

2019/11/10 日本科学哲学会第52回大会

ワークショップ： 深層学習の哲学的意義

オーガナイザ・提題者：鈴木貴之（東京大学）
提題者：植原亮（関西大学）
大塚淳（京都大学）

1

このワークショップは、JST/RISTEX「人と情報のエコシステム」研究開発プロジェクト「人と情報テクノロジーの共生のための人工知能の哲学2.0の構築」の研究活動の一部です。

website: <http://updatingphilosophyofai.net>

RISTEX

HITE
Human
Information
Technology
Ecosystem

2

「ディープラーニングは特徴表現学習の一種であり、その意義の評価については、専門家の間でも大きく2つの意見に分かれている。1つは、機械学習の発明のひとつにすぎず、一時的な流行にとどまる可能性が高いという立場である。これは機械学習の専門家に多い考え方だ。もう1つは、特徴表現を獲得できることは、**本質的な人工知能の限界を突破している可能性がある**とする立場である。こちらは機械学習よりも、もう少し広い範囲を扱う人工知能の専門家に多いとらえ方である。**本書は、後者の立場に立つ。**」（松尾 2015、p. 180）

3

「私は、知能には鳥が飛ぶことと同じように原理があり、それを工学的に利用することもできるはずだと思っています。**すでにディープラーニングで最大の難所が突破されたいま、あとは身体性や記号操作の仕組みを獲得できれば、知能の原理の大方は説明がつくのではないか**—それが私の考えです。」（松尾 2019、pp. 152-153）

4

深層学習の哲学的意義

- 深層学習によって（真の）人工知能は実現するのだろうか？
- 深層学習は人間の認知の本質なのだろうか？
→心の哲学（認知科学の哲学・人工知能の哲学）にとっての意義

5

- 深層ニューラルネットワークはそもそも何をしているのだろうか？
→統計学の哲学にとっての意義

6

- 知的道具としての深層ニューラルネットワークには、どのような可能性があるだろうか？
→認識論にとっての意義

7

おことわり

- 深層学習そのもののくわしい説明はしない（できない）。
- 結論を出すというよりは探索的。
- 取り上げる話題は狭義の深層学習に限定しない。

8

深層学習の哲学的意義： 認知科学の哲学と人工知能の哲学 の場合

鈴木貴之
(東京大学大学院総合文化研究科)

9

問い

- 深層ニューラルネットワーク（以下DNN）を用いた人工知能研究（およびその背景にある理論）は、人間の認知の本質や人工知能の実現可能性をめぐる哲学的論争に、あらたな展開をもたらすか？

10

1 背景

- 2 深層ニューラルネットワーク
- 3 哲学的意義の検討

11

認知科学の哲学と人工知能の哲学

- 認知科学の哲学：人間の認知の本質とは？
- 人工知能の哲学：（汎用）人工知能は実現可能か？

12

認知科学の哲学	人工知能の哲学
記号計算主義	古典的AI (GOFAI)

13

- 記号計算主義：
 - 認知＝形式的な規則に従った記号操作、計算
 - 操作の対象＝意味論的に透明な内的表象
 - 心＝デジタルコンピュータ

14

- 古典的AI (GOFAI)：
 - 知能の本質は計算
 - 知能はデジタルコンピュータによって実現可能

15

- GOFAIへの哲学的批判：
 - 意味理解の問題（中国語の部屋、記号接地問題）
 - 常識、文脈、関連性の問題（フレーム問題、ドレイファスの批判）

16

認知科学の哲学	人工知能の哲学
記号計算主義	古典的AI (GOFAI)
コネクショニズム	ニューラルネットワーク を用いたAI

17

- **コネクショニズム：**

- 認知=ニューラルネットにおける活性化パターンの変換
- 知識は重み付けパターンの全体によって表象される（分散表象）
- 長所：生物学的妥当性、汎化能力、不完全な入力への耐性、ダメージへの耐性

18

- **コネクショニズム：**

- 認知=ニューラルネットにおける活性化パターンの変換
- 知識は重み付けパターンの全体によって表象される（分散表象）
- 長所：生物学的妥当性、汎化能力、不完全な入力への耐性、ダメージへの耐性

19

- **ニューラルネットワークを用いた人工知能研究：**

- 1960年代のパーセプトロン→1980年代のPDPモデル
- パターン認識などに一定の成果を挙げる。
- ただし、第1次と第2次の人工知能ブームはいずれもGOFAIが中心。

20

- 認知科学の哲学におけるその後の展開 (cf. Clark 2014) :
 - 身体・環境の重要性
 - 反表象主義

1 背景

2 深層ニューラルネットワーク

3 哲学的意義の検討

認知科学の哲学	人工知能の哲学
記号計算主義	古典的AI (GOFAI)
コネクショニズム	ニューラルネットワークを用いたAI
	深層ニューラルネットワークを用いたAI

深層学習 (以下DN)

• 位置づけ :

広義の人工知能研究

> 機械学習

> ニューラルネットワークによる機械学習

> 深層ニューラルネットワーク (以下DNN) による機械学習

- 発展の背景：

- コンピュータの性能向上
- ビッグデータの出現
- 新たな手法

25

- 手法：

- **構造**：畳み込みネットワーク、再帰型ネットワーク、ボルツマンマシン、敵対的生成ネットワーク...
- **学習手法**：自己符号化、ミニバッチ学習、ドロップアウト、バッチ正規化...
- **活性化関数**：ソフトマックス、ReLU...
- **誤差関数**：クロスエントロピー...

26

1 背景

2 深層ニューラルネットワーク

3 哲学的意義の検討

27

認知科学の哲学における意義

- コネクショニズムの勝利？

- DLが有効なのはおもに入出力モジュール。
- 体系性をめぐる論争は現在も継続中 (cf. Fodor and Pylyshyn 1988; Pollack 1990; Lake and Baroni 2018)。
- 適切な記述の抽象度とは？ (cf. Buckner 2019)

28

- DLは人間の認知メカニズムか？

- 問題①：生物学的な妥当性（必要なデータの量、ネットワークの構造、重み付けの修正方法...）
- 問題②：adversarial examples

29

- 認知科学の哲学における新しい展開との関係は？

- DLは脳内における情報処理過程。
- 深い層では局在的な表象を見出すことも可能。
→認知科学の哲学のトレンドに逆行？

30

- “predictive mind” (cf. Hohwy 2013; Clark 2016) との関係

- 順伝播型ネットワークは心のよいモデルではない？

31

人工知能の哲学における意義

- DLの限界 (Chollet 2017; Marcus 2018) :
 - 大量のデータが存在しない課題
 - 形式的推論、アルゴリズム的な操作
 - ベクトルの幾何学的変換として扱えない課題

32

- DLによって汎用人工知能は実現可能か？

- DLは記号接地問題の解決をもたらすかもしれない。
- DLは中枢の認知過程にも応用できるか？
→破滅的忘却 (catastrophic forgetting) の問題 (cf. French 1999) と転移学習の問題

33

- 考えられるアプローチ：古典的AIとのハイブリッド (Chollet 2018; Marcus 2018)、複数のDNNの組み合わせ、ロボット工学+DL (+ α)
- 短期的には、領域限定的なDNNを有用な道具として利用するのが現実的。

34

深層学習の哲学的意義とは？

- 新しい心観？

- 80年代からある論争だが...
- コネクショニズム陣営のメジャーバージョンアップ
→より実質的な論争が可能に。
- そもそも何対何の論争なのか？ (推論対パターン認識、論理対統計、演繹対帰納...)

35

- 意義はどこにあるのか？

- 認知の理論としての意義
- 数学的道具としての意義

36

参考文献

Buckner, C. (2019) Deep Learning: A Philosophical Introduction. *Philosophy Compass*, 14: e12625.

Chollet, F. (2017) *Deep Learning with Python*. Shelter Island, NY: Manning Publications.

Clark, A. (2014) *Mindware: An Introduction to the Philosophy of Cognitive Science (Second Edition)*. New York: Oxford University Press.

Clark, A. (2016) *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. New York: Oxford University Press.

Fodor, J., and Pylyshyn, Z. (1988) Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition*, 28: 3-71.

French, R. (1999) Catastrophic Forgetting in Connectionist Networks. *Trends in Cognitive Sciences*, 3(4): 128-135.

Hohwy, J. (2013) *The Predictive Mind*. Oxford: Oxford University Press.

Lake, B., and Baroni, M. (2018) Still not systematic after all these years: On the compositional skills of sequence-to-sequence recurrent networks. <https://openreview.net/forum?id=H18WqugAb¬elid=H18WqugAb>

Marcus, G. (2018) Deep Learning: A Critical Appraisal. arXiv:1801.00631

Pollack, J. (1990) Recursive Distributed Representations. *Artificial Intelligence*. 46(1-2): 77-105.

松尾豊 『人工知能は人間を超えるかーディープラーニングの先にあるもの』 角川書店、2015年

松尾豊 『超AI入門ーディープラーニングはどこまで進化するのか』 NHK 出版、2019年