

# 学習と推論の不可分性について

～IPSモデルのここが好きここが嫌い～

石川 孝 ベンテる(株)電子研究所

## 1. はじめに

$5x = 10$  という方程式を解くのは容易い。しかし一人の人間が解けるようになるには約10年を要し、人類が方程式の概念を作りだすまでには数百万年を要している。さらに代数的解法では解けない方程式が限りなくあることを我々は知っている。このことから明らかなように、人間の推論は高速でも正確でもない。知能は予測できない環境に対処するための生物の機能である。この機能を実現している機構は計算機と比べると非効率的であるが、これは知能の機能からの必然的要請であると思われる。なぜなら環境の変化のために知能にとって絶対真理はありえないので、そのときの環境を反映する知識の学習と知識による推論とは必然的に不可分であると考えられる。

知識ベースシステムという人工知能のパラダイムは、人間は情報処理システムとして機能するというIPS (Information Processing Systems)モデルに基づいている[1]。IPSモデルでは、事物をシンボルによって表象し、そのシンボルを操作する。IPSモデルでは処理と対象とが分離されて認識されているが、この描像はデジタル計算機には当て嵌まっても、生物に当て嵌まるといえるだろうか？ここではこの設問に対して、学習と推論の不可分性という視点から考察してみる。

## 2. 処理と対象は分離できるか

IPSモデルはデジタル計算機からのアナロジーである(図1)。このモデルでは、環境中の事物をシンボルによって表象し、そのシンボルを操作することによって環境に作用する。したがってシンボルによる問題解決に対する初期の人工知能の表面的な成功は、デジタル計算機の有効性と同時に限界を示した。その後のAIの

発展によっても、状況は本質的に変わっていない。何かの間違っているのである。

そもそも知能における処理と対象とは分離できるのだろうか。知能の本質が環境からの知識の学習にあるとするなら、学習における処理と対象の関係を見直す必要がある。学習とは環境の一部である自分自身を制御することであると見ることができるが、この場合学習主体と学習の対象は同一であり、概念的に分離できても実体として分離することはできない。知識は自分自身の推論に関する形式化(概念化)であるから、概念の世界では知能主体を離れて知識が存在するかのように見えても、実体としては分離できないのである。新しい知識ベースシステムのアーキテクチャのヒントはこの辺にあり、AMやEURISKOの成功が"The benefit of having syntax mirror semantics"にあるという指摘[2]は当たっていると思われる。また、精神的過程は意識が存在しなければ起こらず、意識は世界の一部としての自分を制御するためのモデルという考え方[3]も行き詰まりの解決を暗示する。

## 3. 学習と推論は分離できるか

知識ベースシステムにおいて処理と対象とが分離できないとしたら、学習システムのモデル[4](図2)は正しいモデルといえるだろうか？知識と知能主体との関係をCarbonellの仮説[5]をもとに考察しよう。「問題解決と学習は、統一的な認知機構の切り離せない両側面である。別の言葉でいえば、問題解決を行うことなしに必要な認知技法を獲得することはできない。まさに問題解決のプロセス自体が問題解決の技法を獲得し改良するための必要な情報を与えてくれる。」この仮説は学習と問題解決とが切り離せないことを主張しているが、この仮説を認めるとすればなぜそうなのだろうか？

学習アーキテクチャのいくつかを見てみると、例えば Soar[6]では「問題空間計算モデルにおいては、学習は問題空間の間での知識の移転であり」、PRODIGY[7]では「類推的再生による学習は記憶管理機構との相互作用による」とし、記憶を学習の本質的要素と考える選択による学習[8]においても「学習は選択=枝狩りであり、脳の中の超計量性と階層性が知識の本質である」としている。これらのアーキテクチャの本質は、問題解決=問題表現の発見、つまり学習とは表現を見つけることの問題解決[9]、または効率のよい解法を与える表現の発見=表現空間における発見的探索[10]であると見ることができる。さらに、意識、言語、知能の間の関係はつぎのように整理される[11]。「記憶と意識は密接に関連している：ある意味では、記憶の検索が意識なのである」、「知能とは主として表現法を作りだしそれを操作する能力である」、「推論：知識に基づく問題解決 $\leftrightarrow$ 表象の利用」、「学習とは計画的に知識構造を変化させること～状況に応じて自己を修正する能力である」。このように見てくると、学習システムのモデルは、環境がシステム自身を含むように修正する必要がある(図3)、IPSモデルにおける処理装置は環境の一部としてのシステム自身の概念に過ぎない(図4)のではないだろうか。

#### 4. おわりに

人工知能パラダイムとしてのIPSモデルの妥当性について考察したが、これからの人工知能研究はなにを目指すべきであろうか？知能の本質が「表現法を作り出す創造性」にあるとすれば、研究の主題はその機構の解明であろう。まったく類似性が存在していない二つの対象の間に類似性を創造していく解釈的類推[13]はその手掛かりであり、筆者らは類推による学習を基礎とした学習機構の研究を進めている[14]。

#### 参考文献

- [1] Newell, A. and Simon, H. A., "Human Problem Solving", Prentice-Hall (1972)  
 [2] Lenat, D. B. and Brown, J. S.: Why AM and

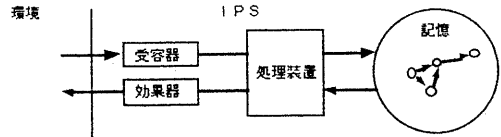


図1. 情報処理システムの一般的構造 [1]

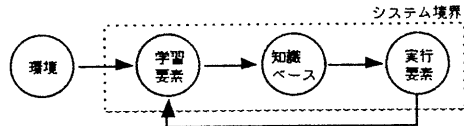


図2. 機械学習システムの単純化したモデル (人工知能ハンドブック)

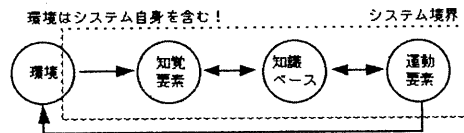


図3. 生物システムの単純化したモデル

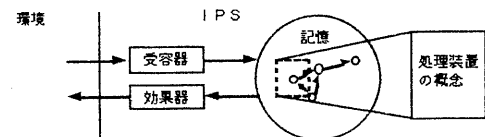


図4. 知能のリフレクティブなモデル

- EURISKO Appear to Work?, Artificial Intelligence, Vol.23, No.3, pp.269-294 (1984)  
 [3] Granovskaya, R. M. and Bereznaya, I. Y.: Consciousness as the Unity of Higher Psychic Processes, Kybernetes, Vol.17, No.2, pp.35-43 (1988)  
 [4] Cohen, P. R. and Feigenbaum, E. A. 編、「人工知能ハンドブック 第三巻」、共立出版 (1884)  
 [5] Carbonell, J. G., 類推学習：過去の経験による計画の定式化と一般化、Michalski, R.S. 他編、「学習と問題解決 (知識獲得と学習 2)」, 共立出版 (1987)  
 [6] Newell, A., "Unified Theories of Cognition", Harvard (1991)  
 [7] Veloso, M. M. and Carbonell J. G.: Learning by Analogical Replay in PRODIGY: First Result, Lect. Notes Comput. Sci., Vol.482, pp.375-390 (1991)  
 [8] Toulouse, G., Dehaene, S., and Changeux, J-P.: Spin Glass Model of Learning by Selection, Proc. Natl. Acad. Sci. USA, Vol.83, No.6, pp.1695-1698 (1986)  
 [9] Amarel, S.: Problems of Representation in Heuristic Problem Solving, Role Lang. Probl. Solving 1, pp.11-32 (1985)  
 [10] Korf, R. E., Toward a Model of Representation Changes, Artif. Intell., Vol.14, No.1, pp.41-78 (1980)  
 [11] M. A. フィッシュラー/O. ファーシャイン著、人と機械の知能。アジソンウェスレイトッパン (1989)  
 [13] Indurkha, B., On the Role of Interpretive Analogy in Learning, New Gener. Comput., Vol.3, No.4, pp.385-402 (1991)  
 [14] 石川 孝、寺野隆雄：戦略構造からの類推による問題解決学習、1992年度人工知能学会全国大会 (1992)