

一般事前知識についての考察 -究極の AGI を目指して-

Speculation about the generic priors for building Physical AI

山川宏¹

Hiroshi Yamakawa¹

¹株式会社ドワンゴ ドワンゴ人工知能研究所

¹Dwango Artificial Intelligence Laboratory, DWANGO Co., Ltd.

Abstract: Artificial general intelligence (AGI) needs machine learning technology acquiring shared knowledge within task region from data and generic prior knowledge. Feasible ultimate AGI will be utilized various constraints of the physical world. Generic prior knowledge, which is based on physical constraints as well as information theory, is considered in this paper.

1. はじめに

近年の急速な人工知能(AI)技術の発展により、既に特定の問題に対処する特化型 AI はしばしば人間の能力を上回りつつある。この背景にはタスク領域毎に安定して存在する知識（以下では”領域内共有知識”と呼ぶ）をうまくシステムに作り込むことの重要性を示しており、同時に設計する知識がシステムの適用領域を制限している。

こうした背景から、経験を積むことで人間のように十分に広範な適用範囲と強力な汎化能力を持つ汎用人工知能 (AGI : Artificial General Intelligence) の将来性にも目が向けられつつある[1]。つまり AGI は多様な問題領域において多角的な問題解決能力を自ら獲得し、設計時の想定を超えた問題を解決できる AI である。AGI はマーク・グブルド氏 (Mark Gubrud) により 1997 年に導入されベン・ゲーツェル氏 (Ben Goertzel) により広められた人間レベルの AI の実現に向けた技術的な目標である。

本稿では AGI 構築にむけて注力すべき研究トピックを明らかにするために、後述する一般事前知識を中心とした議論を行う。

2. 物理制約下における究極の知能

知的システムに限らず、設計者が領域内共有知識を把握していれば、それを設計時に組み込んで特化させたシステムほど領域内では良い性能を得られる。これはノーフリーランチ定理にも関わる。

一方であらゆる知的プログラムの可能性を組み込んだ、ユニバーサル AI [2]は現実的な計算量で実行できない理論的な存在である。

以上の二つの事実から、実用的な性能をもつ AI

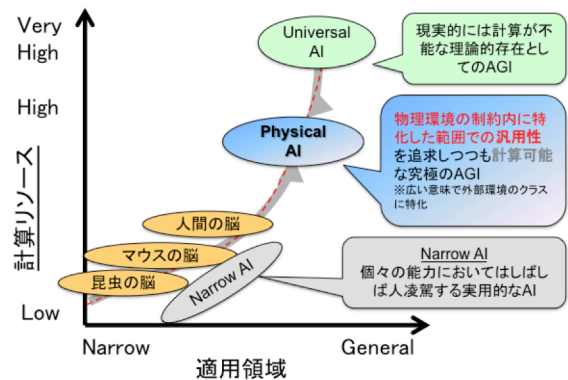


図 1: 実現可能な究極の AGI としての物理的 AI

の実現に必要な計算リソース(データ、計算量)は、理想的なアルゴリズムを用いたとしても、適用領域の拡大に伴って急速に増加するであろう。つまり特化型 AI は領域を狭めることで、少ない計算量で実用的な性能を得ているのである。また人間の脳は、Narrow AI や他の動物に比べて高い汎用性を持つとはいえ、身体的性質や進化の制約からその汎用性はかなりの制限を受ける。

そこで人工知能研究が目指すべき究極の知能は、物理的な制約に特化した範囲内で最大限の汎用性を実現することで計算可能なレベルにダウンスケールした知能と考えられる。私は、これを物理的 AI(Physical AI)と呼ぶことにした。

すると人間を模することで AGI に迫るアプローチを取る限り、その到達点は、物理的 AI の汎用性を下回る。しかし人間と同レベルの AGI は、AGI 自体の開発を加速できるので、実は物理的 AI 構築に向けての良いサブゴールになっている。

3. 領域内共有知識とその学習

私は、AGI が特化型 AI と異なる性質は「領域内共有知識の学習能力」であり、それはより一般的な事前知識と領域固有のデータの組み合わせにより実現されると考えている。ここで領域内共有知識は、特定の領域内においては共通性の高い知識であり、それに引き続く学習プロセスにとっては事前知識となるものである。例として視覚情報処理の場合に、レチノトピーにもとづいて階層的に局所から全体の処理を行う構造は領域内共有知識である。

領域内共有知識を十分に作りこまれた特化型 AI は高い性能を得られる。しかし一般的な事前知識だけを利用しては良い性能は得られない。よって、

- 適用領域が広いと領域内共有知識は減少し、領域を狭めるほど領域内共有知識は豊かになる。
- AI は領域内共有知識の量が増えるほどに実行や学習の性能を高めやすい
- AGI の開発では領域内共有知識の作り込みは最小限に留められる

こうした前提から、AGI が高い性能を持つためには、領域内のデータを用いて領域内共有知識を獲得する必要がある。よって以下が言えることになる、

【結論 1】AGI の特化型 AI と異なる性質は、領域内共有知識の学習能力である。

では AGI は、事前に設計していない領域での対応能力を向上するために、どの様に領域内共有知識を獲得するのだろうか。以下の前提が想定できる。

- 一般的に、学習による知識の獲得は、事前知識とデータの組合せによって実現される。
- 領域内共有知識として想定される仮説空間は非常に大きい。

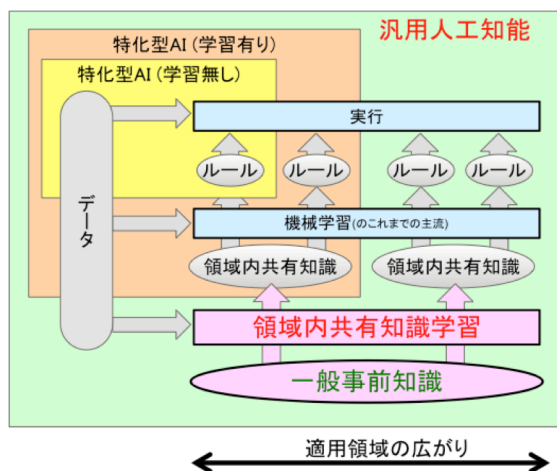


図 2: 汎用人工知能と領域内共有知識

- AGI においても十分に一般性の高い知識（一般事前知識）は設計してもよい
こうした前提から、以下が言えるであろう。

【結論 2】領域内共有知識は、より一般的な事前知識と領域固有のデータの組み合わせた学習プロセスにより獲得される。

以上の結論 2 に立脚するならば、AGI の主要な研究は以下の二つに分解されることになる

一般事前知識: システムが対処すべき世界内で共有される事前知識の研究。この適用範囲が AGI が機能しうる範囲を規定する

領域内共有知識学習: 一般事前知識と領域固有のデータから領域内共有知識を獲得する研究。

J. B. Tenenbaum 氏は「学習者は、”学習を可能とするための知識”をどのように学ぶのだろうか」[3]と述べたが、ここで述べている領域内共有知識学習は文末尾における「学ぶ」に相当するであろう。

この見方をさらに図 2 を用いて説明する。まず機械学習以前の特化型 AI では、人間の設計したルールに従って得られたデータに対して動作していた。そして現在の機械学習を備えた特化型 AI は人手で設計した領域内共有知識にもとづいてルールを獲得／変更する能力をもつ。対して AGI はそうした領域内共有知識さえも、一般事前知識とデータから抽出しうる技術である。この歴史的変化は、計算リソースが増大することで、機械が学習できる範囲が拡大していることに伴うであろう。

例えば初期視覚処理における並進不変性／回転不変性をもつ特徴量は、領域内共有知識を含んでいる。この特徴量を後述するような時間変化の局所性と連続性に関わる物理的基盤をもつ一般事前知識をもとにデータから獲得[4]することは領域内共有知識学習の一例といえるだろう。

如何なる AGI も、一旦領域内共有知識を獲得してしまえば、その先は特化型 AI の機械学習と同様に動作に違いはみられない(図 2 参照)。しかし領域内共有知識が自動学習されるなら、コスト面等の理由により、人出による設計が難しい領域にまで機械学習の適用範囲を拡大可能だろう。Goertzel 氏は、AGI の実現方法は特化型 AI と本質的に異なる (bottom qualitatively different) とも述べているが、その違いは意外にわかりづらいかもしれない。

4. 一般事前知識の考察

AGI が良い性能をもつためには、以上述べたよう、データから領域内共有知識を学習する必要がある。

その学習効率を高めるためには、制約をできるだけ多く「一般事前知識」として盛り込むべきである。

帰納推論に用いられる最も一般性の高い事前知識は、学習理論などでも現れる「単純さ」などである。しかし本稿では我々の住む物理空間を想定世界とした究極の知能(物理的 AI)を考えるため、物理的基盤から議論を出発する。こうした一般事前知識は、システム中に自然な設計として組み込まれる場合に、その存在自体に気づきづらいこともある。

4.1 一般事前知識の物理学的基盤

物理的 AI では生物の進化が辿った自然な外部環境を一般事前知識の前提とする。よって古典物理学の範囲における物理法則や性質を基盤とする。

以下で時間と空間の側面から物理的基盤を考察する。何れにおいても連続性と局所性は重要である。

§1 時間についての一般事前知識

伝統的な時間概念の主要な特性は、あらゆる場所で淀みなく流れてゆくものであり、「連続性」と「一方向性(不可逆性)」がある。哲学的にはクロノス時間と呼ばれるものである。以下では知能が時間に密接に関係することを確認する。

知能を生物の様な「生存し続ける情報」という観点から見ると、その目的自体が時間概念を前提としている。また知能システム一般にアルゴリズムは時間を追って後戻りすることなく手順を実行するという前提が必要だし、タスクや環境に適應するための進化や学習などの概念も、時間にそって機能や構造が改善するという前提に基づいている。

また様々な概念/構造/関係性/コンテキスト/シーン等において、共通の時刻 t を変数に含み時間変化を記述する多種多様な状態 $s(t)$ が用いられる。

§2 空間についての一般事前知識

物理世界において、時間概念と対をなす空間概念は、時間と同様に「連続性」が仮定されるが、時間と異なり任意の方向に進むことができる。

日常的な世界を支配する電磁相互作用においては、その影響は空間中を近接相互作用として伝搬するため、空間には局所性が生み出される。これにより空間的に近い領域からの観測は類似しやすいという空間的連続性の事前知識が得られる。

以上のように、物理世界においては時空間において局所性と連続性を持つため、帰納推論の有効な範囲も局所性を持つ。なお局所性は身体側と外部世界の両方にあらわれる。また不確実性を表す場合には、しばしば局在する連続分布(ガウス分布等)でモデル化することで物理的な性質を反映している。

4.2 帰納推論のための一般事前知識

AI において推論は、演繹推論、帰納推論、アブダクションがあるが、機械学習に関連し、かつ、動物/人間/計算機の何れもが推論の起点として用いるのは帰納推論であることから、これに着目する。

以下ではまず物理的な制約にもとづく一般事前知識を、AI の主体側と外部環境の二つの側面に分けて考察する。その上で改めて、単純さ、多様さ、不変性といった物理的性質によらない一般事前知識はについて検討する。

§1 一般的な身体性としての一般事前知識

帰納推論を行うには比較可能(同一性の判定可能性を含む)な値の集合としての変数が必要である。

さらに多変数の場合には、そこで扱われる変数を互いに指定する構造(属性間指定関係)が必要である。

これに対して、物理センサは時変数として扱えるという一般事前知識を用いることができる。つまり時刻 t を変数とする状態 $s(t)$ において比較可能である。さらに時変数としては物理的センサは、時刻を共有することから、同時刻の測定情報に対して属性間指定関係を規定できる。これは物理世界の情報については同時刻に生起する現象をまとめて扱うことが有効であるという一般事前知識に由来している。また物理センサが一定の時間内において安定に存在し動作するという物理的状況に依存している。

さらに視覚/聴覚/味覚など異なるモダリティ情報は、そこで扱う物理信号の種類毎に独立にある段階までの処理を行ううる、これも一般事前知識のひとつといえるだろう。

おそらく身体性という概念は、特定の身体やそれに伴うセンサに特化した形で議論されることが多い。対して本節では、物理センサの普遍的な特性に着目して、物理的 AI における一般事前知識としている。(つまり身体性の共通項的部分である)

§2 局所性と連続性：物理的外界の制約

身体性がセンサに関わるものであったのに対して、外部世界においても物理的な制約による一般事前知識が存在する。

先に述べた物理学的基盤を反映し、外部世界は時間に対して不変もしくは緩やかに変化し、その局所性から近い時刻の状態ほど似通っているという事前知識を利用できる。実はこれは、存在論の最上位における主要概念である物体/物質(モノ)とイベント(コト)に関わる。いずれの概念も、連続したある程度の長期にわたり空間的に局在したパターンが保存されやすいという物理的背景をもつ一般事前知識に関わる。物体概念では同質/異質に限らず複数の構成要素が物理的に一定の関係をもつ。物質概

念では同質の構成要素が空間的に局在するが特定の関係は持たない。イベント概念では、局在した時間において複数の構成要素が何らかの関係を持つが、空間的な局在という要素は比較的弱い。

反対に物理的制約を受けないソフトウェアにおいては、本来は空間的な局所性は不要である。しかし人による認識を促進するために、たとえばオブジェクト指向プログラミングではオブジェクト(モノ)とメソッド(コト)が導入されている。

§3 改めて「単純さ」：領域内共有知識を形作る

我々の生きる物理世界は必ずしも単純であるとは限らない。しかし帰納推論で予測能力を得るには、事物(データ)を少数の要素で説明する「単純さ」を志向する必要がある。そうした意味からは、単純さは事前知識とは必ずしも言えない制約であるが、事実上は一般事前知識として扱うことができる。つまり世界のある部分領域をある側面から見た個別の領域が、単純な知識やモデルで記述可能とする。

この「単純さ」を求める志向性は、オッカムの剃刀とも呼ばれ、モデル選択基準(例えば、AIC, BIC, MDL など)として定式化されている。さらにスパース性や多様体による説明可能性なども、広くは単純性を志向した一般事前知識に含められ、これらは通常、具体的な知識やモデルの獲得に利用されている。

こうした「単純さ」への志向性は、一般性が高く、領域内共有知識の形成にも利用可能な一般事前知識となる。

§4 多角的に領域を生成する

否定されない多角的な観点やモデルを用いて世界を把握するというアイディアは古くはエピクロスが述べており、ベイズ推定でも採用されている。これは認知アーキテクチャの多様なモジュールの存在に結びつく。

多角的な理解のためには、領域内共有知識によって束ねられる個別の領域を持つ必要がある。個々の領域は、前節で述べた、一定上の広がりを持つ空間である。よって典型的な領域は「類似物体の集合」や「類似イベントの集合」であり、そこに類似した局在パターン群が共有される。よって前節で述べた単純さに関わる一般事前知識が利用されている。

AGIにおいては、こうした領域を自動的に生成したい。しかし領域を形成する手段は、AIにおけるフレーム問題にも関わり、現状では決定的な方法は確立されていない。たとえば状況分解[5]といった手法はあるが計算量的の面で困難さがある。ここで生ずる困難さは、領域は高次元の状態空間上で広がりをもつ多様体でありつつ、その多様体が単純さをもたなければならない点に起因する。ここでこの多様体が領域内共有知識である。そして計算量的な難し

さは領域内の多様体と、領域の境界を同時に決定する必要がある点にある。

領域を決定するために別の方法として、データや変数を分割(セグメンテーション)するというアプローチがある。ここでは相互排他性が一般事前知識である。これは連続性と局所性といった物理的基盤に繋がっており、良い結果を得られることも多い。

なお一般に、ある領域はさらに詳細な領域に分解される階層的な局所性をもつが、それ自体も重要な一般事前知識であろう。

§5 類推と不変性

類推や不変性を用いた処理では、等価に扱える変数集合を利用する。等価に扱われた変数集合間で知識が転移することで予測性が向上する。つまり領域内共有知識の適用範囲が拡大するのである。AGIの出発点のひとつが転移学習であったことを考えればこれは自然かとも思える。

類推においては異なる領域に知識を転移するために夫々の領域における知識の類似性が手掛かりになる。この場合には物理的な背景をもつ一般事前知識の存在はあまり期待できない。

不変性は、例えば視覚情報処理の回転不変性のように、一つの領域内で多数の変数集合を等価に扱う。こうしたセンサに近い部分の不変性は物理的な制約を人が理解できるため、しばしば作りこまれる。一方で、やや抽象度の高いレベルでの不変性の作りこみは難しくなる。この自動発見のために一般的な身体性としての一般事前知識を利用しうらう。

5 おわりに

本稿ではAGIの構築に必要な技術は、一般事前知識とデータを利用した領域内共有知識学習であると考えた。また実現可能な究極のAGIは物理世界の制約を活用した物理的AIであると想定した。そこでまず情報理論だけでなく物理的基盤をも背景とした一般事前知識についての整理を試みた。

参考文献

- [1] Goertzel B.: 汎用人工知能概観, 人工知能学会誌 Vol.29, No. 3, pp. 228-233 (2014)
- [2] Hutter M.: Universal Artificial Intelligence, Springer (2005)
- [3] Tenenbaum, J. B., et. al. :, "How to Grow a Mind", Science, Vol. 331, pp. 1279-1285. (2011)
- [4] 山川 宏, 局所多次元時系列の関係表現としての性質の実験的検討, JSAI2013, 3H4-OS-05c-2in (2013)
- [5] 山川 宏他, ETMIC 基準を用いた状況分解によるカード分類課題での概念獲得と予測過程, 認知科学, Vol. 11, No. 2, pp. 143-154 (2004)