

AIの未解決問題とDeep Learning

松尾 豊^{*1}

Yutaka Matsuo

^{*1}東京大学

University of Tokyo

Deep Learning is a recently-developed set of machine learning techniques which attracts much attention in academia as well as industries in these years. The most prominent part of the techniques is its ability to obtain abstract representations by unsupervised learning, and typically the representation composes a layered-structure. Although deep learning is studied in machine learning communities, its relation to classical AI is not discussed much. In this paper, we try to discuss how the deep learning technology relates to the classical AI problems such as frame problem and symbol-grounding problem. The difficult parts of such AI problems can be reduced into single critical problem: How can we build a good representation from data? Given a large amount of data, we can gradually generate “symbols” which convey information by a small fragment. Such attempts had been impossible in the past because of the lack of data and computational power. By building symbols one by one, layer by layer, the classical AI problems disappear, and there could be many possibilities for AI studies to be practically useful. This paper shows the direction toward bridging the gap between classical AI and recent deep learning development.

1. はじめに

近年、Deep Learning が注目を集めている。Deep Learning とは、深い層を重ねることによって学習精度を上げるように工夫したニューラルネットワークを用いる機械学習技術のことである。2006 年に Hinton らが教師無し学習を反復的に用いることで深い階層のニューラルネットワークで精度を上げることに成功して以来、さまざまな手法が提案され、2011 年には音声認識のタスクで優勝、2012 年には ILSVRC という一般物体認識のコンテストで圧勝するなど、数多くのコンペティションで成果を収めてきた。2013 年には、Google が DNNresearch という Deep Learning の第一人者であるトロント大の Hinton 教授とその学生らが立ち上げた会社を買収、年末には Facebook が同じく Deep Learning の主要研究者であるニューヨーク大の LeCun 教授を招いて人工知能研究所を設立、2014 年初頭には、DeepMind という Deep Learning 技術の会社を Google が 4 億ドルで買収と、産業界周りの動向も慌ただしくなってきた。

さて、Deep Learning とは何であろうか。なぜこのように注目を集めるのだろうか。その立場には大きく 2 つあるように感じる。単なる流行であるという立場と、大きなブレークスルーであるという立場である。私自身は後者だと思っている。しかし、前者の立場から見ると、こうしたトレンドは歴史的には何回も繰り返されている。したがって、今回が真の突破口であると信ずる客観的な理由はない。

それでも著者が、Deep Learning の研究が重要だと考える理由は、Deep Learning が表現学習という AI で最も難しい問題に対する突破口を与えているからである。これまでに AI では多くの研究が行われてきた。AI の初期には夢のようなプロジェクトがたくさんあったわけだが、それがいずれかの時点で壁に当たる、もしくは実用化までに至らないのはさまざまな理由があったとされる。しかし、そうした課題というのは、ほとんどが唯一の難しい問題、すなわち表現をいかに獲得するかという問題に帰結できると考えている。したがって、ここを突破

することができれば、これまでの AI の研究が新しい形で一気に活用されていくのではないだろうか。

本稿では、Deep Learning と AI の未解決問題、特にフレーム問題やシンボルグラウンディング問題との関わりについて述べる。

2. Deep Learning の概要

まず、簡単に Deep Learning の仕組みについて説明しよう。詳しくは、[1] や、人工知能学会誌の深層学習連載特集 ([5, 4, 8, 6, 7, 2]) に詳しい。

Deep Learning では、通常、Restricted Boltzmann Machine、もしくはオートエンコーダと呼ばれる単層の学習モデルが用いられる。これらは、入力されたデータ自身を出力するような学習（データ自身しか使わないので教師無し学習）を行う。このとき、少ない数の中間ニューロンを経由することで、データの圧縮を行う。データを圧縮し（encode）、それを復元（decode）したときに、最も誤差が小さい中間ニューロンが得られれば、それがある種の抽象表現になる。

さらに、この中間ニューロンの値自身を、再度、入力データとして、それを出力するような教師無し学習を行う。すると、もう 1 段、抽象化された表現を生成することができる。こうして、教師無し学習で多段に階層的な構造を構成することができる。最後に、こうして作った中間ニューロンを素性とし、教師あり学習を行うことで、高い精度が得られるというものである。

各階層の中で行っていることは、要するに次元の圧縮であり、最も身近なものは主成分分析であるが、Deep Learning ではさまざまなバリエーションが用いられる。いかに頑健に入力データを復元できるかというところが焦点であり、そのために、ノイズを加える、スパース性を仮定するなどの工夫がある。最近では、DropOut という手法で、毎回ランダムに半数のニューロンを消した状態で学習を行うことがよく行われる。また、MaxOut は、複数の線形関数の中で最大値を関数の値として採用するものであり、多くの分類結果で良い性能を達成している。

3. 人工知能との関わり

さて、Deep Learning は人工知能とどんな関係があるのだろうか。そうした議論は、[1, 8] にも詳しいが、ここでも少し議論を展開してみたい。

表現とは何であろうか。現象は何らかの形でデータ化することができ、その上でのパターンも記述することができる。一方で、全てを記述できることと、それが経済的であることは両立しない。この表現の可能性と、それがどのように経済的であるかのトレードオフによって、良い表現かどうかが決まってくる。こうした表現のトレードオフについては、[3] でも議論されている。したがって、良い表現かどうかを議論する際に、それがさまざまなタスクに対してどのくらい汎用に有用であるか、それによってどのくらい表現力を制限されてしまうのか、その両面から考える必要がある。そうして見たときに、ひとつの方法として「自分自身のデータを復元するようなコンパクトな表現はよい表現である」と考えるオートエンコーダの考え方は妥当である。

表現は、現実の事象を捉え、意思決定主体が合理的な判断を行っていくために必要なものである。その背景には、生物が変わりゆく環境のなかで、いかに「早く」不偏量を見つけ、それに適応するかの競争をしていることがある。ここでいう早いというのは、絶対的な時間として早いのではなく、与えられたサンプル数に対して早いということである。2、3の事象から、周りの地形を把握できれば、行動の上で有利であるし、2、3の事象から敵がいることを察知することができれば、すばやく逃れることができる。

実際、Deep Learning 系の表現学習は、教師データが豊富にあるときにはその効果が現れない。層を重ねることによる膨大な表現力の空間を探索するには、現実のデータはあまりに少ない。少ないデータで素早い学習を行うには、さまざまな仮定を置かなければならない。例えば、複数のタスクに共通する要因は、真の要因である可能性が高い。簡単な概念の組み合わせで概念を作るほうが、うまくいく確率が高い。こうした「なぜか分からないがこの世界では真であることの多い前提知識」、すなわち prior をいかに使うかが世界の謎を早く解くためのヒントになる。Deep Learning は、こうした不偏量を早く見つけるための仕組みとして、合理的で妥当であると感じる。

人工知能における長年の問題は、表現に関わる問題である。それはしばしばフレーム問題と呼ばれたり、シンボルグラウンディング問題であるとされたりする。フレーム問題は、状況に応じて適切な知識を記述することができないという問題であり、環境において何を表現すべきかに関わる。これを Deep Learning の観点から考えると、さまざまな現象を記したデータがあり、そこにおいて何を表現すべきかという問題をボトムアップに解くのであれば、表現されたものの範囲外にあるような事象というのはそれほど発生しないはずである。(なお、人間の場合には、シンボルを操作の系列として組み合わせることで、例外的な事象を表現する方法であったり、その解決策を産み出しているが、それはまた操作の組み合わせによる解空間の捉え方であり、別の話である。)

一方、シンボルグラウンディング問題は、記号で指し示されるものを計算機がどのように認識するかという問題であり、何を特徴として捉えるべきか(表現されるべきか)に関わる。これは、シンボルを、データから生成されたものとして扱わずに、トップダウンに与えるために発生する問題である。つまり、データからボトムアップに何が表現されるべきなのかを捉え、それに対して名前をつけるという過程を経るのであれば、

シンボルがそれが指すものと切り離されて存在するということは起こりえない。

こうしたシンボルをシンボルとして扱わざるを得なかった背景には、計算能力の限界があるだろう。以前の計算機の能力で、シンボルの操作を実現しようと思えば、それはシンボルが指すものとは別のものとしての系を構築せざるを得ない。(そして、そのことはシンボルのひとつの大きなメリットでもある。)しかし、多くのデータと計算機の資源がある現在、データ由来で表現を生成し、それに名づけるという過程を踏むことで、こういった問題は解消できるはずである。

表現を巡る議論は根深い。ここで議論した以外でも、例えば、そもそも生物がセンサーとして獲得している情報の取捨選択という段階と、それをどのように加工するのかという段階でも少なくとも2段階のステップのバランスを取る必要がある。時間に関する扱い、行動に関する扱いなど、足りない点が多々ある。表現を巡る旅はこれで終わりではなく、これが始まりである。Deep Learning は認識の構造として、抽象化のしくみをうまく実現しているが、人間の知能には、「思考する」という行為を行う。この思考というのはいったいなんだろうか。例えば、AIの研究で長く行われてきた推論やプランニングなどは、Deep Learning の枠組みとどう関わるのだろうか。この答えは単純ではないと思う。

今後、さまざまな形で Deep Learning の研究は進んでいくだろう。Deep Learning とプランニングや推論を組み合わせた研究が次に大きなインパクトをもたらすかもしれない。つまり、主体の行動という観点から見た表現とは何なのだろうか。チューリングマシンのような原始的な行為の系列としてすべてのプログラムが実現できるとする立場と、この表現学習はどのような関わりをもつのであろうか。おそらく、我々が行動とか思考と言っているものの自体、抽象化されている。少し分りにくい表現になるが、encoder、decoder 自体が抽象化されているなかで、どのような抽象表現が得られるのだろうか。さらには、短期記憶やイメージの生成能力は Deep Learning の抽象化とどのような関係にあるのだろうか。こうした部分が解き明かされてくれば、時系列のデータ、行為のデータが対象にはいってくるはずである。名詞概念だけでなく、動詞や副詞、形容詞など、幅広い言語概念とのバインディングが可能になる。その際に、バインディングされた概念からどのように意味理解をすることができるのだろうか。

4. おわりに

私自身、2000年には、教師無しデータを教師あり学習に用いて精度を上げる研究 [Nigam00] に感心し、また、2005年には入力データをそのまま正解データとして用いる手法 [Ando05] を見て、これだと思った。その後、私自身も Deep Learning の趣旨とほぼ同じ内容の提案書をさがし研究で書き進めていたが、このブレークスルーにはいたらなかった。同様の狙いの研究は、国内でも私が知る範囲でも、非常に多くの研究者が試みている。実はこの先の展望をさまざまに描いている研究者も多いのではないかと思います。Deep Learning に限らず、2013年から新しい国際会議 International Conference on Representation Learning (ICLR) が開かれ、100名を超える参加者が参加した。国内では、著者を含む3人のオーガナイザによって、「全脳アーキテクチャ勉強会」が開かれ、第1回は100名、第2回は200名を超える参加者を得た。今後、この分野の発展を多くの研究者とともに盛り上げていきたい。

参考文献

- [1] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 2013.
- [2] ボッレーガラ ダヌシカ. 自然言語処理のための深層学習. 人工知能学会誌, 29(2), 2014.
- [3] 松原 仁. 一般化フレーム問題の提唱. 哲学書房, 1990.
- [4] 岡野原 大輔. 大規模 deep learning (深層学習) の実現技術. 人工知能学会誌, 28(5), 2013.
- [5] 安田 宗樹. ディープボルツマンマシン入門-ボルツマンマシン学習の基礎-. 人工知能学会誌, 28(3), 2013.
- [6] 岡谷 貴之. 画像認識のための深層学習. 人工知能学会誌, 28(6), 2013.
- [7] 久保 陽太郎. 音声認識のための深層学習. 人工知能学会誌, 29(1), 2014.
- [8] 麻生英樹. 多層ニューラルネットワークによる深層学習の学習. 人工知能学会誌, 28(4):649–659, 2013.