

ガートナー「日本における未来志向型インフラ・テクノロジーのハイプ・サイクル：2020年」に登場するエッジ AI 関連インフラ用語の解説

溝口 隼人

Edge AI Research Center 研究員
h.mizoguchi@edge-arc.jp

工藤 俊太郎

Edge AI Research Center 研究所長
kudo@edge-arc.jp

1 はじめに

先日、ガートナーから「日本における未来志向型インフラ・テクノロジーのハイプ・サイクル：2020年」が発表された。当該ハイプサイクルにおいてエッジ AI は「黎明期」に分類されているが、エッジ AI の普及は単独で成しえるものではなく、他のインフラ技術の発展と協調も必要であると考えられる。そこで本稿では、エッジ AI と関連すると思われるインフラテクノロジー用語をピックアップし、用語の概要とエッジ AI との関わりについて解説する。

2 ハイプサイクルとエッジ AI 関連用語

「日本における未来志向型インフラ・テクノロジーのハイプ・サイクル：2020年」を図 1 に示す。

以下では、図中にプロットされている用語のうち、「エッジコンピューティング」、「AIOps プラットフォーム」、「5G」、「OT と IT の融合」の 4 つを取り上げ、エッジ AI との関連という観点から解説する。

2.1 エッジコンピューティング

2.1.1 概要

エッジコンピューティングとは、ネットワークの終端（エッジ）で計算を実行できるようにするアーキテクチャを指す。末端のデバイス上や、終端付近に設置したサーバ上で計算処理を行うような構造が該当する。

モノのインターネット（Internet of Things, IoT）の概念の普及により、様々なデバイスがネットワークに繋がれるようになってきているが、その数は今後さらなる増加が見込まれている。例えば [2] のレポートでは、IP ネットワークに接続されるデバイス数は、2018 年の 184 億から 2023 年には 293 億まで増加すると予想されている。

現状 IoT デバイスの活用においては、デバイスで収集したデータをクラウド環境に送信して活用するクラウドコンピューティングのアーキテクチャの採用が一般的である。しかし、デバイス数の増加により、クラウド環境との通信に起因する通信量、通信遅延、またセキュリティや機器管理コストなどが問題

となり得ることから、近年はエッジコンピューティングが注目を集めている。[3] では、エッジコンピューティング市場規模は 2027 年までに 434 億米ドルに達し、その期間の年平均成長率は 37.4 % であると予測している。

ネットワークを経由してクラウド上の計算資源を利用するクラウドコンピューティングと比較すると、以下のようなメリットが存在する。

- 通信遅延が小さいため、超低遅延の用途に利用可能
- エッジ側で処理が行われるため、通信コストが削減可能
- ネットワークを経由させたくない情報を扱う処理が可能

2.1.2 エッジ AI との関連

エッジコンピューティングのうち、エッジ側で行う計算処理が AI による推論である場合を、特にエッジ AI 呼ぶ。AI での推論、とりわけ深層学習の推論をエッジデバイス側で行う場合には、計算量やメモリなどの使用可能なリソースが限定されていることが問題となる。そのため、Google の Edge TPU (Tensor Processing Unit) [4]、NVIDIA の Jetson [5] など、各社がエッジ側での深層学習処理を高速化するためのハードウェア開発を進めている。また推論アルゴリズムの面でも、深層学習や従来の機械学習手法を軽量化しエッジ側での推論に最適化する試みは数多く行われている [6][7][8]。

2.2 AIOps プラットフォーム

2.2.1 概要

AIOps は、人工知能 (AI) と運用 (Ops) を組み合わせた用語であり、IT システムの運用において AI を活用することを意味する。近年、システムの規模や扱われるデータの量が大きくなるとともに、システムを効率よく運用することは困難となってきており、その対策として運用に AI を活用する AIOps が注目を集めている。ガートナーの調査では、AIOps や監視ツールを使用する大企業の割合は、2018 年の 5% から 2023 年には 30% まで上昇すると予測している [9]。

AIOps を実現するための基盤が AIOps プラットフォームであり、データに基づいて可視化、パフォーマンス分析、異常検知などの様々な運用に関する機能を AI により提供する。

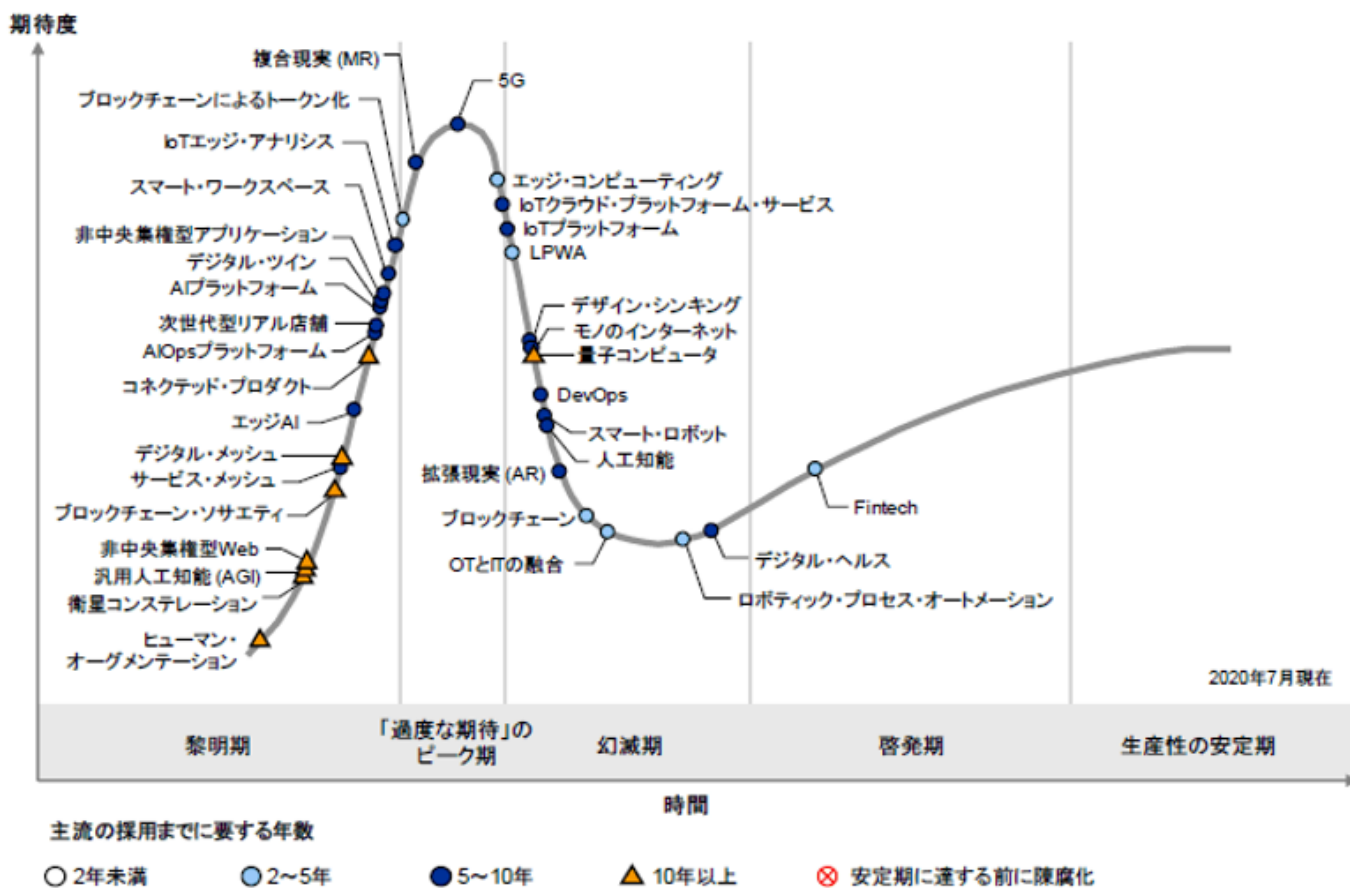


図1 日本における未来志向型インフラ・テクノロジーのハイプ・サイクル：2020年 [1]

2.2.2 エッジ AI との関連

エッジ AI との融合という観点では、データ通信量の節約や処理遅延の削減のために、エッジ AI による処理を組み合わせることが考えられる。例えば IDC の調査では、IoT エッジインフラの利用用途は「AI によるデータ分析」が今後増加し、その最も重要な理由は「データ処理が速い」ことだとされており [10]、AIOps においてもエッジ AI の特性が行かせる場面が存在すると推測される。

2.3 5G

2.3.1 概要

5G は「第 5 世代移動通信システム」という無線通信の規格を指す。従来規格と比較し、「高速・大容量」、「低遅延」、「多数端末接続」という特徴があり、4K/8K 高精細映像や AR/VR を活用した高臨場感のある映像の伝送、自動運転サポートや遠隔医療などを実現し、様々なサービス、産業を革新すると期待されている [11]。

2.3.2 エッジ AI との関連

5G は高速や低遅延などの特性は備えているものの、ネットワークを介する以上は通信の遅延が発生し、またパケットの損失など信頼性の問題は残る。そのため、5G が普及した場合でも、超低遅延や高信頼性といったエッジコンピューティングやエッジ AI が持つ特性が必要となる場面は残ると考えられる。

例えば、エッジコンピューティングの規格の 1 つである MEC (Multi-access Edge Computing) を 5G ネットワークに導入するための検討は数多く行われており [12] [13]、5G とエッジ側処理の共存は重要な課題と認識されている。

2.4 OT と IT の融合

2.4.1 概要

OT は運用・制御技術 (Operational Technology)、IT は情報技術 (Information Technology) を指し、OT と IT の融合は、これらの技術を統合することを指す。Industry-4.0 [14] が提唱されるなど、世界の潮流として製造業等の産業と情報技術との融合は必須の課題である。

OT と IT の融合の重要な要素の一つとしては、産業分野向け IoT である IIoT (Industrial Internet of Things) などが存在すると考えられ、近年大幅に進展している。文献 [15] では、産業用途での IoT デバイス数の年平均成長率を 34.1% と予測しており、他分野と比較し最も高い値としている。図 2 に年平均成長率のグラフを示す。

2.4.2 エッジ AI との関連

産業用途においても AI の活用は伸びてきており、例えば AI 導入が急速に進む中国では、製造業での AI 市場規模が 2025 年までに 20 億ドルを超えると予想されている [16]。図 3 に中国での製造業の AI 市場規模予測のグラフを示す。

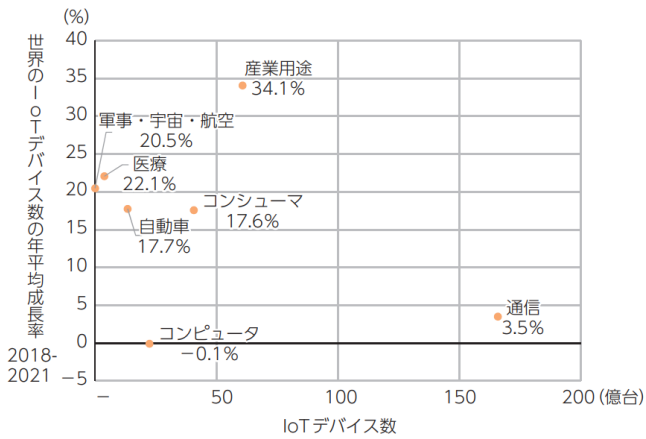


図2 分野・産業別の世界のIoT デバイス数及び成長率予測 [15]

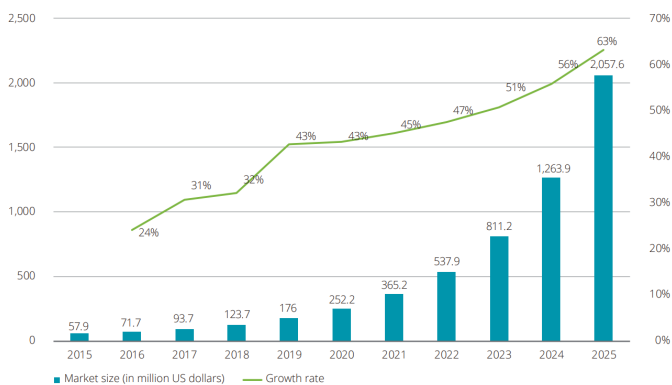


図3 分野・産業別の世界のIoT デバイス数及び成長率予測 [16]

AI の活用が進むにつれ、計算量やメモリ等の使用可能なリソースの問題、通信遅延の問題が生じると予想されるため、エッジ AI 関連技術が必要となる場面も増加していくと考えられる。

3 まとめ

本稿では、「日本における未来志向型インフラ・テクノロジーのハイプ・サイクル：2020年」に登場するエッジ AI 関連インフラテクノロジー用語について解説した。エッジ AI とインフラテクノロジーの関連についての理解を深める一助となれば幸いです。

参考文献

- [1] ガートナー, “日本における未来志向型インフラ・テクノロジーのハイプ・サイクル：2020年”
- [2] Cisco, “Cisco Annual Internet Report (2018 - 2023)”
- [3] Grand View Research, “Edge Computing Market Worth \$43.4 Billion By 2027 | CAGR: 37.4%”, <https://www.grandviewresearch.com/press-release/global-edge-computing-market>
- [4] Google Cloud, “Edge TPU”, <https://cloud.google.com/edge-tpu/>
- [5] NVIDIA, “Buy the Latest Jetson Products”, <https://developer.nvidia.com/buy-jetson>
- [6] Zhang, Xiangyu, et al. “Shufflenet: An extremely efficient

convolutional neural network for mobile devices.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

- [7] Rastegari, Mohammad, et al. “Xnor-net: Imagenet classification using binary convolutional neural networks.” European conference on computer vision. Springer, Cham, 2016.
- [8] Kumar, Ashish, Saurabh Goyal, and Manik Varma. “Resource-efficient machine learning in 2 KB RAM for the internet of things.” International Conference on Machine Learning. 2017.
- [9] Smarter With Gartner, “How to Get Started With AIOps”, <https://www.gartner.com/smarterwithgartner/how-to-get-started-with-aiops/>
- [10] IDC, “国内 IoT インフラ市場インテリジェントエッジ利用状況ユーザー調査結果を発表”, <https://www.idc.com/getdoc.jsp?containerId=prJPJ46186920>
- [11] NTT Docomo, “5G (第 5 世代移動通信システム)”, <https://www.nttdocomo.co.jp/corporate/technology/rd/tech/5g/>
- [12] Kekki, Sami, et al. “MEC in 5G networks.” ETSI white paper 28 (2018): 1-28.
- [13] Pham, Quoc-Viet, et al. “A survey of multi-access edge computing in 5G and beyond: Fundamentals, technology integration, and state-of-the-art.” IEEE Access 8 (2020): 116974-117017.
- [14] Hermann, Mario, Tobias Pentek, and Boris Otto. “Design principles for industrie 4.0 scenarios.” 2016 49th Hawaii international conference on system sciences (HICSS). IEEE, 2016.
- [15] 総務省, “情報通信白書令和元年度版”
- [16] Deloitte, “Deloitte Survey on AI Adoption in Manufacturing”