2E4-OS-12a-5

群知能型意識ネットワーク構築に向けて

Constructing Consciousness Space by Deep Collective Intelligence

栗原 聡*1 Satoshi Kurihara　坪井一晃*1 Kazuki Tsuboi　藤田直哉*1 Naoya Fujita　芦原祐太*1 Yuta Ashihara

*1電気通信大学
The University of Electro-Communications

In this paper, we will discuss how to construct consciousness space by using deep collective intelligence technology. Deep Collective Intelligence consists of complex multi-layered architecture and complex network of diverse multimodal information. Behavior selection is also important function for the consciousness space for interaction with environment, and we will propose a multiagent based approach.

1. はじめに

AlphaGO が見せつけた AI の高い可能性により，引き続き一般社会を巻き込んでの盛り上がりが加速している AI であるが，AAAI2016 での基調講演においても，Demis Hassabis は「Deep Mind 社はこれから GAI (General Artificial Intelligence) 開発を本格化させる」と宣言している．これまでの，徹底的な先読みに基づく対チェスや対将棋の AI と異なり，我々人のように画像から特徴を抽出し，学習により最善の手を考える AlphaGO*1が，これからの AI の発展に与える影響は極めて大きい．

これから先の AI の発展については，主に次の 3 つのルートに分けらると考えている．1 つ目は，人とのインタラクションが必須な AI である.「場の空気を読む」「人との阿吽の呼吸」といった関係を構築できることが課題であり，人が AI に対して「感情」や「意識」を感じられる AI の構築が目的となる．人と同じ仕組みではなく，人が AI に対して一方的に感情や意識を感じることで十分なのかもしれないが，少なくとも，AI には何らかの意図 (行動を誘発させるエンジン) を埋め込む必要がある．重要なのが目的指向をどのように組み込むかである．

2 つ目は汎用性を持つ AI である．汎用性とは文字通り使い勝手のよい AI ということであり，具体的には学習の追加や転移，そしてプランニングであればリアクティブ性・熟考と即応の両立など,「高い適応性・柔軟性」や「頑健性」といった性質を持つ AI の構築が目標となる．なぜ汎用性を持つ AI が必要となるのか？　力任せでよいのなら Narrow AI の寄せ集めでも実現可能という考え方もあるかもしれない．しかし，寄せ集めて適宜使い分けるメカニズムが肥大することになる．複数の narrow AI をまとめたモジュール化や階層化といった展開も必要となる．

人の知能が汎用性を持つ理由は自明である．身体という限定されたリソースで全てに対応しなければならないからである．当然，ロボットを動かすなら，Wifi 等を使い，頭脳部分はロボット本体には搭載されていないサーバー等を利用することが可能であろう．しかし，我々生物にはそれができない．加えて一つ一つの神経細胞の動作速度はコンピュータに比べてはるかに遅く精度も低い．この状況で地球という自然環境で生き抜くために獲得したのが，脳という汎用性の高いデバイスであり，超多数かつ並列に動作する神経細胞におる大規模複雑ネットワーク化により汎用性を獲得したと言えよう．つまりは汎用性の特徴は「限られたリソースを最大限に活用する効率性や省エネ」ということになる．特に前者が重要な能力であり，この能力の工学的な有用性は高い．

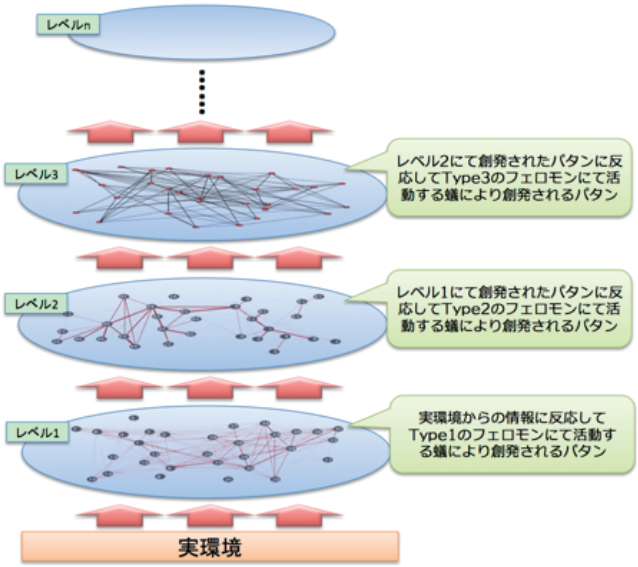

図 1: 多段創発構造

最後がスーパー AI と呼ぶべきルートで，超ビッグデータを対象としたデータマイニング能力であり，膨大な論文から新たな関係や科学的発見の手かがりを探索するなど，ハイパフォーマンスコンピューティングも活用する必要があろう．スーパー AI は現在において膨大なデータと計算基盤を有する Google や Facebook が有利な立場にあることは否定できない．そして，前者 2 つの AI についても，今回の AlphaGO の登場等により水を開けられつつあるが，まだ意識や汎用性などに迫る競争において決定的な優劣がついてはいないと考えている．本論では，前者 2 つの AI の実現に向けたいくつかの可能性について議論する．

連絡先: 栗原 聡，電気通信大学，〒 182-8585 東京都調布市調布ヶ丘 1 － 5 － 1，042-443-5660，skurihara@uec.ac.jp

*1　技術的な本当の成果は DQN であろう．

2. 注目すべき要素

無論，唯一の手本は「人」であり，その中心的モジュールが脳である．コンピュータに比べてはるかい遅く精度の低い脳というデバイスが現在においてもコンピュータと同等の性能を維持できる要因は，

1. 並列に動作する超多数自律分散システムであり，
2. 超多数ノードが大規模複雑ネットワークを構成しており，
3. そのネットワークが階層性とスモールワールド性を有していること，

であると考えている．なお，項目 3. についてであるが，脳はスモールワールド性のみを有しているが，AI を実現するハードウェア環境はスケールフリー性も有していることから，脳にはない新たな能力の実現も期待できるかもしれない．

2.1 基盤は複雑ネットワーク

近年，自然界に存在するネットワークから人工的に組織されたネットワークに至る様々なネットワークがスモールワールド性やスケールフリー性 (総称して複雑ネットワークと呼ぶ) を有していることが分かってきた．これは，特徴的なネットワーク構造がその状況において好ましいということを意味している．興味深いのは，スケールフリーなどのネットワーク構造に着目しているのが，コンピュータネットワークや社会学などに携わる研究者だけでなく，物理，化学，生物 (脳科学)，薬学など様々な分野の研究者であることである．ただし，脳神経ネットワークはスモールワールド性を有するものの，空間グラフであることから，関係グラフが有するスケールフリー性は有していない．スモールワールド性の特徴は，高いクラスタ性を有しつつ，ネットワークの直径が短いということであり，この特徴が脳の機能の基盤の一つとなっていると考えている．脳がスケールフリー性を有しないという事実は，『スケールフリー性を有するコンピュータネットワークにて脳のダイナミクスを実現できた場合，脳が獲得できなかったスケールフリー性という特徴が脳を超える機能を創発させる可能性』を示唆している．

2.2 DNN での中間層での汎化機能

DNN が興味深いのは，その特徴抽出能力の高さはもとより，中間層でのダイナミクスにある，入力段階でのミクロレベルの情報から中間層にて抽象度の高い情報が形成される過程は，身体や社会システムにおける多段創発現象を彷彿させるが (図 1)，DNN のような規則的なネットワークではなく，『スモールワールド性と動的性を兼ね備えたネットワーク構造にて多段創発構造を構築できるかどうか』が課題となる．

2.3 潜在・顕在意識空間と Thought Vectors

人同士であればごく当たり前の「場の空気を読む」「阿吽の呼吸」といった行為を，人と AI との間で構築するにはどうすればよいか？　Siri[*2]を始めとして様々な対話システム (Q & A システム) が登場しているが，それらでは阿吽の呼吸的関係の構築は不可能である．このような関係は，自律的な非線形系同士がアトラクタが形成できる状態と考えると分かりやすい．トラクタには引き込み能力とノイズに対する頑健性があるが (図 2)，息の合った会話においても同様の現象を見ることができる．つまり人と対話するシステム側にも自律性，言い換えれば目的が必要となる (図 3)．

*2　http://www.apple.com/jp/ios/siri/

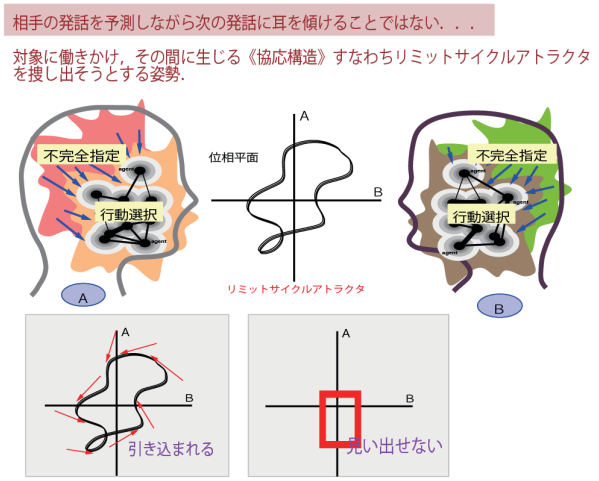

図 2: アトラクタの形成

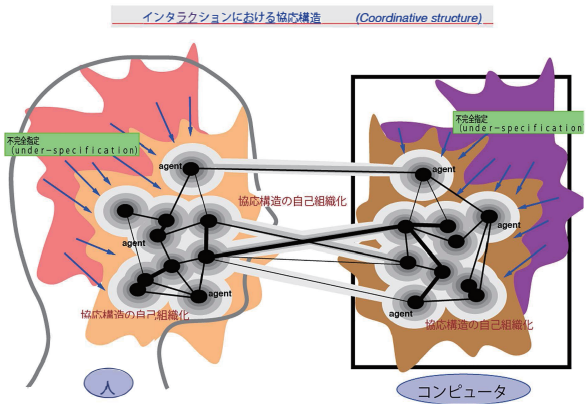

図 3: 人と AI とのインタラクション

そして，AI 側が対話相手の背景やその場の状況に基づいた対話をいかにして行うかが課題である，AI に与える目的は「対話相手を楽しませたい」「Q & A のように対話相手の疑問に答えたい」「対話相手の利得向上」などが考えられよう．優先純度と，達成させるまでに許容される時間が異なる様々なゴールが考えられる．

その課題において，「seq2seq[3]」といった手法も対話対の学習に過ぎないわけであるが，「Thought Vectors[2]」という考え方は同じ方向だと推察されるが，顕在・潜在意識空間という考え方まで考慮されているかどうかは不明である．

2.4 マルチモーダル性

DNN は高い性能を発揮するものの，相当数の学習データを必要とする．しかし，例えば，我々が生涯に見る猫は多くてもせいぜい数百匹程度であろう．人は 2 次元的な猫の画像のみで学習しているのではなく，時系列的な猫の動作に加え，鳴き声や猫を見た時の情景など，膨大な情報と関連させて学習している．AI においてもマルチモーダルな時系列情報を，各モーダル同士の関連性も含めた学習が出来ることが必須である．これにより学習量の低減化も期待できる．

2.5 エージェントネットワークアーキテクチャ

では，上記課題を解決するアーキテクチャをどのようにデザインするかであるが，1991 年の Patte Maes のモデルは今に

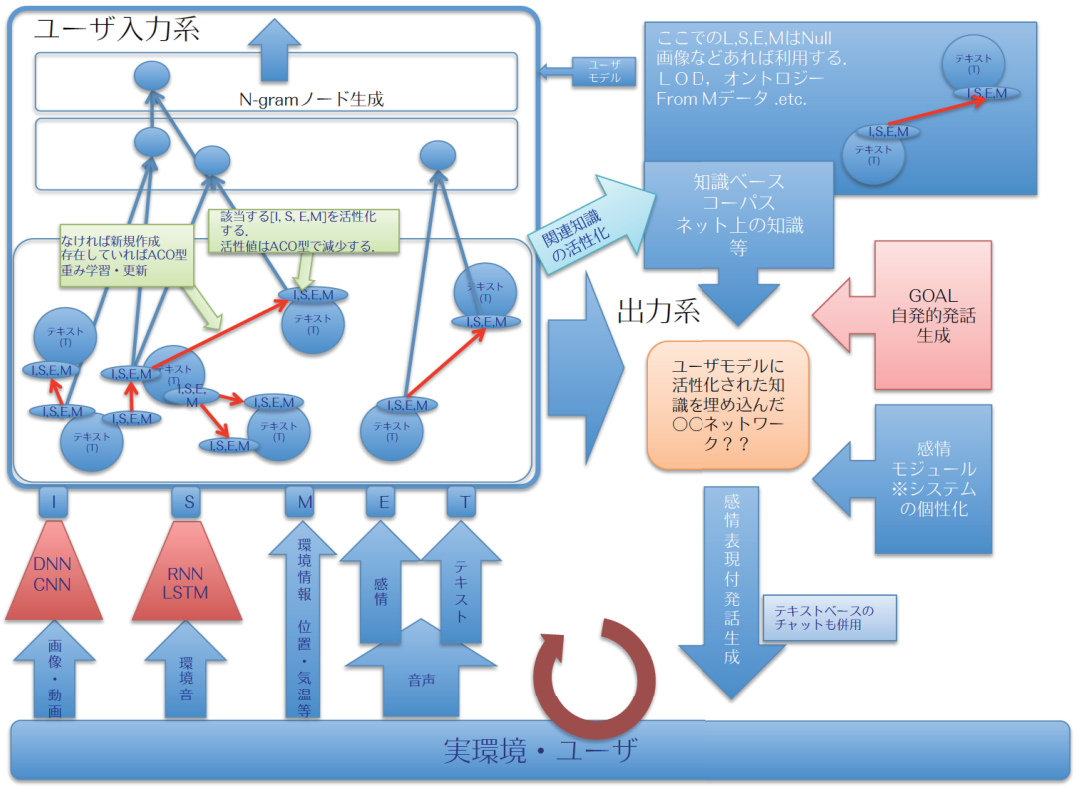

図 4: 認知アーキテクチャ案

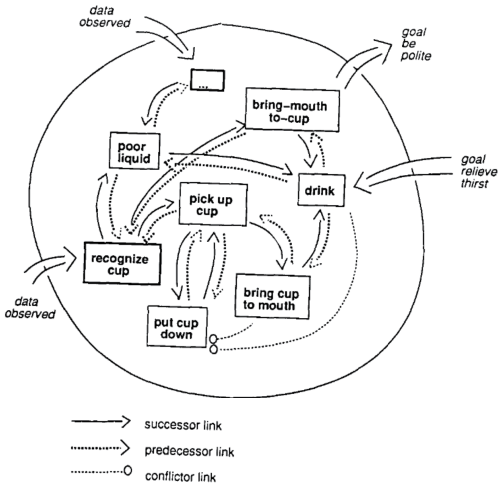

図 5: Agent Network Architecture(ANA)

至っても新鮮である [1]．古典的プランニング法である STRIPS の各単位プランをネットワーク化したものであり，環境側とゴール側から活性伝搬を継続的に送り込む中で活性値が閾値を超えた単位プランが実行される仕組みである (図 5)．脳はこのネットワークが大規模複雑化させたものであり，加えて記憶や優先度や達成までに要する時間が異なる超多数ゴールからの様々な活性伝搬が並列に発生すると考えられる．そして，このモデルは『意識はモニタである』という考え方 [4] とも親和性が高い．

3. 意識空間の構築

図 4 に示すアーキテクチャの構築を開始したところである．入力はマルチモーダル型とし，画像等は CNN 型を介して中間層にて概念化された情報を入力とする．環境音なども含め CNN+LNTM 型 [5] での取り込みが適切であろう．人とのインタラクションにおいて対話が最も重要な入力であるが，音声信号から直接的に意図に変換する方法ではなく，音声から発話テキストに変換しての取り込みとする予定である[*3]．その際，音声信号から感情情報を抽出しこれも感情情報として入力する．図 4 左上段は入力されたマルチモーダル情報を時系列順にネットワークを構築させる．各ノードは入力された発話テキストであり，個々のテキストに，そのテキストが発話された状況での画像や環境情報，感情などが付加される．ノード同士は時系列発生順にネットワーク化され，その強度は発生頻度に基づいて強化される．Ant Coloney Optimization(ACO)[*4] 型のネットワーク構築等を検討する．さらに，この複数のノードが規則的に発火する状況を，メタレベルノードとして上位階層化させ，これを多段階層とすることで，上位階層にて抽象度の高いノードを創発させる．無論，これだけはユーザが発話した情報以外の情報が獲得できないことから，図 4 右上段にて，背景知識や辞書的情報を用意する．こちらも発話ノードネットワークと同様のネットワークとして用意しておき，発話にて生性されたノードと関連するノードとネットワークを構築することで，発話に際して関連するネットワークが発火できるようにする．これらネットワーク全体が潜在意識空間であり，以上が入力系となる．

*3 究極は音声信号そのものを直接入力されることであろう
*4 http://iridia.ulb.ac.be/ mdorigo/ACO/ACO.html

上述したように，人同士で形成されるインタラクションを実現させるにはシステム側もゴール指向型システムとする必要がある．入力系により，関連するノードが発火する様は，意識モデルにおける様々な関連ネットワークが独立かつ並列に発火している状態である．このネットワークに対して，ANA の要領にて，環境とゴールか活性伝搬を加えることで，ゴールを達成するノードを発火させ，具体的な出力とする．具体化された部分が顕在意識という位置づけとなる．複数ゴール同士の優先順位付けは AI のシステムとしての「個性・性格」を位置付けることとなり，人の反応からのフィードバックに基づいてその順位を学習させることで，より人とのインタラクションを自然な形に適応させる．

4. 行動選択と対話

2.3 節でも述べたように，音声をインタフェースとするシステムも増加しつつあるものの，必ずしも的確に答えてくれるわけでもないし，能動的に話しかけてくれるわけでもない．人同士の会話では，お互いが単にお互いのしゃべった文字列に対して反応し合っているのではなく，会話でのテーマをお互いが共有し，お互いがお互いの表情や発言の抑揚などを含む反応を見ながら，そして，会話のテーマにおける相手の背景を考慮しつつ，目的に基づいて発話している．ある時は相手の健康や感情の状態の安定化や，より好ましい状態への遷移が目的であり，ある時はその場の雰囲気の維持が目的かもしれない．

我々も初対面の人が相手では，最初は場の空気を読む会話は難しい，しかし，我々は相手との会話を通して相手の個性や背景知識を徐々に学習し，時には他の人から学習した知識と融合させ，相手との親近感のあるインタラクションを当たり前のようにすることができる．例えば，相手が「のどが渇いた」としゃべっても，その相手の健康への気遣いと状況によっては「自販機を探そう」と言う場合もあれば「今は我慢して」という言う場合もあろう．このような会話は過去の膨大な会話対のデータを学習しても実現できない．

4.1 プランニング

ロボットにおける行動プランニングであれば，初期状態を目的状態に遷移させるための振る舞いを規定する個々の単位行動の最適な適応系列の探索が行われる．個々の単位行動は，STRIPS のようにその単位行動が発動できるための状態や，発動したことで新たに追加される状態等で定義されるのが一般的である．この単位行動の定義付けにより，図 5 のような単位行動間の依存関係を表現するネットワークを構築することができる．

一方，対話における単位行動は動きを表現する「動詞」である．つまり，構築される意識空間の状態に対して発言可能な動詞を同定できればよい．ただし，STRIPS のように個々の動詞に対して，必要となる状態，例えば,「立つ」という動詞において前提条件として「座っている」といったレベルでの定義付けでは不十分である．言葉な多様な状況において使用される．つまりは動詞が使用された時の意識ネットワーク全体の状態そのものを前提条件とする必要がある．ただし，意識ネットワークは大規模複雑ネットワークであることから，そのままではサイズが大きすぎる．次元圧縮が必要となり，CNN のような Deep Learning を利用する方法が有効であると考えられる．対話は時系列データであり，RNN 等の方法を適用しがちであるが，現状において Deep Learning として効果を発揮しているのは CNN である．CNN の特徴抽出能力をいかに利用するかが鍵である．この能力を効果的時系列データに適用した研究例として，時系列を分類する方法 [6] や,「座る」や「歩く」といった振る舞いを認識する 3DCNN[7] などがある．本研究であれは，ネットワークトポロジーをいかにしイメージ化して CNN への入力とするか，ということになる．

また，AlphaGO における DQN のような強化学習を併用することも有用であろう．ただし，AlphaGO と異なり，対話などにより人とインタラクションを行うシステムにおいては，報酬の獲得が困難であるなどの課題がある．プランニングにて粗い行動選択方策を獲得し，継続的に強化学習を行い方策の精緻化と環境適応を行うという組合せが妥当であろう．

5. まとめ

以上，人とインタラクションを行う領域での AI 構築についての概観を述べた．脳は現状にコンピュータに比べて，大規模層的複雑ネットワークの個々のノード全体が並列処理を行うアーキテクチャという意味で全く異なっており，このアーキテクチャが創発する能力が，個々の神経細胞の処理速度の遅さや不確実性を補っている．脳を手本とする取組において，この並列性をどのようにノイマン型アーキテクチャにて実現させるかも大きな課題となる．また，図 4 において，感情を生性させる仕掛け，学習と学習を転移させる仕掛けなどは盛り込まれてはおらず，アーキテクチャとして不完全であるが，本研究のような場合は構成論的手法に基づくアプローチで進めるのが適切であり，まずは実装から開始している状況である．

参考文献

[1] Pattie Maes, The agent network architecture (ANA), ACM SIGART Bulletin Homepage archive Volume 2 Issue 4, Aug. 1991 Pages 115-120

[2] Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Antonio Torralba, Raquel Urtasun, Sanja Fidler, Skip-Thought Vectors, CoRR, abs/1506.06726, 2015.

[3] Ilya Sutskever, Oriol Vinyals, Quoc V. Le, Sequence to Sequence Learning with Neural Networks, CoRR, abs/1409.3215, 2014.

[4] 前野隆司，ロボットの心の作り方 受動意識仮説に基づく基本概念の提案 ，日本ロボット学会誌 23 巻 1 号，pp. 51-62, 2005.

[5] Jeff Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko and, and Trevor Darrell, Long-term Recurrent Convolutional Networks for Visual Recognition and Description, CoRR, abs/1411.4389, 2014.

[6] http://pr.fujitsu.com/jp/news/2016/02/16.html

[7] http://www.ntt.com/release/monthNEWS/detail/20151007_4.html