

# 地球科学分野における機械学習の近年の研究動向について

磯 真一郎

深田地質研究所

## Recent Research Trends in Machine Learning in Geosciences

ISO Shinichiro

Fukada Geological Institute

要旨：近年、人工知能、特に機械学習、深層学習の理論と技術が飛躍的な発展を見せており、多種多様な領域においてその適用手法についても研究が活発になされている。この傾向は深田地質研究所が対象としている固体地球科学の研究分野でも例外ではない。本稿では2020年と2021年に開催された地球科学研究に関連する二つの国際学会での研究講演と地球科学に限らず過去5年の発表論文の件数を検索エンジンによって調査することで、機械学習を用いた研究活動をまとめた。その結果、固体地球科学研究においても適用目的は様々であるが、弾性波探査関連分野、衛星画像や航空写真による地形関連分野で比較的多くの機械学習の適用研究が多く発表されていた。機械学習のアーキテクチャ・アルゴリズムに関しては畳み込みニューラルネットワークを用いた事例が多かったが、強化学習など新たな手法についての活発化も見られた。さらに、これらの論文発表件数を調査した結果から、今後の地球科学分野における研究動向について考察した。

キーワード：人工知能、深層学習、機械学習、地球科学、固体地球科学、研究活動

Abstract: In recent years, the theory and technology of artificial intelligence, especially machine learning and deep learning, have shown dramatic development, and research on its application methods in various fields has been actively conducted. This trend is not an exception in the research field of solid earth science, which is the interest of the Fukada Geological Institute. This paper summarizes the research activities using machine learning by using a search engine to survey the number of research presentations given at two international conferences related to geoscience research held in 2020 and 2021 and the number of papers published in the past five years, not limited to geoscience. As a result, although the purpose of applying machine learning in solid earth science research varies, a relatively large number of studies involving machine learning were published in the fields related to seismic explorations and topography studies using satellite images and aerial photographs. As for the architecture and algorithms of machine learning, most of the papers were based on convolutional neural networks, but there was also a growing interest in new methods such as reinforcement learning. In addition, based on the survey results of the number of papers published, future research trends in earth science are discussed.

Keywords: artificial intelligence, deep learning, machine learning, solid earth science, research trend

## 1. はじめに

今日、自動車の自動運転や顔認証などの報道ほかで、人工知能、機械学習、深層学習と呼ばれる技術を目にすることが多い。これらの技術は近年、飛躍的な発展をしており、幅広く様々な産業や研究領域に適用が進んでいる。

地球科学分野も例外ではなく、機械学習技術を適用することによって、今まで不可能とされていたことの実現化や、新しい発想が生まれると考えられる。

本稿では、最近の国際学会での講演発表事例を基に、特に機械学習が地球科学における貢献状況を示し、今後の研究の方向性を洞察する一助になればと考えている。

なお、地球科学には一般に気圏、水圏、地圏など様々な分野を含むが、本稿では本研究所が主として対象としている「地球システム」を「地球科学」あるいは「固体地球科学」として表現している。この分野には、探査工学、土木工学、地質学、地球物理学、固体惑星科学をはじめ多くの学問研究領域を含め、気象や海洋、恒星をはじめとする宇宙に関連する領域は対象としていない。

また、多くの人工知能・機械学習のアーキテクチャ・アルゴリズムについての略称については補遺を参考願いたい。

## 2. 人工知能・機械学習の概要

人工知能 (AI; Artificial Intelligence) という用語が1955年に開催されたダートマス会議で初めて使用され、研究が長く続けられてきた。その後数度の隆盛と低迷の時期を繰り返し、近年は第4次人工知能ブームと言われている。その大きな理由の一つは、深層学習と呼ばれるニューラルネッ

トワークの多層化を特徴とする機械学習のアプローチである。深層学習は大量のデータを用い、予測などを行う。2000年代のインターネットの普及とコンピュータなどの計算機資源の価格の低下により、大量のデータを容易に利用できることとなったことも要因とされている (松尾, 2015)。

### 2.1 人工知能と機械学習、深層学習

人工知能、機械学習、深層学習はこの順序で包含関係を持つ異なる概念ではあるが、時に厳密な区別をされず言及されることも多い。

人工知能 (AI, Artificial Intelligence) に関連する分析技術として機械学習 (ML, Machine Learning) があり、その一つの技術に深層学習 (DL, Deep Learning) がある。深層学習は論理や言語で説明が困難、すなわち従来のアルゴリズムや仕様では記述が困難である特徴をコンピュータ自身がデータに基づき学習し、動作することが可能となることが特長である。深層学習は今世紀に入り、物体認識でのブレークスルーを皮切りにいわゆるAIブームを作り出した。(松尾, 2015)

本稿では深層学習を包含した機械学習を論ずる。各々のアーキテクチャ/アルゴリズムの種別をみれば深層学習あるいは機械学習であるかは判別できる。

### 2.2 学習方法による機械学習の分類

機械学習には様々な観点からの分類方法があるが、一般には学習の方法によって、「教師なし学習」、「教師あり学習」、さらに「強化学習」を加えて以下の3つの類型にまとめられる。

教師あり学習とは、正解が与えられているデータ、例えば、犬の写真に「犬」という名称 (ラベル)、を学習することにより、未知のデータに対して解答を予測する。これは、写真の判定とい

った分類や数値を予測する回帰問題に適用される。

教師なし学習とは、決まった解がないデータ自体を学習する機械学習であり、予定されている正解を予測するものではない。多くの場合データの特徴を示すクラスタリングや次元削減といったことに用いられている。(岡谷, 2015)

強化学習とは、ある環境下で、エージェントが行動を自ら決定し得られる報酬を大きくすることで学習を行う。

これらの異なる3つの学習に対してそれぞれ様々な機械学習のアーキテクチャやアルゴリズムが発見・発明されている。

### 2.3 人工知能の研究動向

図1はAI関連主要国際会議への参加者数推移を示すものである、2010年代に入り、指数関数的に参加者が増加しており、この顕著な増大が現在の人工知能のブームを反映している。

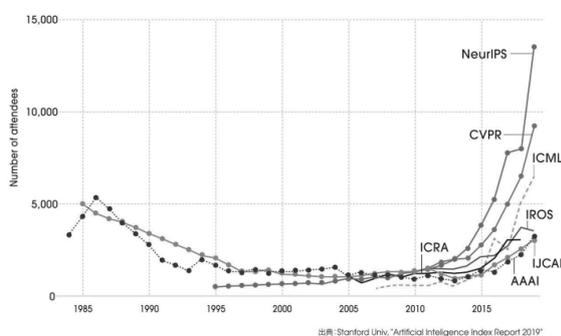


図1 AI関連主要会議への参加者数推移 (AI白書2020 抜粋, 凡例のAI関連主要会議の略称は補遺表7を参照)

研究活動の一つ成果として、特許出願がある。図2に示す通り、特許出願件数においても特許出願件数も2010年代から大幅な伸びを示しており、人工知能学会での参加者と同様な現象である。

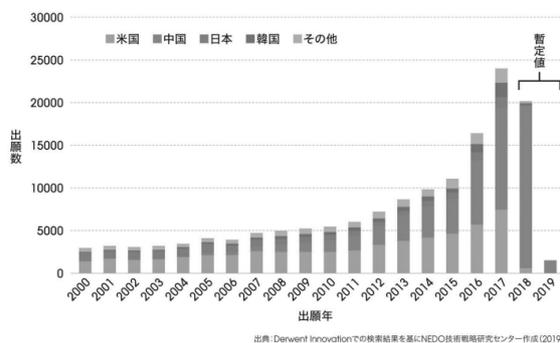


図2 正解の特許出願数推移 (AI白書2020 図2-6-1より抜粋)。研究論文と同様、特許出願件数も同様な指数関数的に増加していることがわかる

### 3. 地球科学分野での機械学習適用

前項では2010年代から現在まで人工知能に関して研究活動の活発化を示したが、これはすべての研究領域を合計したものである。

本稿で対象としている「地球科学」でも様々な専門分野で機械学習が適用されていると考えられる。

#### 3.1 機械学習の適用に対しての地球科学分野の特徴

近年の機械学習の特徴の一つは大量のデータを扱うことである。この点で地球科学分野がほかの研究分野と異なる点は、研究対象が複雑、砂渡井単純な物理モデルでは説明が困難であることに加えて、教師あり学習に不可欠である正答(ラベル)がある良質なデータの取得が難しい、あるいは多額の労力や資金を要することが多いことであると、考える。特に正答を作成する際に多くの人的バイアスが含まれる問題もあるようである。

例えば地下の空洞の有無を物理探査的手法により推測したとしても、実際に開削するまでは空洞の有無は結論付けられることはなく、当然なが

ら開削には多くの費用を要する。また、弾性波探査や空中写真や衛星画像のように、情報量自体は大量に取得する事例も多く、おのずと計算機資源の必要量も大きくなる。同じデータ、同じ解析者が対応しても、往々にして同じ結果にはならないケースもある。このような特徴から地球科学分野、特にフィールドデータに対しての機械学習の新たな技術の適用は、比較的ハードルが高いと考える。

しかし、他分野での成果・成功事例から、ツールとしての機械学習美術が本分野でも多く取り入れられてきていることを次項以下にしめす。

### 3.2 地球科学領域での研究発表に見る機械学習

機械学習がどのように適用されているかを2020年と2021年に開催された2件の大規模な研究講演会、日本地球惑星科学連合2021年大会(JpGU, 2021), (以下 JpGU2021), および米国物理探査学会2020年大会(SEG, 2020), (以下 SEG20)を対象に、適用の主目的、機械学習の手法(アーキテクチャ・アルゴリズム)の観点ほかで研究発表件数の調査を行った。なお、適用の主目的については、必ずしも一般に認知された分類ではなく、著者が本調査のため便宜的に整理したものである。

JpGU2020は日本地球惑星科学連合が2021年6月に開催された学術講演会である。Covid-19感染対策のため、オンラインにて実施された。その研究対象範囲は地球惑星科学のみならずに関連する領域を含む幅広い分野である。

SEG20は米国物理探査学会が2020年10月にもオンライン(バーチャル)大会で開催された。SEG20は開催主体が示すように、米国における地球物理学、特に物理探査工学に関する講演発表が主体である。

本稿では先に述べた通り、そのうち、当研究

所を対象とする研究分野についての調査結果を表1から表4にまとめた。調査結果である論文件数は著者が、実際に講演に参加、あるいは講演論文要旨ほか目を通して判別したものである。一講演で複数の課題を取り扱っている事例もおおく、講演総数は各の表の間で必ずしも一致しない。

表1はJpGU2021において講演された適用目的にまとめたものである。講演件数は専門分野の重複を許すと40件あったが、実際は29件の講演が機械学習を主たるテーマとして研究発表を行っていた。大会全体で講演数をしらべたところ152件(分野での重複あり)であった。多くは気象や海洋でのデータ解析に機械学習が適用されていた。地震(弾性波)関連のデータ解析に適用されて講演が最も多く(合計14件)、次いで地形解析に適用されて講演(6件)であった。

地形解析は衛星や航空機から測定された画像や標高データを利用しているケースが多くみられた。地震関連のデータ解析とともに、どちらも大量のデータが取得される分野である。大量データの取り扱いに威力を発揮する機械学習が適用されていることを示していると考えられる。

表2にはこれらの講演についてさらに学習モデルとアーキテクチャ・アルゴリズムごとに発表件数を分類した。これから、教師あり学習の一つであり、特に画像データに対してよく用いられている畳み込みニューラルネットワークに関する講演が多数発表されていることがわかる。畳み込みニューラルネットワーク、CNN、(Convolutional Neural Network)の着想(Fukushima,1988)は比較的以前からあったが、AlexNet(Krizhesky, et. al., 2012)が発表されてから、その優位性が広く認識され、以来様々な研究結果、成果が発表されている。

また、教師なし学習では様々なクラスタリング手法が多く適用されている。

表 1 JpGU2021 における機械学習を適用した講演発表での研究領域別分類 (著者しらべ).

機械学習適用の主目的	講演件数 (件)
波形解析 (地震)	6
震源解析	4
初動検出 (地震)	2
測定システム (地震)	1
地震予知	1
地形解析	6
地質解析	3
組成解析 (岩石)	3
応力解析	1
作業効率 (掘削)	1
津波予測	1
合計	29

表 3 から 5 には SEG20 での講演件数の調査結果をまとめた. SEG20 では 760 件の講演があり (SEG, 2020) そのうち, 著者が調べたところ, 機械学習を含む講演は項目による重複を含めて 86 件であった.

SEG は地球物理と物理探査工学に関する研究が多く, 対象とするデータの計測・探査手法について表 3 に示した. 一般に弾性波探査に関連する研究講演が多いこともあるが, 機械学習も弾性波探査に適用する事例がほかの手法に比べても多いことがわかる.

表 2 表 1 で示された JpGU2021 講演での機械学習のモデルとアーキテクチャ・アルゴリズムの分類 (著者調べ).

教師データの有無	学習モデル	アーキテクチャ・アルゴリズム	講演件数 (件)
教師あり学習	従来型 NN	NN, ANN	2
		DNN	2
	畳み込み NN	CNN	7
		VGG19	1
		U-Net	1
	再帰的 NN	RNN	1
LSTM		2	
教師なし学習	生成モデル	GAN	1
		クラスタリング	k-means
		t-SNE	1
		HCA	4
		RF	1
		GBDT	1
	MDS	1	

表 3 SEG20 での機械学習に関する講演が扱った計測・探査データの種別

対象とするデータの計測・探査手法	講演数 (件)
弾性波探査	71
超音波	1
検層	4
電磁探査	1
地磁気地電流法 (MT 法)	2
断層撮影画像	4
その他・不明	3

また、機械学習を適用した主目的について、を表4に示す。逆解析（波形逆解析を含む）、およびマイグレーションを目的とする講演が合わせて30件あり、多数を占めている。これは、これまでの数値計算法に、機械学習の手法を導入することで、計算能力の向上を図るという試みも含まれると考えられる。次いで岩相分析（7件）に用いられているのは、画像認識など機械学習の提供が先行した領域で大きな成果があったものを応用したと考える。

表4 SEG20での機械学習に関する発表の主目的毎の講演数、ただし分類不可能・不明な講演については含まない。（著者調べ）

機械学習適用の主目的	講演数 (件)
弾性波速度	3
初動検出	3
SNの改善	4
波形逆解析	10
マイグレーション	10
逆解析一般	10
岩相分析	7
岩石物理	2

SEG20にて各発表講演で適用されていた機械学習のアーキテクチャ・アルゴリズムについてJpGU2021と同じく講演件数について著者が調査した結果を表5に示す。

表5 SEG20の講演での適用機械学習のアーキテクチャ・アルゴリズムの分類（著者調べ）。

学習タイプ	学習モデル	アーキテクチャ／アルゴリズム	講演数 (件)
教師あり学習	従来型 NN	DNN	2
	畳み込み NN	CNN	28
		U-Net	9
		再帰型 NN	RNN
		LSTM	2
		SVM	1
	*	PINN	2
教師なし学習	クラスタリング	SOM	2
		RF	1
		AE	2
	生成モデル	GAN	7
強化学習		RL	2

\*) PINN は自らシミュレーションにより教師データを作成する場合がある。

最も多用されているのは JpGU2021 と同様、畳み込みニューラルネットワーク（CNN）であり、次いで U-Net であった。

U-Net (Ronneberger, et. al, 2015) は画像セグメンテーションのためのアーキテクチャ・アルゴリズムであり、この分野でのブレークスルーをもたらしたものであり、様々な派生アーキテクチャ・アルゴリズムが存在する。

その他では、GAN (Goodfellow et. al, 2014) ,に関する講演が多いことがわかる。これは JpGU2021 では教師なし学習のほとんどがクラスタリングを適用していたことと比べて大きく SEG20 で大きく異なっている点である。

### 3.3 全ての領域での機械学習アーキテクチャ・アルゴリズムに関する論文件数動向

人工知能・機械学習の技術は、本質を抑えたうえで手を加えた上で、幅広く多種多様な研究領域に適用可能なツールであろう。前節にて、様々な機械学習のアーキテクチャ・アルゴリズムが最近の地球科学分野の国際学会で適用されていたかを示した。今後新しいアーキテクチャ・アルゴリズムも創造されていくだろうが、どのようなアーキテクチャ・アルゴリズムが地球科学分野に限らず、活用状況の把握を試みた。

今日、極めて多くの研究論文や日々発表されており、そのすべての論文で人工知能・機械学習の適用の有無やそのアーキテクチャ・アルゴリズムについてすべて正しく判定するのは困難である。そこで、本稿では、検索エンジンを用い、論文のタイトルにこれら特定のアーキテクチャ・アルゴリズムが含まれる論文を数え上げることで近似的に活動度の理解を試みた。使用した検索エンジンは Google 社が提供している Google Scholar (Google, 2021) である。Google Scholar の正確な検索アルゴリズムやデータベースサイズは不明であること、論文タイトルに正しくアーキテクチャ・アルゴリズムが記述されるとは限らないこと、また応用研究ではない論文も含まれることなど、調査の精度には問題点はあるが、一定の信頼度はおけるとされている (Howland, et al., 2009)。本稿の目的とする各々のアーキテクチャ・アルゴリズムの一般的活動度とみなすことは十分に可能だと考える。

今回、アーキテクチャ・アルゴリズムの検索に用いたワードは完全一致で、それぞれ “deep neural network”, “convolutional neural network”, “recurrent neural network”, “long short-term memory”, “support-vector machine”, “self-organizing map”, “random forest”, “autoencoder”, “generative adversarial

network”, “reinforcement learning”, “t-sne, および t-distributed stochastic neighbor embedding”, “U-net” の 12 パターンである。これらは必ずしもお互いに独立なアーキテクチャ・アルゴリズムではないが、著者が日々目にしてのものから重要とおもわれるものを選択し、論文件数を調査した。図 3 に過去 5 年間 (2016 年から 2020 年まで) の上述の各アーキテクチャ・アルゴリズムをタイトルに含む論文件数の増減度をしめした。2016 年の論文件数を 1 としており、発表論文件数の伸びを示す縦軸が対数表記であることに留意されたい。

さらに各々のアーキテクチャ・アルゴリズムを含む論文件数がこれらの中でどの程度の割合を占めているかについて図 4 に示した。

生成モデルである GAN が 2016 年の論文件数の 500 倍近くと最も増加している。続いて U-Net の約 120 倍、LSTM, (Hochreiter, & Schmidhuber, 1997), はおおよそ 8 倍となっている。図 4 は、各アーキテクチャ・アルゴリズムの占有率の推移を示している。2016 年ではおおよそ 3 割を占めていた SVM, (Vapnik, & Lerner, 1963) が約 8 パーセントと大幅にシェアを減らしている一方、2006 年には占有率が 2 割程度であった強化学習に関する論文がおおよそ 3 割となり、次いで CNN が 23% を超え、この 2 つのアーキテクチャ・アルゴリズムで過半数を占める。

また、同期間でのデータが不十分であったため、本稿の調査には含めなかったが、PINN, Physics-Informed Neural Networks, (Raissi et al., 2019) は 2019 年には 7 件、翌 2020 年に 23 件、2021 年中途で 56 件の論文が被検索可能であった。PINN は支配方程式による拘束プロセスを含んでおり得られた解が物理的な条件を満足すると考えられ、特に物理探査領域での利用がさらに進むことが期待できる。

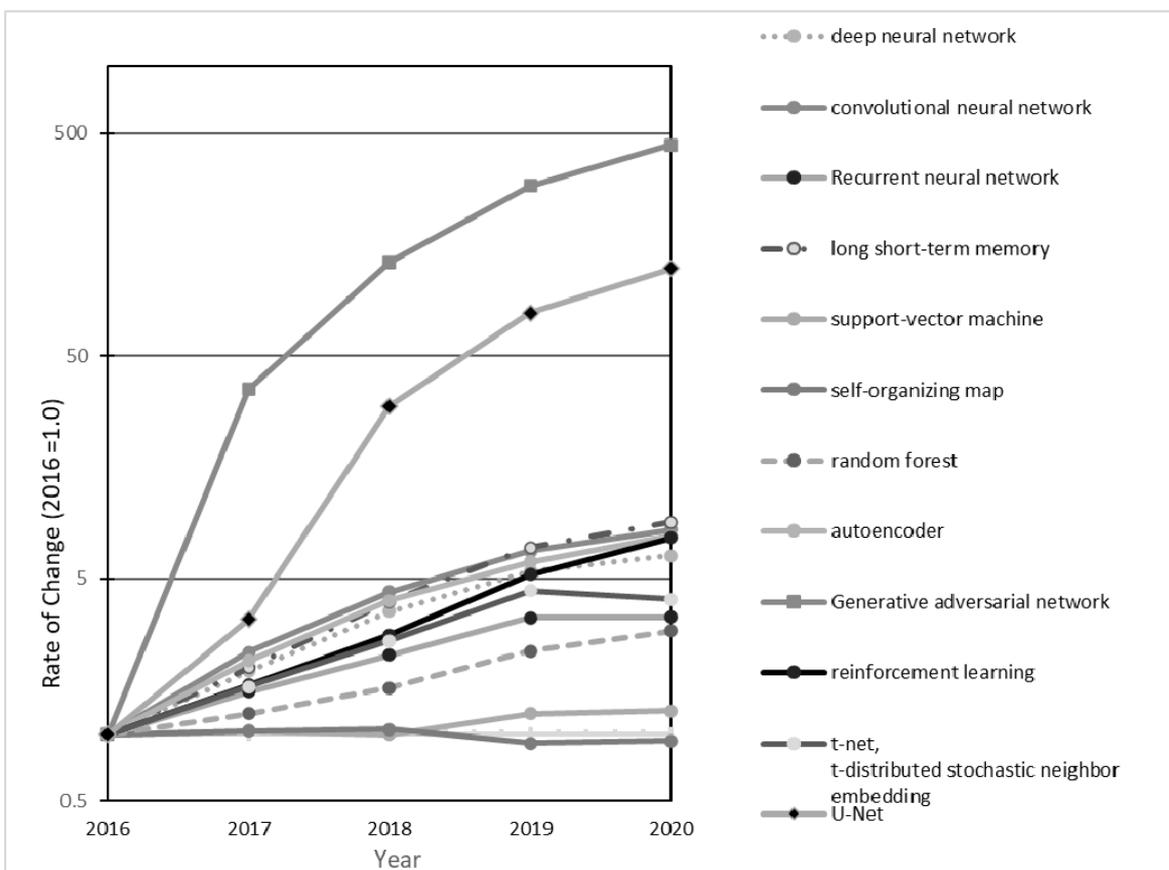


図 3 論文タイトル中のアーキテクチャ・アルゴリズム毎の発表件数の増減.

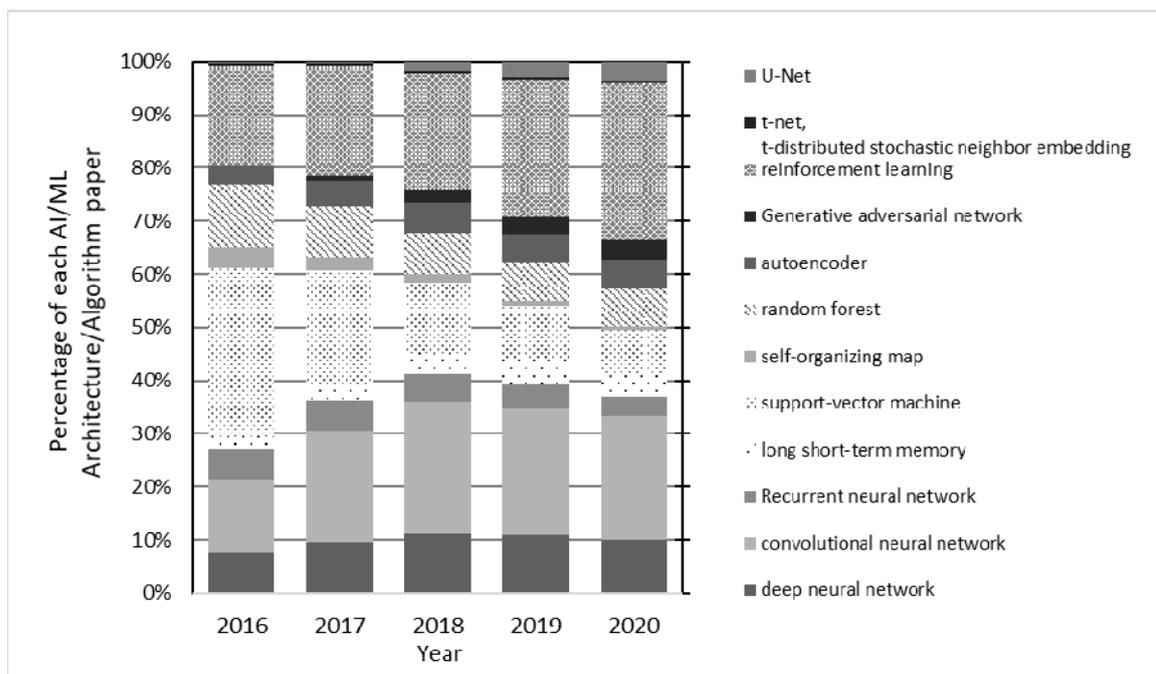


図 4 論文タイトル中のアーキテクチャ・アルゴリズム毎の発表件数の占める割合の変化.

#### 4. まとめと今後の動向

今回は、限られた事例による調査ではあるが；

- 畳み込みニューラルネットワークは依然として地球科学および全般において多数の発表件数がある
- 地球科学分野での機械学習と関係した研究活動は様々な領域で多数行われている
- 強化学習に関する論文件数の割合は、地球科学分野では他分野と比べてちいさい

といった、傾向があることが講演件数と論文件数から観察した。

特に地球科学の研究領域で厄介な質の良いデータ、特に正答（ラベル）付きデータ、を用意する労力が、強化学習といった教師なし学習の手法の適用によって、大きく軽減される可能性がある。生成モデルである GAN もこの問題の解決に有用であろう。PINN も同様である。

今後もすべての地球科学分野において、機械学習の利用は進み、より重要性を増してくると考えている。

#### 文献

岡谷貴之 (2015) : 深層学習 (機械学習プロフェッショナルシリーズ), 講談社サイエンティフィック, 東京, 25-27.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012): ImageNet Classification with Deep Convolutional Neural Networks, *Advances In Neural Information Processing Systems*, 1-9.

公益社団法人日本地球惑星科学連合 (2021); 日本地球惑星科学連合 2021 年大会, [http://www.jpogu.org/meeting\\_j2021/](http://www.jpogu.org/meeting_j2021/), (2021 年 10 月 1

日閲覧).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y., 2014, Generative adversarial nets. *Advances in neural information processing systems*, 27.

Google, (2004), Google Schola, <https://scholar.google.com/>, (2021 年 12 月 28 日閲覧).

Society of Exploration Geophysicists (2020), SEG2020 Annual Meeting, <https://seg.org/AM/2020/> (2020 年 11 月 1 日閲覧).

独立行政法人情報処理推進機構 AI 白書編集委員会. AI 白書 2020(単行本). 株式会社角川アスキー総合研究所.

Hochreiter, S., and J. Schmidhuber., 1997, Long short-term memory. *Neural computation*. **9**, 8,, pp.1735-1780.

Howland, Jared L.; Wright, Thomas C.; Boughan, Rebecca A.; Roberts, Braiam C., 2009, How Scholarly Is Google Scholar? A Comparison to Library Databases. *College & Research Libraries*, **70**-3, 227-234.

松尾豊 (2015) : 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, KADDOKAWA / 中経出版, 東京.

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational Physics*, **378**, 686-707.

Ronneberger, O., Fischer, P. & Brox, T., 2015, U-net: Convolutional networks for biomedical image segmentation. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 9351, 234-241.

Vapnik, V., & Lerner, A., 1963. Pattern recognition using generalized portrait method. *Automation and remote control*, 774-780.

## 補遺

表 6 機械学習アーキテクチャ・アルゴリズムに関連する略称と名称 (英文/和文).

略称	名称 (英文/和文)
NN	Neural Network / ニューラルネットワーク
DNN	Deep Neural Network / ディープニューラルネットワーク
CNN	Convolutional Neural Network / 畳み込みニューラルネットワーク
RNN	Recurrent Neural Network / 回帰型ニューラルネットワーク
LSTM	Long Short-term Memory / 長・短期記憶
SVM	Support-Vector Machine / サポートベクターマシン
PINN	Physics-Informed Neural Networks / 物理情報ニューラルネットワーク
SOM	Self-Organizing Map / 自己組織化写像
RF	Random Forest / ランダムフォレスト
AE	Autoencoder / 自己符号化器
GAN	Generative Adversarial Network / 敵対的生成ネットワーク
RL	Reinforcement Learning / 強化学習
t-SNE	T-distributed Stochastic Neighbor Embedding / t分布型確率的近傍埋め込み法
k-means	k-means clustering / k平均法
HCA	Hierarchical Clustering Analysis / 階層的クラスタリング解析
GBDT	Gradient Boosting Decision Tree / 勾配ブースティング決定木
MDS	Multi Dimensional Scaling / 多次元尺度構成法

表 7 著名人工知能関連学会の名称

略称	名称
AAAI	Association for the Advancement of Artificial Intelligence
IJCAI	International Joint Conference on Artificial Intelligence
NeurIPS	Conference on Neural Information Processing Systems
CVPR	Conference on Computer Vision and Pattern Recognition
ICML	International Conference on Machine Learning
ICRA	IEEE International Conference on Robotics and Automation
ACL	Association for Computational Linguistics