，本論文では可能世界，または単に世界と呼ぶ．犯人に関する情報が何もなければ４つの世界すべてが現実世界として可能である．容疑者AとB以外に犯人がいないという証拠は論理式$p \vee q$を支持する．$p \vee q$を支持する証拠が確実なものであれば，α_4の可能性はなくなり，現実世界の候補は

$\{\alpha_1, \alpha_2, \alpha_3\}$

に限定される．また，単独犯であるという確実な証拠は論理式$\neg p \vee \neg q$を支持し，候補を

$\{\alpha_2, \alpha_3, \alpha_4\}$

に限定する．これら２つの証拠をあわせれば，現実世界の候補は

$\{\alpha_2, \alpha_3\}$

まで限定できる．このように様々な証拠が現実世界の候補を限定し，候補となる世界が一つに絞られたとき，その世界は現実世界の確実な部分的記述，すなわち知識として認識体系に組み込まれると考えられる．以上が証拠の蓄積に基づく知識獲得の基本的考え方である．

3．可能世界限定モデル

前章で述べた可能世界の限定という考え方を定式化するために，測度に基づく様相論理[4,5]，中でもD-S理論で展開されたplausibility関数に基づく様相論理[3,4]を利用する．まず，文の集合である言語$\mathcal{L}$を

$\mathbf{P}$：原子文の有限集合
演算子：$\top$(恒真)，$\neg$(否定)，$\vee$(選言)，$\mathbf{B}$(信念)
かっこ：(，)

から通常通り形成する．演算子$\bot$(恒偽)，$\wedge$(連言)，$\rightarrow$(質料含意)，$\leftrightarrow$(同値)も通常通り定義する．

可能世界の集合をD-S理論の識別の枠とみなしてモデルを定義する．可能世界限定モデルとは

$\mathcal{M}=\langle W, \mathrm{bpa}\rangle$

である．ここで

$W=\mathfrak{P}(\mathbf{P})$：$\mathbf{P}$のべき集合
$\mathrm{bpa}：\mathbf{P}(W)\rightarrow[0,1]$　基本確率割当[6]

である．Wの元を可能世界，または単に世界と呼ぶ．

文pがモデル$\mathcal{M}$の世界αで真であることを「$\models^{\mathcal{M}}_{\alpha} p$」で表わす：

$\models^{\mathcal{M}}_{\alpha} p \iff p \in \alpha$　（p：原子文）

「$\models^{\mathcal{M}}_{\alpha}$」は$\mathbf{B}$を除いて通常通り複合文へ拡張される：

$\models^{\mathcal{M}}_{\alpha} \top$
$\models^{\mathcal{M}}_{\alpha} \neg p \iff \models^{\mathcal{M}}_{\alpha} p$でない
$\models^{\mathcal{M}}_{\alpha} p \vee q \iff \models^{\mathcal{M}}_{\alpha} p$　または　$\models^{\mathcal{M}}_{\alpha} q$

モデル$\mathcal{M}$のすべての世界でpが成り立つとき，pは$\mathcal{M}$において妥当であるといい，「$\models^{\mathcal{M}} p$」で表わす．

可能世界限定モデルでは，現実世界の候補を限定する情報を証拠と見なして基本確率割当に変換する．例えば，前章で論じた$p \vee q$を支持する証拠は

$$\mathrm{bpa}(X)=\begin{cases}1 & X=\|p \vee q\|=\{\alpha_1, \alpha_2, \alpha_3\}\\ 0 & \text{その他}\end{cases}$$

に変換される．ここで，文pに対して，$\|p\|$はpが成り立つ世界の集合を表わす．したがって，システムが持つモデルは固定的なものではなく，新たな証拠の獲得とともに結合規則によって変化する．

基本確率から導かれるplausibility関数を利用して，信念演算子$\mathbf{B}$を定義する[3,4]：

$$\models^{\mathcal{M}}_{\alpha} \mathbf{B}p \iff \mathrm{Pl}(\|p\|)=\sum_{\|p\|\cap X\neq\phi} \mathrm{bpa}(X)=1.$$

信念演算子$\mathbf{B}$は推論規則

RE．$p\leftrightarrow q \implies \mathbf{B}p\leftrightarrow\mathbf{B}q$

を満たした上，公理

M．　$\mathbf{B}(p\wedge q)\rightarrow(\mathbf{B}p\wedge\mathbf{B}q)$
N．　$\mathbf{B}\top$
P．　$\neg\mathbf{B}\bot$　　$(\bot\leftrightarrow\neg\top)$
4．　$\mathbf{B}p\rightarrow\mathbf{B}\mathbf{B}p$
5．　$\neg\mathbf{B}p\rightarrow\mathbf{B}\neg\mathbf{B}p$

を妥当とし，体系$EMNP45$の演算子となる[3,4]．ここで，公理等の名称はChellas[2]に従った．尚，諸体系$E\cdots$はミニマル・モデルに対応する様相論理体系

Evidence-Theory-Based Doxastic Logic and Knowledge Acquisition

T.Murai[1)], M.Miyakoshi[2)], and M.Shimbo[2)]
[1)]Hokkaido University of Education
[2)]Hokkaido University

であり，クリプキ・モデルに対応する諸体系K･･･を特別な場合として含む(cf.[2]).

Plausibility関数の特別な場合として，Zadehの可能性測度および確率，Dirac測度があり(cf.[7]),それぞれ特徴ある公理を妥当とする[4]．まず,可能性測度は基本確率の焦点要素が入れ子であるplausibility関数であり，これに基づく**B**は先の公理に加えて

F． $\mathbf{B}(p \vee q) \to (\mathbf{B}p \vee \mathbf{B}q)$

D_c． $\neg\mathbf{B}\neg p \to \mathbf{B}p$

を妥当とし，体系$EMFNP45$の演算子である．確率に基づく**B**も先の公理に加えて

C． $(\mathbf{B}p \wedge \mathbf{B}q) \to \mathbf{B}(p \wedge q)$

K． $(\mathbf{B}p \to q) \to (\mathbf{B}p \to \mathbf{B}q)$

D． $\mathbf{B}p \to \neg\mathbf{B}\neg p$

を妥当とし，体系$KD45$の演算子である．Dirac測度は可能性測度であると同時に確率でもあり，それに基づく**B**は以上すべての公理を満たし,体系$KD!45$の演算子となる．

4．信念形成

4．1．証拠が形成する信念集合の性質

可能世界限定モデル$\mathcal{M}$における信念文の集合$\mathfrak{B}$を

$$\mathfrak{B} \triangleq \{p \in \mathcal{L} \mid \models^{\mathcal{M}} \mathbf{B}p\}$$

によって定義する．$\mathfrak{B}$の性質は演算子**B**の性質，したがって基礎となるplausibility関数の性質に依存する．

まず，一般のplausibility関数が構成する$\mathfrak{B}$に対して，演算子**B**の性質から

$\models^{\mathcal{M}} p \leftrightarrow q$ かつ $p \in \mathfrak{B} \Longrightarrow q \in \mathfrak{B}$

$p \wedge q \in \mathfrak{B} \Longrightarrow p, q \in \mathfrak{B}$

$\top \in \mathfrak{B}$

$\perp \notin \mathfrak{B}$

$p \in \mathfrak{B} \Longrightarrow \mathbf{B}p \in \mathfrak{B}$

$p \notin \mathfrak{B} \Longrightarrow \neg\mathbf{B}p \in \mathfrak{B}$

が成り立つ．しかし，一般に演算子**B**は公理KおよびDを満たさないので，$\mathfrak{B}$において，それぞれ

$p, p \to q \in \mathfrak{B} \Longrightarrow q \in \mathfrak{B}$

$p \in \mathfrak{B} \Longrightarrow \neg p \notin \mathfrak{B}$

は成り立たない．したがって,一般に$\mathfrak{B}$はモダス・ポネンスに関して閉じておらず，しかも

$p, \neg p \in \mathfrak{B}$

である可能性もある．それにもかかわらず，$\perp \notin \mathfrak{B}$であることは，矛盾した信念を持つことと矛盾自体を信じることは別であることを示す．

可能性測度および確率に対しては，更に，それぞれ

$p \vee q \in \mathfrak{B} \Longrightarrow p \in \mathfrak{B}$ または $q \in \mathfrak{B}$

$\neg p \notin \mathfrak{B} \Longrightarrow p \in \mathfrak{B}$

および

$p, q \in \mathfrak{B} \Longrightarrow p \wedge q \in \mathfrak{B}$

$p, p \to q \in \mathfrak{B} \Longrightarrow q \in \mathfrak{B}$

$p \in \mathfrak{B} \Longrightarrow \neg p \notin \mathfrak{B}$

が成り立つ．確率が構成する$\mathfrak{B}$はモダス・ポネンスに関して閉じており，しかも矛盾した信念を持たない．一方，可能性測度が構成する$\mathfrak{B}$はすべての$p \in \mathcal{L}$に対し

$p \in \mathfrak{B}$ または $\neg p \in \mathfrak{B}$

が成り立つのが特徴である．Dirac測度が構成する$\mathfrak{B}$は前述のすべての性質を持つ．

4．2．信念形成の非単調性

新たな証拠が見つかれば,今までの証拠とDempsterの結合規則によって結合される．この際，信念の対象であった文が信念から外れることがある．例えば，第３章で例示した容疑者ＡとＢ以外に犯人がいないという証拠を表わす基本確率の下で

$p \wedge q$， $p \wedge \neg q$， $\neg p \wedge q \in \mathfrak{B}$

となる．新たにＡが犯人である強い証拠

$$\mathrm{bpa}'(X) = \begin{cases} 0.8 & X = \|p\| = \{\alpha_1, \alpha_2\} \\ 0.2 & X = W \\ 0 & \text{その他} \end{cases}$$

を得たとき，先の基本確率と結合すると新たに

$$\mathrm{bpa}''(X) = \begin{cases} 0.8 & X = \|p\| = \{\alpha_1, \alpha_2\} \\ 0.2 & X = \|p \vee q\| = \{\alpha_1, \alpha_2, \alpha_3\} \\ 0 & \text{その他} \end{cases}$$

を得る．この基本確率の下では

$p \wedge q$， $p \wedge \neg q \in \mathfrak{B}$， $\neg p \wedge q \notin \mathfrak{B}$

となり，Ｂの単独犯行であるという信念が外れる．このように，新たな証拠による基本確率の更新は信念文の集合を非単調に形成する．

5．知識獲得

第２章で論じたように，本研究で採用する信念が知識に転化する規準は現実世界の候補世界が１つに絞られることである．Plausibility関数が世界を１つに絞るための必要十分条件は

$$\cap\{X \mid \mathrm{bpa}(X) > 0\} = \{\alpha\}$$

となるただ１つの世界αが存在することである．このとき，plausibility関数が構成する演算子**B**は公理KおよびDを満たすので,信念文の集合$\mathfrak{B}$はモダス・ポネンスに関して閉じ，しかも矛盾した信念を持たない：

$p, p \to q \in \mathfrak{B} \Longrightarrow q \in \mathfrak{B}$

$p \in \mathfrak{B} \Longrightarrow \neg p \notin \mathfrak{B}$.

また，原子文$p \in \mathbf{P}$に対して

$p \in \mathfrak{B} \Longleftrightarrow p \in \alpha$

$\neg p \in \mathfrak{B} \Longleftrightarrow p \notin \alpha$

が成り立つので，すべての原子文pに対して

$p \in \mathfrak{B}$， $\neg p \notin \mathfrak{B}$

のいずれか一方が成立する．知識の整合性の観点から，これらの性質は信念集合$\mathfrak{B}$が知識の集合としての資格を備えたことを示すと考えられる．例えば，前章第２節の例に加えて，更にＢが犯人でない証拠

$$\mathrm{bpa}'''(X) = \begin{cases} 0.7 & X = \|\neg q\| = \{\alpha_2, \alpha_4\} \\ 0.3 & X = W \\ 0 & \text{その他} \end{cases}$$

を獲得し，先の基本確率割当bpa”と結合して

$$\mathrm{bpa}''''(X) = \begin{cases} 0.7 & X = \{\alpha_2\} \\ 0.24 & X = \{\alpha_1, \alpha_2\} \\ 0.06 & X = \{\alpha_1, \alpha_2, \alpha_3\} \\ 0 & \text{その他} \end{cases}$$

を新たに得たとき，この基本確率割当は

$$\cap\{X \mid \mathrm{bpa}(X) > 0\} = \{\alpha_2\}$$

を満足するので，候補世界はα_2に絞られ,Ａの単独犯であることが知識として獲得される．このとき，原子文p，qに対して

$p, \neg q \in \mathfrak{B}$， $\neg p, q \notin \mathfrak{B}$

が成り立ち，原子文に対する信念が矛盾なく確定し，信念集合$\mathfrak{B}$は知識としてふさわしい整合性を持つ．

文　献

[1] R.Carnap, Logical Foundations of Probability. The University of Chicago Press, 1962.

[2] B.F.Chellas, Modal Logic: an Introduction. Cambridge University Press, 1980.

[3] 村井,宮腰,新保,Dempster-Shafer理論に基く知識と信念の論理．情報処理学会第44回全国大会講演論文集(3), 1992, pp.67-68.

[4] 村井,宮腰,新保，種々のファジィ測度の族が決定する様相論理体系．第８回ファジィ・システム・シンポジウム講演論文集, 1992, pp.29-32.

[5] T.Murai, M.Miyakoshi, and M.Shimbo, Measure-Based Semantics for Modal Logic. R.Lowen and M.Roubens(eds.), Fuzzy Logic：States of the Art, Kluwer Academic, 1992, to appear.

[6] G.Shafer, A Mathematical Theory of Evidence. Prinston University Press, 1976.

[7] 菅野，室伏，ファジィ測度論入門[１]． 日本ファジィ学会誌，2(1990), pp.174-181.