基般

2. エッセイ集

4 深層学習から汎用人工知能への進化に向けて

栗原　聡（電気通信大学大学院情報理工学研究科／人工知能先端研究センター）

深層学習と人工知能

昨今の人工知能ブームの火付け役となった深層学習法が，ここまで高く注目される理由は何なのであろうか？　これまでにも HMM（Hidden Markov Model）や SVM（Support Vector Machine）など，さまざまなブレークスルーがあった．しかし，いずれも「人の知能を超えるかもしれない」といった話題が出ることはなかった．これに対し，2016 年 2 月の AlphaGO による世界チャンピオンからの勝利や，Rembrandt と同じ画風にて油絵の作成に成功した話題など，深層学習法によるブレークスルーは，人工知能の能力を，囲碁や絵画という限定されたタスクではあるが，人を超えるレベルに押し上げるきわめてインパクトの強い衝撃的なものであった．「人工知能が進化しその能力を劇的に加速させることで，本当に人を超えてしまう可能性があるのではないか?」というある種の悲壮感を秘めた人間側からの推測から，深層学習法によるブレークスルーは世界中で話題となったのである．

しかし，深層学習のみで，いわゆるシンギュラリティが実現されるわけではない．そういった流れの中，Deep Mind 社の Demis Hasabiss 氏はこの春の AAAI-16 で実施した基調講演において，これからの人工知能研究では汎用人工知能（Artificial General Intelligence：AGI）が肝となると発言している．

今後の人工知能の進化の方向性は 2 つある．1 つは，膨大な論文や文献を理解し，新たな科学的発展に寄与するためのスーパー人工知能の実現を目指す方向性である．またそれに加え，日常生活に浸透し人に寄り添う人工知能を実現する，というグランドチャレンジの要素を持った方向性がある．そしてそのような人工知能を実現させるには，汎用性の高い人工知能が必須となる．

なぜ汎用人工知能なのか？

一方で，急速に発展する人工知能に関しては，人工知能により職業が奪われてしまう，といった懸念をよく耳にする．当初は，知的作業が要求されるデスクワークなどに比べて，接客業などから先に人工知能に置き換えられる，という見方が多かったものの，最近はその逆であろうという意見も増えつつある．なぜなら，従来の人工知能にとっては，囲碁や将棋といった，論理的思考が要求される，一般に人にとって苦手なタスクの方が得意であり，人，特に子供でも当たり前のようにできる会話や振舞いの方が苦手だからである．

人が当たり前のようにでき，人工知能が苦手とする顕著なタスクの代表が，「効率的な学習」と「対話」であろう．ここでの学習とは，あるタスクのために獲得した学習結果を別のタスクに応用し，無駄な新規の学習を省く能力や，これまで獲得してきた複数の学習結果を組み合わせて新しい局面に対応する能力のことである．そして，対話においては，現在においても，すでにスマートフォンに搭載されている Siri などの音声対話システムを始め，さまざまな用途で音声を用いたナビゲーションシステムが導入されてはいる．Pepper や，最近発表された RoBoHoN なども音声による対話ができることが特徴である．しかし，現状は，人同士のような違和感のない自然な対話ではなく，文字でのやりとりを音声に置き換えたレベルであったり，質問に対する回答という，定型的なタスクがほとんどである．そして，すでに Siri が利用するデータ量は人の記憶量を大きく凌駕していると推測されることから，データ量を大きくすれば自然な対話が実現されるわけではないことは自明である．

人にあって深層学習系人工知能に不足している重要な機能としては，「マルチモーダル性」と「目的指向」の 2 つが挙げられる．今回の第 3 次人工知能ブーム

は過去 2 回と異なり，研究主導型ではなく実用技術主導型であると言える．深層学習の基本的な手法は 10 年ほど前までに揃っており，技術主導型であればブームは 10 年前に発生していたはずである．深層学習法はその弱点として，学習するために大量のデータが必要であることが指摘されている．10 年前に比べ，現在はビッグデータの利用が容易となり，そして GPU などの計算機環境も劇的に進化している．深層学習の性能が発揮できる状況が整い，来るべくして来ているのが今回のブームであろう．

学習に大量のデータを必要とする点について，我々はどうであろうか？　幼少期に猫を見分けられるようになるまでに 100 万匹もの猫を見てはいまい．おそらくは数匹であろう．だからといって人は少ないデータで学習できると言い切れるであろうか？　人は 1 匹の猫であっても 3 次元動画としてその動き方を捉え，同時に鳴き声や触った場合にはその柔らかさ，そして猫を見たときの情景，場所や気温，風の強さなど，五感を通して入るすべての情報を関連させて猫の概念を獲得している．よって「猫」と言われたときに各自が想起する猫は猫の真正面の顔ではなく，それぞれが経験した情景として想起される．しかし，深層学習においては，猫の顔の 2 次元画像しか与えられていない．よって，大量のデータが必要になると思えば納得もできる．では，仮に深層学習において 1 種類のデータを 100 個で学習させることができるとして，人は 10 種類のそれぞれ異なるセンシング能力でこれを学習するとしたら，個々のセンシングで何個のデータがあればよいであろうか？　単純に考えればそれぞれ 10 個と言いたくなるが，実際は 7 個かもしれない．残りの 30 個分の学習が足りなくなるが，そこに仕掛けがある．我々人間は複数のセンシングチャネルからのデータをそれぞれ独立に記憶しているのはなく，お互いに複雑にネットワーク化して関連付けている．そしてこのネットワーク化による効果が残り，それらが 30 個分のデータに相当するのである．実際に我々の脳は，マルチモーダルな情報収集により，より少ないデータ数で高い学習効果を発揮し，学習においても，ほかの用途に転移させて利用したり，学習結果同士を組み合わせて新たな局面に対応したりする．脳は省エネ型のシステムなのである．

そして，人工知能システムに目的を与えることも必須である．スマートフォンの音声対話システムに対して「喉が渇いた」と発言すれば，直近のコンビニや自販機の場所が回答として返ってくるであろう．しかし人同士の場合，「今は我慢して」などと返答する場合もあるであろう．たとえばこれは，直近の自販機には水以外のジュースしかなく，相手の糖分取り過ぎを防ぐための健康を気遣った回答である．別の言い方をすれば，「相手の幸福度を向上させる」という目的を達成するために「今は我慢して」という発言をしたと言えよう．また，相手への気遣い以外にも，その場の雰囲気を維持したいという目的や，自らの目的を達成するための発言もあり，個々の目的の優先度や達成させるまでの時間的猶予を考慮して適切なプラン立案と実行を行う並列リアクティブプランニングの枠組みが必要となる．プランニングであれば，古典的な STRIPS 型で実現させることも可能かもしれない．しかし，前提条件が，従来であれば状態を表す論理式であるのに対して，今回は，マルチモーダルデータにより構成される大規模複雑ネットワークに対する外部入力により，ネットワークの一部分が活性化したときの，活性化にかかわった部分ネットワークが前提条件となる．無論，その同定の実現は容易ではないものの，取り組むべき重要課題である．

以上本稿では，2 ページという制約の中，詳細までを議論することはできないものの，人の相棒として日常生活にて動作する人工知能に必須な，そして深層学習を汎用人工知能に進化させるために必要な 2 つの要素であるマルチモーダル性と目的指向について概説した．

（2016 年 7 月 2 日受付）

栗原　聡（正会員）　kuri@is.uec.ac.jp

慶應義塾大学大学院理工学研究科卒業．NTT 基礎研究所，大阪大学大学院情報科学研究科／産業科学研究所を経て，2012 年より電気通信大学大学院情報システム学研究科教授．同大人工知能先端研究センターセンター長．大阪大学産業科学研究所招聘教授．ドワンゴ人工知能研究所客員研究員．内閣府科学技術・学術政策研究所客員研究官．博士（工学）．人工知能，複雑ネットワーク科学等の研究に従事．著書『社会基盤としての情報通信』（共立出版）．翻訳『群知能とデータマイニング』，『スモールワールド』（東京電機大学出版）等．人工知能学会理事・編集長などを歴任．