

心理学の知見に基づく意識モデルの構築

Building a model of consciousness based on psychological findings

渡邊俊樹¹⁾

指導教員 亀田 弘之²⁾, 研究協力者 安藤大智³⁾

- 1) 東京工科大学大学院 バイオ・情報メディア研究科 コンピュータサイエンス専攻 思考と言語研究室
- 2) 東京工科大学 コンピュータサイエンス学部
- 3) 東京工科大学大学院 バイオ・情報メディア研究科 コンピュータサイエンス専攻 岩下研究室

人工知能, 心理学, モデル化, 心のモデル, 意識

1 はじめに

1.1 背景

近年、AI は急速に発展し、スマートスピーカーや自動運転技術など様々なものに搭載され、より身近なものになってきたといえる。しかしそれらと同時に、AI の判断ミスや、人間の倫理にそぐわない選択をしてしまい、問題になっている事例も取り上げられるようになった。[1][2]

1.2 本研究の目標

今後 AI が更に発展し、より広く普及するためには、倫理的な問題を考慮する事のできる AI が必要になると考えられる。

本研究では倫理的な判断を行える AI を制作するための一つの基準として、倫理的な判断を行える、コンピュータに搭載可能なモデルの構築を目標とする。

2 既存研究

2.1 心理学

心理学の分野においての意識の取り扱いは、フロイトやユングなどの心理学が挙げられ、ユングは心を意識、個人的無意識、普遍的無意識の三層で説明し、フロイトは局所論（意識、前意識、無意識）と後期では構造論（イド、自我、超自我）で説明した。両者共に人の心を構造的に説明し、内部の構造

や動きなどからどのような原理で人が行動しているのかについて研究した。[3][4]

2.2 脳科学・神経学

脳科学、神経学の分野では、意識を医学的な意識と心理学的な意識の両方を取り扱い、機械が意識をもつために必要な要素を論じた論文も存在。[5]

2.3 情報工学

この分野では、心のモデルを用いてコンピュータに実装しようとする試みがいくつかあり、受動意識仮説に基づいた心のモデルや、心の全体像を意識したモデルを構築した研究など、コンピュータに実装可能なモデルが提案されている。[6][7]

2.4 医学

医学分野では意識を覚醒している、や意識がある、などの覚醒レベルとしての意識があり、意識の鮮明度によってレベルが存在する。[8]

1 本研究

1.1 モデルの構築法

本研究では、人間の価値観・倫理観・地域固有の文化・風習を理解、考慮した行動ができるような、AI に搭載できる意識モデルの構築を目指す。倫理観や文化などを含んだ人間の考え方や判断を基準とした意思決定システムとしての意識の構築を目指す。それらを実現するためにそれらを考慮したユ

ング心理学で扱われる意識をモデル化する。

1.2 意識の機能

ユング心理学で用いられている意識、個人的無意識、普遍的無意識を意識の機能として捉え、それぞれがどのような役割を持っているのかをまとめると、表1が得られた。

表 1 意識の機能

機能	役割
意識	認識した情報に関連する情報を無意識から引き出すことができる
	欲求, 欲望を引き出すことができる
	欲求, 欲望を抑圧することができる
個人的無意識	コンプレックス (欲求, 欲望が経験, 教育などによって強く抑圧された物)
	抑圧された欲求, 欲望 (現在意識されていない欲求, 欲望)
普遍的無意識	人間の本能 (母性本能や性欲などの人類が共通して持っている元型)

表1に示した役割は、相互に影響しあっていると考えられ、ソフトウェアとして考えた際のデータの流れを考え、必要に応じ機能を追加した結果、図1のコンピュータへの実装を想定した意識のモデルが得られた。

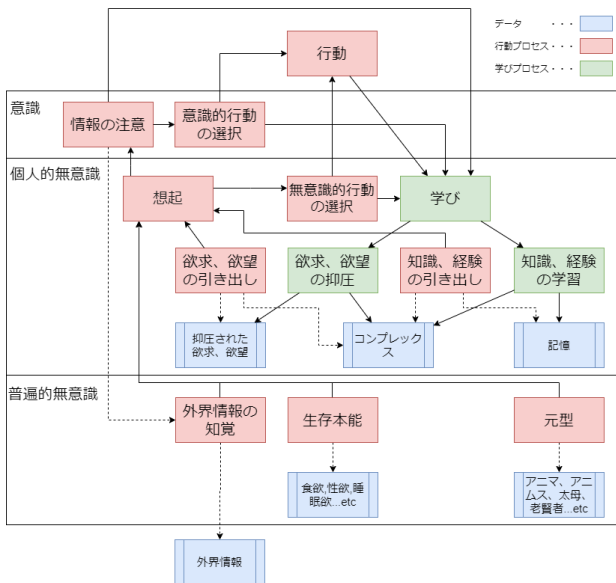


図 1 意識モデル

1.3 意識モデルの想定動作

提案したモデルが動作した場合の想定動作を考える。例えば、空腹の場合生存本能が空腹を知覚し、想起によって食べ物やその在処などの連想情報を引き出し、プライミング記憶などが無意識の行動に送られて行動が選択され、エピソード記憶が情報注視に送られ、外界情報の知覚とエピソード記憶の情報が結びつけられる。その後、意識的

行動選択が行動を選択し、行動プロセスが意識的行動と無意識的行動の中から最終的な行動を選択する。選択された行動と行動結果は、学習に送られ、欲求欲望の抑圧と知識経験の学習に分けられ、それぞれの記憶の形に整えられてから格納される。

1.4 検証

現在、共同研究者協力の元、プロトタイプ型開発にて ML エージェントへのモデルの実装、検証を行っている。

2 おわりに

今回得られたモデルは、現時点では検証中のため、評価はできていないが、心理学からコンピュータへ実装を可能とする手法はソフトウェアの新たな開発手法になると考えている。

参考文献

- [1] “Moral Machine,” *Massachusetts Institute of Technology*. <https://www.moralmachine.net/> (accessed Jul. 26, 2021).
- [2] Y. Matsuo *et al.*, “人工知能と倫理,” *AI Mag.*, vol. 7, no. 2, pp. 70–79, 1986.
- [3] 河合隼雄, *ユング心理学入門*. 培風館, 1967.
- [4] 妙木浩之, *フロイト入門*. ちくま新書, 2000.
- [5] S. Dehaene, H. Lau, and S. Kouider, “What is consciousness, and could machines have it?,” *Science* (80-.), vol. 358, no. 6362, pp. 486–492, Oct. 2017, doi: 10.1126/science.aan8871.
- [6] N. OKADA, “A Computational Model of the Mind,” *IEICE ESS Fundam. Rev.*, vol. 1, no. 4, pp. 30–45, 2008, doi: 10.1587/essfr.1.4_30.
- [7] T. Maeno, “How to Make a Conscious Robot-Fundamental Idea based on Passive Consciousness Model-,” *J. Robot. Soc. Japan*, vol. 23, no. 1, pp. 51–62, 2005, doi: 10.7210/jrsj.23.51.
- [8] H. K. Walker, W. D. Hall, and J. W. Hurst, “Clinical Methods, 3rd edition,” Butterworth, 1990.