

データ循環としての知能と社会

Intelligence as Data Circulation



東京大学大学院情報理工学系研究科ソーシャル ICT 研究センター教授

橋田 浩一

1981年東京大学理学部情報科学科卒業。1986年同大学院理学系研究科博士課程修了。理学博士。1986年電子技術総合研究所入所。1988年から1992年まで(財)新世代コンピュータ技術開発機構に就任。2001年から2013年まで産業技術総合研究所。2013年から東京大学大学院情報理工学系研究科ソーシャル ICT 研究センター教授。専門は自然言語処理、人工知能、認知科学。現在の主な研究テーマはパーソナルデータの分散管理と意味的構造化およびそれに基づく人工知能。

✉ hasida.koiti@i.u-tokyo.ac.jp

1 | はじめに

価値は直接的または間接的に個体や種のサバイバルに帰着され、意味は価値に由来する。また、各々の価値を具現するのは仮説検証サイクル（データ循環）である。ゆえに、認知主体（生物や人工物や社会）がさまざまな意味（価値）を具現するとすれば、それらはさまざまなサイクルの組み合わせだろう。

各サイクルは何らかの制約（constraint；情報処理の手順を捨象した仕様）の充足を指向すると考えられる。つまり意味とは制約であり、認知主体全体の挙動は何らかの制約充足と見なせる。サイクルの機能は最終的には環境への適応（による個人や種のサバイバル）であるから、データ循環（制約 = 意味 = 価値）はしばしば認知主体の内外にわたり、環境への適応を実現する。

代表的なデータ循環（意味・価値の具現化）には下記のような種類がある。

- (1) センシング（知覚）と行動のカップリング
- (2) 学習（知識や技能の修正・拡張）
- (3) 進化や市場における淘汰やイノベーション

これらは時空的スケールの昇順に並べてある。つまり、(1)よりも(2)、(2)よりも(3)の方が、サイクルの1回転に要する時間が長く、またサイクルが空間的に大きい。生命や社会などの有機体は、このように多様な時空

的スケールに及ぶ重層的な構造にわたって意味と価値を具現するものと考えられる。以下本稿では、人工知能およびその社会との関わりについて、主に(1)と(2)の観点から述べる。

2 | 人工知能

人工知能の技術がさらに進歩して総合的に人間のレベルに至るには、上記のように意味（制約）の重層的な組み合わせとして自律システムを構成する方法を明らかにする必要がある。しかし現在の人工知能は、必ずしも上記(1)~(3)のようなサイクル（意味）を基本単位として構成されていない。たとえば、(1)の意味（データ循環）は多くのモジュールを含むと考えられるが、現在のニューラルネットワーク（NN）はそのようなモジュール構造を学習できていないと思われる。

一般に、複雑なシステムは必然的にモジュール構造を持つと考えられる（Simon, 1981）。そして、意味（知識や技能）の各モジュールはデータ循環（意味）の束であろう。たとえば図1のように、 m 層からなるNNを考えよう。このNNでは層 i から層 $i+1$ ($0 < i < m-1$)に活性が直接伝播するものとする。各モジュールはシステム（NN）の内外を巡るサイクルの束である。このNNが学習した意味がモジュール構造を持つとすれば、

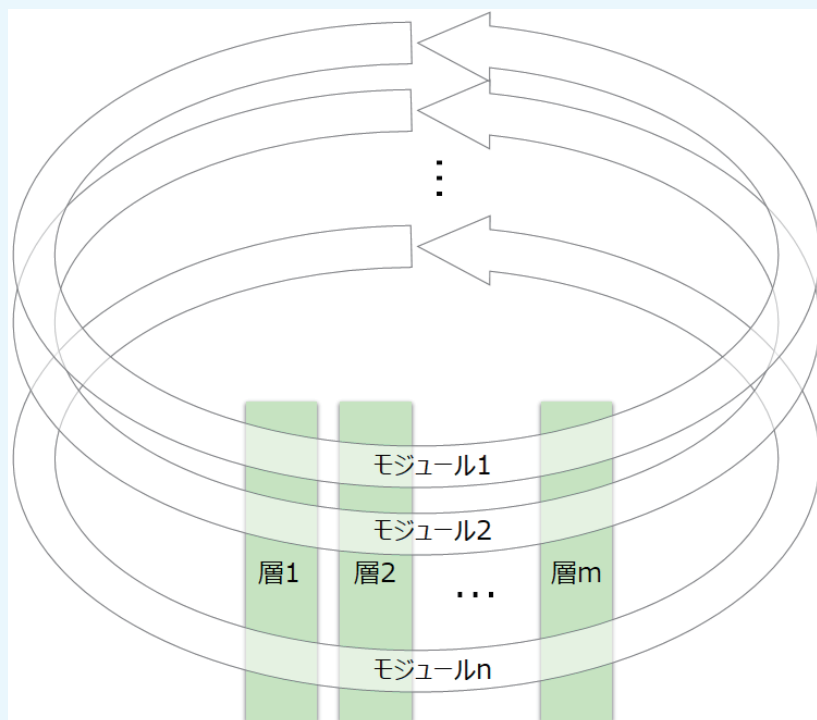


図1 層状ニューラルネットワークと意味のモジュール

各モジュールはすべての層を横断して環境にも及ぶサイクルの束と考えられる。

このようなモジュール構造の存在を想定しなければ、きわめて莫大なデータが学習に必要となり、実際には学習が成立しないと考えられる。つまり、モジュール*i*の学習に必要なデータのサイズを S_i とすると、モジュール1～モジュール*n*をすべて学習するのに必要なデータのサイズは $S_1 \times S_2 \times \dots \times S_n$ だが、それでは一般常識等の学習は現実的にはほぼ不可能であり、現在のNNの技術によって自然言語理解等における長足の進歩がもたらされるとは考えられない。

このようなモジュール構造を機械学習によって自動的に獲得するには、モジュール内の相互作用よりもモジュール間の相互作用の方が弱いことを使えば良いのではないか。それにはたとえば、NNのリンクの重みの絶対値の分散のようなものを表わす項をコスト関数に含めることによりモジュール間のリンクの重みの絶対値をゼロに近付ける等の方法が考えられる。しかしこれはきわめて基本的なアイデアに過ぎず、このアイデアを発展させてモジュラリティを学習する方法を確立するには今後数十年を要するかも知れない。

3 | スマートソサエティ

人工知能にはデータが必要だという議論は、ほとんどの場合、(2)の2次利用(人工知能のトレーニング)に必要な大量のデータ(多数回のサービスの受容者や文脈に関するデータ)に関するものである。しかしそれよりもはるかに重要なのは、(1)の1次利用に必要な個票データ(特定の回のサービスの受容者や文脈に関するデータ)である。

人工知能を実社会でのサービスのために運用する際、そのサービスの個別の受容者や文脈に関する正確で詳細な個票データがサービスの質を効率良く高める(仮説検証サイクルによって精度の高い仮説に効率良く到達するため)に必要なので、そのような個票データがサービスの現場で容易に取得できなければならない。つまり、(1)のためのリッチな個票データがスムーズに流通する社会でなければ人工知能はあまり役に立たない。さらに、(1)用の個票データが容易に取得できるならば、(2)のための大量のデータの収集も容易である。ゆえに、人工知能の振興のために達成すべき最重要課題は、(2)の学習用に大量のデータを集めることではなく、(1)の実サービスのためのリッチな個票データが簡単に利用できるようにすることである。

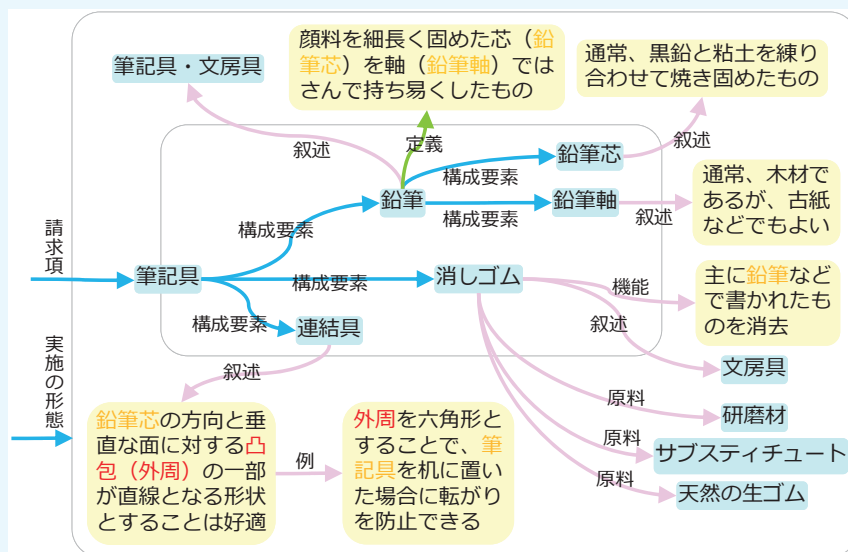


図2 意味構造化文書の図式表現

日本を含むほとんどの国で GDP の大半を個人消費 (小売を含む個人向けサービス) が占める。さらに、個人向けサービスには家事や育児など無料のものも多い。つまり、社会における価値のほとんどは個人向けサービスによって生み出されている。したがって、リッチなパーソナルデータがスムーズに流通し活用される社会を構築することが人工知能の進歩と普及に必須である。とりわけ、(1)のための個票パーソナルデータがサービスの現場で容易にアクセス可能でなければならない。パーソナルデータの利用には原則として本人同意が必要だから、パーソナルデータが本人の意思に基づいて自由に流通することがスマートソサエティ (人工知能と融合することにより価値創造の効率を極端に高めた社会) の構築において本質的である。

さらに、(1)においても(2)においてもデータは標準オントロジーに基づいて意味的に構造化されているべきである。それは、非構造化データの意味を十分に理解できない現在の人工知能に必要であるだけではない。

たとえ人工知能の能力が十分に高まって、そのような構造化は人間にとってこそ必要である。多くの人間は文章読解能力において人工知能に劣る (新井・尾崎, 2017) からである。社会は契約や法律等の文書による意味内容の共有と合意に基づいて運用されているから、多くの人々がそれらの文書を十分に読解できていないということは、その読解を支援して精度と効率を高めることにより社会全体の生産性を向上させられる余地が大き

いということである。

意味的に構造化した文書 (特許の請求項とその実施例) の例を図2に示す。たとえばこのような構造化の方法とそれに基づくユーザインタフェースの開発が求められる。

このような構造化は、人間同士の合意形成の精度を高めるために必要だが、それだけではない。人間の希望を人工知能に伝えたり人間と人工知能が目的を共有して協調したりするには、両者が広義の文書を介して意味内容を共有し合意を形成せねばならないが、人間の読解能力が低い場合にはこうした構造化が必須である。そして、そのような場合は無視できないほど多いと考えられる。

4 | おわりに

データ循環の観点から人工知能の技術的側面と社会的側面を論じた。前記(3)の類のデータ循環については十分に論じられなかったが、他日を期したい。

参考文献

Herbert Simon (1981) The Sciences of the Artificial (2nd Edition) , MIT Press. (稲葉 元吉・

吉原 英樹 訳 (1999). システムの科学. パーソナルメディア)

新井 紀子・尾崎 幸謙 (2017) デジタイゼーション時代に求められる人材育成. NIRA Opinion Paper, 31. <http://www.nira.or.jp/pdf/opinion31.pdf>