

2019年11月10日

日本科学哲学会@慶應義塾大学

WS「機械学習・深層学習の哲学的意義」

# 機械学習・深層学習と 知的創造性

植原亮

(関西大学総合情報学部)

【略語】

ML: Machine Learning

DL: Deep Learning

# 1. はじめに

- ML・DLの認識論的問題
- 創造的な？ML・DL

# 2. 創造性の哲学

- Boden, Kronfeldner
- Guatら, 徳認識論、多元主義へ

# 3. ML・DLと知的創造性：課題

# 1. はじめに

- ML・DLの認識論的問題
- 創造的な？ML・DL

# 2. 創造性の哲学

- Boden, Kronfeldner
- Guatら, 徳認識論、多元主義へ

# 3. ML・DLと知的創造性：課題

# ML・DLの認識論的問題

## 1. 知識をもてるか？

- ML・DLの本性
- 知識とは何か(定義)、知識概念の改訂？

## 2. Black Box 化にまつわる問題(とくにDL)

- 人間には理解できないプロセス
- 正当化・信頼性の確保 (cf. Schubbach 2019)
- 科学の自動化の可能性 (cf. 北野2019)
- 認識的依存、利用者に必要な知識・スキル

(Ratti forthcoming)

# ML・DLの認識論的問題

## 3. 人間の認識のモデルとして

- ベイズ的予測機械としての心
- 生得説 vs. 経験主義: e.g. 抽象概念の起源
- 信念の本性
- 古典的懐疑論の新形態? (仮説としての世界)  
知覚や錯覚、経験や夢の捉え直し

cf. Clark 2016, Hohwy 2016

## 4. 知的創造性をもてるか?

- 人間に固有か? そもそも創造性とは?

# 創造的な？ ML・DL

## 1. DeepMind

- AlphaGo (2016) : イ・セドルに勝利
- 独創的な手
  - チャンピオン「美しい。これは人間の手ではない。人間がこの手を打つのは見たことがない。」
  - 序盤は教師あり学習 → その後、自分と対局
- AlphaGoゼロ (2017) : ルールのみからスタート
- Alpahaゼロ : ルールなしにスタート
- AlphaFold : タンパク質の3D構造解析

# 創造的な？ ML・DL

## 2. 物理学・天文学（cf. セイノフスキー 2019）

- 素粒子衝突実験や大型望遠鏡の巨大データ
- MLで処理可能に
- 遠方の銀河系を検出・発見など
- 近似計算の高速化：三体問題等（Breen et al. 2019）
- 10層のDNNで従来の1億倍の速度

## 3. 芸術

- バッハと違いが分からないような楽曲の生成
- レンブラント風絵画など

# 創造的な？ ML・DL

- 稲葉振一郎(2019)『AI時代の労働の哲学』
  - 模倣にすぎないという批判もあるが、人間の創造性もよくわかっていないので焦るべからず
- 松尾豊(2015)『人工知能は人間を超えるか』
  - 現在：特徴量／概念の獲得は創造性そのもの
  - 将来：環境との相互作用と試行錯誤が可能になれば、これまで誰も考えたことのないアイデアも
  - (さらに発展すれば知識も獲得可能という見通し)



# 1. はじめに

- ML・DLの認識論的問題
- 創造的な？ML・DL

# 2. 創造性の哲学

- Boden, Kronfeldner,
- Guatら, 徳認識論、多元主義へ

# 3. ML・DLと知的創造性：課題

# 創造性の哲学：Boden

創造性についての影響力ある規定（2004など）

- ①新しい ②価値ある ③驚くべき

アイデアや人工物を生み出す能力

# 創造性の哲学：Boden

創造性についての影響力ある規定（2004など）

①新しい ②価値ある ③驚くべき

アイデアや人工物を生み出す能力

心理的創造性（個人）／歴史的創造性（人類）

②は文脈による

# 創造性の哲学：Boden

創造性についての影響力ある規定（2004など）

①新しい ②価値ある ③驚くべき

アイデアや人工物を生み出す能力

心理的創造性（個人）／歴史的創造性（人類）

②は文脈による

- 結合的 (combinational)
- 探索的 (exploratory)
- 変形的 (transformational)

# 創造性の哲学: Boden

結合的 (combinational)

- 馴染みのアイデアの組み合わせ

探索的 (exploratory)

- 既存の概念空間 (制約) 内での新しさ
- ベンゼン誘導体など

変形的 (transformational):

- 概念空間 (制約) そのものの変更
- それまでは考えることが不可能に見える (最も驚く)
- 非ユークリッド幾何学、ベンゼンの構造 (ケクレ)

# 創造性の哲学: Boden

結合的 (combinational)

- 馴染みのアイデアの組み合わせ

探索的 (exploratory)

- 既存の概念空間 (制約) 内での新しさ
- ベンゼン誘導体 など

鎖状のみから  
環状構造も可能に

変形的 (transformational):

- 概念空間 (制約) そのものの変更
- それまでは考えることが不可能に見える (最も驚く)
- 非ユークリッド幾何学、ベンゼンの構造 (ケクレ)

# 創造性の哲学: Kronfeldner

ケクレの例の検討から (esp. 2009, cf. 2018)

- 化学の知識と経験は豊か
- にもかかわらずベンゼンの構造は把握できず
- きっかけ: 動物園の蛇、輪になって踊るダンス

# 創造性の哲学: Kronfeldner

ケクレの例の検討から (esp. 2009, cf. 2018)

- 化学の知識と経験は豊か
- にもかかわらずベンゼンの構造は把握できず
- きっかけ: 動物園の蛇、輪になって踊るダンス



- 斬新さ: 知識や経験からの距離
- 驚き: 予測不可能 (本人にも)
  - ⇒ 創造の過程の自発性 (spontaneity) の重要性
    - 意図的な制御やプランから独立
    - 本人にとってはむしろ受動的



# 創造性の哲学：Kronfeldner

Boden:

①新しさ ③驚き  
についての補足・深化

プロセスに注目

muse-likeな  
靈感的側面

- 斬新さ: 知識や経験からの距離
- 驚き: 予測不可能(本人にも)
  - ⇒ 創造の過程の自発性(spontaneity)の重要性
    - 意図的な制御やプランから独立
    - 本人にとってはむしろ受動的

# 創造性の哲学：Gaut, Paul, Kieranら

## 行為者性 (agency) の重視

人に注目

- 真正の理解や知識の所有・行使
- 創造の過程に対する適切な判断・評価 etc.

(e.g. Gaut 2018, Paul and Stokes 2018)

# 創造性の哲学: Gaut, Paul, Kieranら

## 行為者性 (agency) の重視

人に注目

- 真正の理解や知識の所有・行使
- 創造の過程に対する適切な判断・評価 etc.

(e.g. Gaut 2018, Paul and Stokes 2018)



- 主体の自発性 (spontaneity) による: 非偶然的  
⇒ 創造性が称賛に値する理由
- 卓越性 → 徳という方向へ (Kieran 2014)  
(知的徳 Intellectual Virtue)

# 創造性の哲学：徳認識論

焦点：主体に備わる知的徳

人に注目

- 開かれた心
- 知的謙虚さ
- 認識的勇気 etc.

△ 信念そのもの  
△ 形成プロセス

知識の定義

- 知的徳の発揮を通じて形成された真なる信念

知識に「真なる信念」以上の価値がある理由

- 知的徳の発揮を通じてなされる「達成」だから
- 徳の発揮は善き生を構成するから

# 創造性の哲学：徳認識論

## 1. 創造性は知的徳ではない

- 生得的なものなので涵養できない(esp. 靈感)

cf. Baehr 2018

## 2. 創造性は知的徳の一種である

- 科学的発見などさまざまな認識活動に関わる
- 教育の重要な目標として設定されうる
- 創造性の発揮に価値がある

## 3. 古典的な意味では存在しない

- 通状況的な頑健性がない (Alfano 2013)

## 4. 徳を涵養すべきかどうかそもそも疑問 (Prinz 2009)

# 創造性の哲学：徳認識論

1. 創造性は知的徳ではない

- 生得的なものなので涵養できない

cf. Baehr 2018

2. 創造性は知的徳の一種である

- 科学的発見などさまざまな認識活動に関わる
- 教育の重要な目標として設定されうる
- 創造性の発揮に価値がある

3. 古典的な意味では存在しない

- 通状況的な頑健性がない (Alfano 2013)

4. 徳を涵養すべきかどうかそもそも疑問 (Prinz 2009)

# 創造性の哲学：多元主義へ

## 2つの方向

- ① プロセス重視：新しい、驚くべき
- ② 主体重視：称賛に値する卓越性（徳）、涵養の対象

## 考慮すべき点

- 霊感的なケース（esp.②）
- （程度差あり）

## 存在論的身分の候補

- ？ 単一の自然種、複数の自然種の集合
- 多元主義：さまざまな特徴のどれかを満たせばOK

# 1. はじめに

- ML・DLの認識論的問題
- 創造的な？ML・DL

# 2. 創造性の哲学

- Boden, Kronfeldner
- Guatら, 徳認識論、多元主義へ

# 3. ML・DLと知的創造性：課題



# ML・DLと創造性：プロセス重視

Bodenの分類が適用可能

- 既存の特徴量 ⇒ 結合、概念空間の探索
- 新たに抽出した特徴量 ⇒ 概念空間の変形

◎ 人間より高速、人間には困難な特徴量抽出も  
△ ただし、現在は領域特異的(タスク特化型)

# ML・DLと創造性：プロセス重視

Bodenの分類が適用可能

- 既存の特徴量 ⇒ 結合、概念空間の探索
- 新たに抽出した特徴量 ⇒ 概念空間の変形

◎ 人間より高速、人間には困難な特徴量抽出も  
△ ただし、現在は領域特異的（タスク特化型）



1: 人間の方が得意な領域・タスクはあるか？

→ あればその要因は何か？

2: いずれ霊感的ケースも扱えるようになるか？

→ 複数の領域の横断に加えて何が必要か？

# ML・DLと創造性：靈感的ケース

Boden (2014) 松尾 (2015) など：

外部との偶然的な相互作用が生じるシステム

- ケクレのような驚くべきケースも可能なのか？
- そもそも spontaneity 等をどう評価するのか？

3: 知識や経験からの距離の評価

→ 人間よりML・DLの方が得意かもしれない

△ ただし、Black Box 問題

# ML・DLと創造性：靈感的ケース

Boden (2014) 松尾 (2015) など：

外部との偶然的な相互作用が生じるシステム

- ケクレのような驚くべきケースも可能なのか？
- そもそも spontaneity 等をどう評価するのか？

3: 知識や経験からの距離の評価

→ 人間よりML・DLの方が得意かもしれない

△ ただし、Black Box 問題



4: ML・DLの理解可能性についての理論構築

5: 集団内の知的オーソリティへの比較(認識的依存)

# ML・DLと創造性：主体重視

将来は、自律性・人格、本物の理解 etc.を備えるかも  
・・・よくある話？

これまでの考察を踏まえて、別の方向で



## 6：ML・DLとの知的創造活動の分業

プロセス面のML・DLで実現可能な範囲を明らかに

⇒ 人間が主体面で発揮しうる創造性の余地を把握

⇒ 人間とML・DLをうまく組み合わせる

- 組み合わせ最適化問題はML・DLが得意？
- 善き生の構成要素として何をどこまで残す？（倫理へ）

# まとめに代えて

- ML・DLと知的創造性をめぐる哲学的考察はスタートしたばかり
- 課題も数多く、洗い出しの段階
  - いくつかは、ML・DLについての他の認識論的問題（Black Box化、知識は持てるか）とも関わる
  - また、従来の認識論的問題（知識の定義、知的権威、認識的分業・依存）と関わるものもある
  - 課題の中には、ML・DLによってこそうまく解決されるものもあるかもしれない
- こうした点がML・DLにまつわる認識論を興味深いものにしていく

## 文献

- Alfano, M. (2013). *Character as Moral Fiction*, Cambridge University Press.
- Baehr, J. (2011). *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. Oxford University Press.
- Baehr, J. (2018). Intellectual Creativity. In B. Gaut and M. Kieran eds. (2018).
- Boden, M. A. (1995). Could a Robot Be Creative—And Would We Know? In K. M. Ford, C. Glymour and P. J. Hayes eds. (1995). *Android Epistemology*. AAAI Press/The MIT Press.
- Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms, second edition*. Routledge.
- Boden, M. A. (2014). Creativity and Artificial Intelligence: A Contradiction in Terms? In E. S. Paul and S. B. Kaufman eds. (2014).
- Boden, M. A. (2016). *AI: Its Nature and Future*. Oxford University Press.
- Gaut, B. and Kieran, M. eds. (2018). *Creativity and Philosophy*. Routledge.
- Gaut, B. (2018). The Value of Creativity. In B. Gaut and M. Kieran eds. (2018).
- Kieran, M. (2014). Creativity as a Virtue of Character. In E. S. Paul and S. B. Kaufman eds. (2014).
- Kronfeldner, M. E. (2009). Creativity Naturalized. *The Philosophical Quarterly*, Vol. 59, No. 237: 577-592.
- Kronfeldner, M. E. (2018). Explaining Creativity. In B. Gaut and M. Kieran eds. (2018).
- Paul, E. S. and Stokes, D. (2018). Attributing Creativity. In B. Gaut and M. Kieran eds. (2018).
- Prinz, J. J. (2009). The Normativity Challenge: Cultural Psychology Provides the Real Threat to Virtue Ethics. *The Journal of Ethics*, 13: 117-144.
- Ratti, E. (forthcoming). Phronesis and Automated Science. In Sterpetti, F. and Bertolaso, M. eds. *Will Science Remain Human?* Springer
- Schubbach, A. (2019). Judging Machines: Philosophical Aspects of Deep Learning. *Synthese*.  
<https://doi.org/10.1007/s11229-019-02167-z>.
- 稲葉振一郎『AI時代の労働の哲学』、講談社メチエ、2019年
- テレンス・J・セイノフスキー『ディープラーニング革命』、銅谷賢治監訳、ニュートンプレス、2019年
- 松尾豊『人工知能は人間を超えるか—ディープラーニングの先にあるもの』、角川EPUB選書、2015年