

【新学術領域研究（研究領域提案型）】 複合領域



研究領域名 人工知能と脳科学の対照と融合

沖縄科学技術大学院大学・神経計算ユニット・教授

どうや けんじ
銅谷 賢治

研究課題番号：16H06561 研究者番号：80188846

【本領域の目的】

本領域の目的は、それぞれの研究の高度化のなかで乖離して行った人工知能研究と脳科学研究を結びつけ、両者の最新の知見の学び合いから新たな研究ターゲットを探り、そこから新たな学習アルゴリズムの開発や脳機構の解明を導くことである。

感覚運動情報の背後にある構造を捉える表現学習、それらの変化を予測する内部モデル学習、さらに予測結果の評価による探索学習について、それぞれを確実に効率よく実現するアルゴリズムとその脳での実現を探るとともに、それらをつないだ全脳アーキテクチャにならった柔軟な人工知能システムの実現をめざす。

【本領域の内容】

人工知能研究と脳研究は、「電子回路で知能を実現するには脳での実現法にこだわる必要はない」という視点と、「脳のような高度な知能の実現例があるのだからそれに学ばない手はない」という視点から、接近と乖離を繰り返してきた。例えばパターン認識の黎明期には、脳の視覚野の解剖学、生理学的知見をもとに、多層神経回路モデルとその学習アルゴリズムが開発された。そこから芽生えた機械学習理論は、カーネル法やベイジアンネットといった形で、脳とは離れた形で高度化して行ったが、近年、深層神経回路モデルが大量のデータのからの学習で非常に高い性能を示したことから、脳に学んだ人工知能への期待が再び高まっている。

例えば、これまでの脳の実験的・理論的研究は、小脳、大脳基底核、大脳皮質はそれぞれ、与えられた手本を再現する教師あり学習、探索により報酬を最大化する強化学習、データの構造を捉える教師なしの表現学習にそれぞれ関わっていることを示唆している（図1）。一方、囲碁の世界王者を破るという画期的な成功をおさめたプログラムは、棋譜からの教師あり学習、自己対戦による強化学習、深層神経

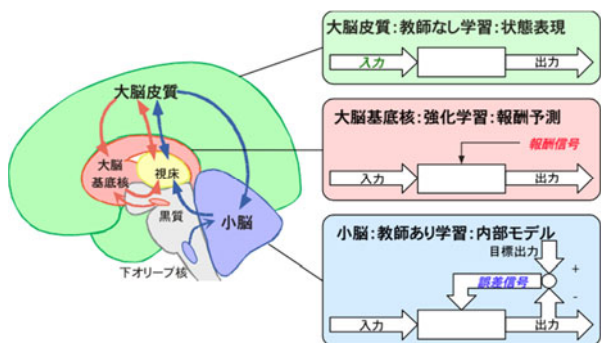


図1：小脳、大脳基底核、大脳皮質の学習アルゴリズム (Doya, 1999)。

回路による表現学習を巧妙に組み合わせることで、非常に高度な知能が実現され得ることを実証した。このことは、異なる学習要素を状況に応じて柔軟に組み合わせる脳の仕組みを解明することが、特定の課題だけでなく多様な問題に対処可能な汎用人工知能の実現につながることを示唆するものである。

本領域は、人工知能と脳科学の先端的な研究者が両者の最新の知見を学び合い、緊密な議論を行うなかで新たな研究ターゲットを探り、新たな学習手法の開発や脳機構の解明を導くことをめざす。教師あり学習による内部モデルの獲得、強化学習による行動探索と評価、多層神経回路による表現学習など、それぞれの学習機構の高度化と脳での実現の解明を進めるとともに、それらを統合する脳の仕組みを探り、人間のように柔軟な行動やコミュニケーションが可能な人工知能の実現のための研究を展開する。

【期待される成果と意義】

人工知能と脳科学の融合研究により、深層神経回路の学習過程の理解と学習パラメタの自動制御、データ効率の高い学習による人型ロボットの制御、階層予測モデルによる人の行動意図の推定などの実現が期待される。より長期的には、予測モデル学習、探索と評価、深層表現学習を組み合わせた全脳アーキテクチャの理解により、汎用人工知能の設計指針の導出と、学習の乱調としての精神疾患の理解への貢献が期待される。

また人工知能と脳科学を融合する新たな人材育成と国際ネットワーク形成のため、サマースクール、ハッカソン、若手研究者の海外派遣などを企画し、新たな融合領域を切り開く人材の輩出をめざす。

【キーワード】

深層学習 (deep learning): 多層の人工神経回路モデルによって、データの持つ統計的な特徴を単純なものから高次なものへと順次抽出し、画像や音声などの複雑なパターン識別や予測を可能にする学習方式。

強化学習 (reinforcement learning): 動物、ロボット、プログラムなどが、外界の状態に応じて行動を行い、結果の良し悪しを表す報酬信号をもとに適切な行動則を探索的に学習する方式。その理論をもとに、大脳基底核とドーパミンの機能解明が進んだ。

【研究期間と研究経費】

平成 28 年度－32 年度
1,119,100 千円



Title of Project : Correspondence and Fusion of Artificial Intelligence and Brain Science

Kenji Doya

(Okinawa Institute of Science and Technology Graduate University, Neural Computation Unit, Professor)

Research Project Number : 16H06561 Researcher Number : 80188846

【Purpose of the Research Project】

The purpose of this research project is to bring together artificial intelligence research and brain science research, which separated apart while their own sophistication, and to promote developments of novel learning algorithms and deeper understanding of brain mechanisms. We aim to develop efficient algorithms and clarify brain's realization of **supervised learning of internal models, reinforcement learning by exploration and evaluation**, and representation learning to facilitate them. We further try to realize flexible artificial intelligent systems based on the understanding of the whole-brain architecture for flexibly linking those learning modules.

【Content of the Research Project】

Artificial intelligence and brain science have had a swinging relationship of convergence and divergence. In the early days of pattern recognition, multi-layer neural networks based on the anatomy and physiology of the visual cortex played a key role, but subsequent sophistication of machine learning promoted methods that are little related to the brain. Recently, however, the remarkable success of deep neural networks in learning from big data has re-evoked the interests in brain-like artificial intelligence.

Evidence suggests that the cerebellum, the basal ganglia, and the cerebral cortex are respectively specialized in supervised learning, reinforcement learning, and unsupervised representation learning, (Fig. 1). On the other hand, the recent success of artificial intelligence in beating a Go world champion has demonstrated that exquisite

combination of supervised learning, reinforcement learning, and representation learning by deep neural networks can achieve human level or higher intelligence.

This project brings together leading researchers in artificial intelligence and brain science, let them learn the latest developments in each other's fields, and promote identifying the seeds of novel discovery and developments. In addition to developing novel algorithms for supervised learning, reinforcement learning, and deep representation learning and clarifying their implementation in the brain, we try to elucidate how those learning modules are flexibly integrated in the brain depending on the behavioral needs, and utilize the knowledge for developing artificial intelligence that allows human-like flexible actions and communication.

【Expected Research Achievements and Scientific Significance】

We expect to achieve better understanding of the mechanisms of deep neural network learning, data-efficient learning algorithms for humanoid robots, and hierarchical models to understand human intentions, for example. In a long run, we try to understand the self-organizing mechanism of learning modules in the brain and to derive the design principles for realizing artificial general intelligence.

The project will also aim to produce young scientists who can lead the crossing edges of artificial intelligence and brain science, through training programs such as summer schools, hackathons and international exchange programs.

【Key Words】

Deep neural network: a multi-layer network for discovering statistical features hidden in the data, from simple ones to gradually complex ones. It is widely used for image and speech recognition.

【Term of Project】 FY2016-2020

【Budget Allocation】 1,119,100 Thousand Yen

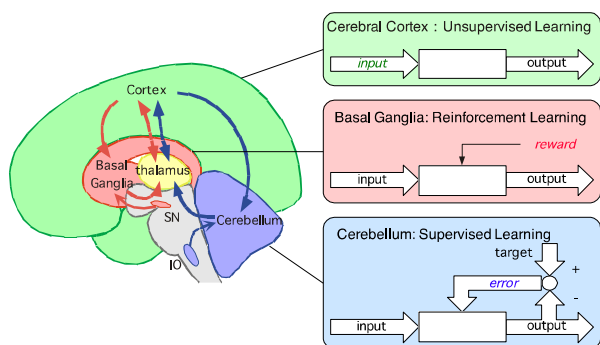


Fig. 1: Learning algorithms of cerebellum, basal ganglia, and cerebral cortex (Doya, 1999).