

通信業界の AI/ML 技術動向

赤間 洋祐 Yosuke Akama, 滝沢 正則 Masanori Takizawa

[要 旨]

6G を含む次世代通信システムにおいて、通信の複雑性の増大や多様化するユースケースに対応するために AI/ML(Artificial Intelligence/Machine Learning)が不可欠な要素と位置付けられている。特に、ネットワークのリアルタイム適応やリソースの効率的な割り当て、通信環境への柔軟な適応などにおいて、その活用が進んでいる。本稿では 5G の無線インタフェース規格である NR(New Radio)エアインタフェースにおける AI/ML 技術の統合を中心に、国際標準化を推進する 3GPP(3rd Generation Partnership Project)における標準化の最新動向を整理し、AI/ML 技術の適用がもたらす可能性と課題を紹介する。また、6G 時代を見据えた技術革新への取り組みについても紹介する。

1 はじめに

次世代通信システム(6G)の商用化が 2030 年頃に見込まれる中、国際電気通信連合(ITU)は「IMT(International Mobile Telecommunications)-2030 フレームワーク(6G ビジョン)」を策定した¹⁾。このフレームワークは、通信の複雑性の増大や多様化する社会的ニーズに対応し、次世代無線通信の可能性を広げる指針を提供している。IMT-2030 は、図 1 に示す主要な利用シナリオと、それを支える設計理念を掲げ、技術の進化を通じて持続可能な社会の実現を目指している。

IMT-2030 では、「持続可能性(Sustainability)」、「セキュリティとレジリエンス(Security and Resilience)」、「ユビキタスインテリジェンス(Ubiquitous Intelligence)」、「デジタルデバイド解消(Connecting the Unconnected)」という 4 つの設計理念を提唱している。これらは、次世代通信が直面する社会的課題を解決し、技術革新と持続可能な価値創出を両立させるための指針である。

さらに、IMT-2030 は、IMT-2020 のシナリオの拡張に加え、以下の新しい利用シナリオを提示している¹⁾。

・人工知能と通信の融合:

AI アプリケーションや分散コンピューティングを支える通信基盤を提供し、自動運転支援や医療支援など多様なユースケースが期待される。また、低遅延、高信頼性、高データレートを備え、AI モデルの分散学習や推論、リソース管理の統合的な支援が期待される。

・センシングと通信の統合:

センシング機能と通信を統合することで、ナビゲーション支援、動作検知、環境モニタリングなどの新たなユースケースが可能となり、高精度な位置特定や効率的なデータ収集・処理が期待される。

・ユビキタス接続性:

デジタル格差を解消し、未接続または接続が不十分な地域にも公平な通信サービスを提供し、さらに他の通信システムと連携して接続性を広げ、すべての人々に通信の恩恵が行き渡る社会が期待される。

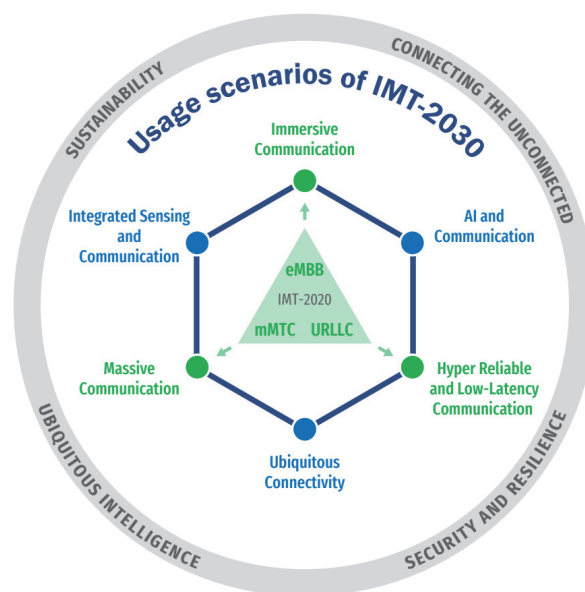


図 1 IMT-2030 の利用シナリオと方針¹⁾

IMT-2030 の設計理念や利用シナリオを実現するには、通信技術の進化に加え、AI/ML などの革新的なデジタル技術の導入が不可欠である。AI/ML は、通信システム全体の効率化、適応性の向上、リソースの最適管理を実現し、IMT-2030 を支える基盤技術として期待されている。その導入の主な動機として以下の点が挙げられている²⁾。

・従来技術の限界と新たなアプローチ:

増大するネットワークの複雑性に対して、従来の通信技術だけでは対応が困難となりつつある。そのため、AI と通信理論の融合による新たなブレークスルーが求められている。

- ・ AI の特性と無線通信の適合:

AI は、学習能力、適応性、およびリアルタイムでの意思決定に優れており、これらの特性は、動的かつ複雑な無線ネットワークが求める要件に適合している。

- ・ 通信システムの最適化と性能向上:

AI 技術により、リソース配分の最適化やネットワーク混雑の予測が可能となり、これにより通信システム全体の性能向上が期待される。

- ・ シームレスで効率的な通信環境の実現:

AI 技術により、途切れることのない効率的な無線接続が実現し、エネルギー効率の向上および通信基盤の高度化が促進されることが期待される。

このような背景を受け、6G に向けた技術革新を進める上で、現在の 5G 技術における AI/ML の適用とその標準化が重要なテーマの一つとなっている。その中で、3GPP では、5G における AI/ML 導入に向けた規格策定を進めている。そこでは、NR エアインタフェースやネットワーク管理、オーケストレーションなど多岐にわたる領域に AI/ML 技術を適用し、通信システムの効率性および柔軟性を向上させるための方策が検討されている。

これらの取り組みや技術動向を踏まえ、本稿では AI/ML 技術に関する最新の動向を整理し、その概要を紹介する。まず、3GPP 標準化における AI/ML の現状を概観し、続いて NR エアインタフェースへの AI/ML 適用に関する動向を中心に説明する。特に、AI/ML 適用の主要なユースケースであるビーム管理やチャネル状態情報 (CSI: Channel State Information) フィードバック強化に焦点を当て、その技術的可能性と利点について紹介する。さらに、6G を見据えた革新的アプローチとして、「セマンティック通信」や「生成 AI」の動向を紹介し、これらが通信業界にもたらす潜在的な影響を俯瞰する。

2 3GPP 標準化における AI/ML の動向概要

5G のさらなる高度化、すなわち 5G-Advanced の実現および将来の 6G を見据え、3GPP をはじめとする標準化団体は、ネットワークおよび無線アクセス技術分野への AI/ML の適用に重点的に取り組んでいる。本章では、3GPP 標準化における AI/ML の動向を概説する。

2.1 5G ネットワークにおける AI/ML の動向

AI/ML は、ネットワーク運用効率の向上や新たなユースケースの実現に向けた重要な技術基盤として位置付けられている。

3GPP の Release 15 以降では、AI/ML を活用したデータ分析機能の導入が進められ、5G コアネットワーク (5GC) を担う NWDAF (Network Data Analytics Function) や、運用・管理・保守 (OAM: Operations, Administration, Maintenance) 領域を担う MDAF (Management Data Analytics Function) といったデータ分析の基盤が整備されてきた。Release 17 では、新たに AnLF (Analytics Logical Function) と MTLF (Model Training Logical Function) の 2 つの論理ネットワーク機能が導入された。MTLF で AI/ML モデルの訓練や再トレーニングを行い、AnLF はその AI/ML モデルを活用してネットワークの推論や予測を行うことにより、高度なデータ分析と予測機能を実現する。さらに、Release 18 では、エンド・ツー・エンドのデータ転送時間分析や移動行動分析などのネットワーク分析機能が拡充され、NWDAF は 5G-Advanced の重要な基盤技術としての役割を担うに至っている。これにより、ネットワークのパフォーマンス向上や新たなユースケースの実現に貢献している^{3)~5)}。

2.2 NG-RAN における AI/ML の動向

3GPP RAN WG2 では、Release 17 から NG-RAN (Next Generation Radio Access Network) に対する AI/ML 技術導入の検討が開始された。Release 18 以降では本格的な適用に向けた仕様策定が進められている。

主なユースケースは以下のとおりである。

- ・ 省電力化 (Energy Saving):

ネットワークデータ分析によるエネルギー効率の最適化を行う。

- ▶ セルのオンオフ制御:

トラフィック需要に応じて、基地局がカバーするセルを動的に有効化または無効化することで、不要なエネルギー消費を削減する。

- ▶ トラフィックオフロード最適化:

トラフィックを効率的に分散し、負荷の集中を緩和することで、システム全体のエネルギー効率を向上させる。

- ・ 負荷分散 (Load Balancing):

5G ネットワークのトラフィック分配の最適化を行う。

- ▶ トラフィック分配効率化:

UE (User Equipment) やネットワークノードから得られるリアルタイムの RAN データを解析し、トラフィックを効率的に分散させる。

- ▶ ネットワーク性能向上:

負荷の集中を緩和し、全体的なネットワーク性能とユーザ体験を改善する。

- ・ モビリティ最適化(Mobility Optimization):

5G ネットワークにおけるモビリティ管理の効率化を行う。

- ハンドオーバー最適化:

RAN データを活用し、UE の位置や動作を予測することで、最適なタイミングと方法で基地局を切り替えるハンドオーバーを実行し、通信の中断や無線リンク障害を最小化する。

- トラフィック最適化とネットワーク効率化:

リアルタイムでトラフィック最適化を行い、混雑を回避しながら適切なルートへ誘導することでネットワーク全体の負荷を軽減し、より効率的なリソース利用と高品質な通信体験を実現する。

さらに、Release 19 以降では、カバレッジおよび容量最適化(CCO: Coverage and Capacity Optimization)やネットワークスライシングといった新たなユースケースへの適用が検討される一方で、これまでに検討されてきたユースケースにおける AI/ML 機能の強化も進められており、NG-RAN における AI/ML の適用範囲は着実に拡大している^{6)~8)}。

2.3 NR エアインタフェースにおける AI/ML の動向

5G の無線インタフェース規格である NR エアインタフェースでは、Release 18 で AI/ML ベースのアプローチが本格的な研究アイテムとして取り上げられた。2023 年 12 月に発行された 3GPP TR 38.843⁹⁾では、3 つのユースケースを通じて、AI/ML の NR エアインタフェース統合による性能向上、複雑性、現行仕様への影響が従来手法と比較して検討されている。加えて、性能評価のための指標(KPI: Key Performance Indicator)の検討や、AI/ML モデルの訓練、展開、推論、監視、更新を管理するライフサイクル管理(LCM: Life Cycle Management)、AI/ML モデルの特長、および関連する用語の定義についても議論されている。この取り組みは、AI/ML フレームワークの構築に向けた基盤を形成することを目的としている。

本節では、この 3GPP TR 38.843 で取り上げられている 3 つのユースケースの概要、AI/ML モデルの配置検討、そして導入する際の重要な課題となる相互運用性確保および性能評価の概要について説明する。なお、本稿で扱う AI/ML モデルとは、通信ネットワークにおけるデータを学習し、最適な判断や予測を行うためのアルゴリズムや手法を指す⁹⁾。

2.3.1 各ユースケースの概要

本項では各ユースケースについて概説する。それぞれの AI/ML モデルの配置位置(例:two-sided モデル)については次項で説明する。

- ・ CSI フィードバック強化:

マルチパス無線チャネル情報である CSI を UE から基地局(gNB)へフィードバックするシステムに対して、AI/MLを活用した最適化を行う。具体的には、CSI データを圧縮する CSI 圧縮モデルと、CSI データを予測する CSI 予測モデルの 2 つの手法がある^{9)~12)}(詳細は 3.2 章を参照)。

- CSI 圧縮モデル:

AI/ML ベースの Auto Encoder(AE)を用いて CSI 圧縮モデルにより、CSI データのフィードバックオーバーヘッドを削減する。UE 側と NW(Network)側の AI/ML モデルが相互連携し、データを圧縮と復元を行うことで、高い復元精度を維持しつつ、効率的に CSI を伝達し、フィードバックデータ量を最小化する(two-sided モデル)。

- CSI 予測モデル:

過去の CSI を AI/ML で学習し、将来の CSI を予測することで、動的なチャネル環境への適応を実現し、通信品質が向上する(UE-sided モデル)。

- ・ ビーム管理:

5G ミリ波システムのビームフォーミングに AI/ML を適用することで、無線環境を解析し、最適なダウンリンク(DL: Downlink)ビームペアを予測・選択できる。これにより、従来の方式よりもオーバーヘッドを削減し、遅延を低減するとともに、ビーム選択精度を向上させる。AI/ML を活用したビーム予測には、空間領域と時間領域の 2 つのアプローチがある^{9)~11), 13), 14)}(詳細は 3.1 章を参照)。

- BM-Case1(空間領域 DLビーム予測):

測定用ビームセット B の測定結果に基づいて予測用ビームセット A の中から最適な DL ビームを選択することで、空間的なビーム探索を効率化する(UE-sided モデルまたは NW-sided モデル)。

- BM-Case2(時間領域 DLビーム予測):

測定用ビームセット B の過去の測定結果に基づいて予測用ビームセット A の中から最適な DL ビームを選択することで、ハンドオーバー遅延やビーム追跡のオーバーヘッドを低減する(UE-sided モデルまたは NW-sided モデル)。

- ・ 位置精度の向上:

見通し外(NLOS: Non-Line-of-Sight)環境を含む多様なシナリオでの測位精度の向上を図る。AI/ML を用いた測位には、直接測位とアシスト測位の 2 つの手法がある^{9)~11), 15)}。

➤ 直接 AI/ML 測位:

位置情報を直接 AI/ML モデルが推定する。これには、パワー遅延プロファイル(PDP: Power Delay Profile)やチャネルインパルス応答(CIR: Channel Impulse Response)など、予め測定したチャネル観測値に基づくフィンガープリンティングベースの測位を含む(UE-sided モデルまたは NW-sided モデル)。

➤ AI/ML アシスト測位:

LOS(見通し内)/NLOS(見通し外)識別情報、測定タイミングや角度といった中間的な測位用測定情報を AI/ML モデルによって提供することで、LOS 環境を前提とした従来の NR 位置測位の精度向上を支援する(UE-sided モデルまたは NW-sided モデル)。

2.3.2 AI/ML モデルの配置

AI/ML の配置位置について、ユースケースの特性や運用条件に応じた 2 つのモデル、one-sided モデルと two-sided モデルの標準化が検討されている^{9)~11)、16)}。

・ one-sided モデル:

AI/ML モデルによる推論が UE 側または NW 側のいずれか一方でのみ行われるモデルである。この場合、推論を UE 側で実行する UE-sided モデルと、NW 側で実行する NW-sided モデルに分類され、以下の検討がされている。

➤ UE-sided モデル:

UE 側でのユースケース実現に必要な新しいシグナリングや AI/ML 対応機能の制御手順の標準化について議論されている。

➤ NW-sided モデル:

NW 側(gNB または、LMF: Location Management Function)のユースケースでは、UE 側からの補助情報の提供など、NR エアインターフェース上で追加のシグナリングが必要となる場合がある。

・ two-sided モデル:

UE 側と NW 側の両方に AI/ML モデルを搭載し、相互連携して推論を行う構成である。このモデルでは、異なる環境で開発・トレーニングされたモデル間の連携において、トレーニングデータやモデル設計情報の共有が課題となる。さらに、マルチベンダー間の互換性と相互運用性を確保する必要があり、これが標準化における大きな課題となっている。

2.3.3 相互運用性とテスト容易性

相互運用性とテスト容易性は、標準化された AI/ML ベースの機

能をセルラーネットワークに導入する際の重要な課題であり、3GPP の RAN4 ワーキンググループでは、AI/ML ベースの機能の性能検証に向けて、推論やデータ収集、汎化性能の確認を含むテストの要件や手順を検討している。特に、AI/ML モデルはデータを学習して動作するデータ駆動型のため、従来の物理法則や数学的基準では評価が難しい。そのため、AI の特性に適した新しい検証方法が求められている。

図 2 は、3GPP RAN4 が提案する AI/ML ベースの機能をテストするための試験装置と被試験装置(DUT: UE または gNB)の参照ブロック図を示している。この図は one-sided モデルおよび two-sided モデルの双方を対象としており、特に two-sided モデルでは、試験装置側にも AI/ML モデルを搭載し、DUT と共同で推論を行うことが想定されている。ただし、試験装置内の AI/ML モデルは、DUT に搭載された AI/ML モデルの性能を評価するためのリファレンスとなるモデルであり、このモデルをどのように設計し、DUT 内の AI/ML モデルの性能を効果的に評価するかについては、現在も議論が続けられている。

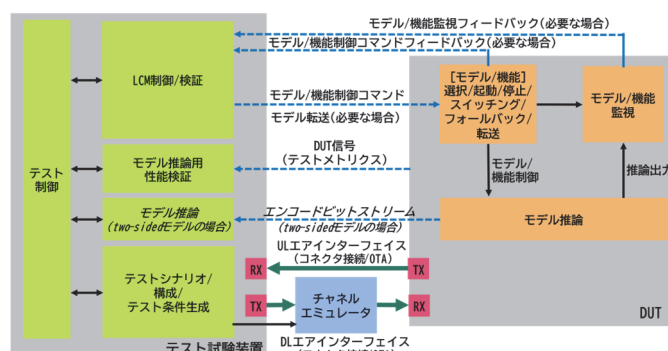


図 2 AI/ML ベースの機能テストの参照ブロック図 9)、10)、17) を参考に編集

さらに、AI/ML ベースの機能のテストにおいては以下のような課題が存在する。

・ データ収集・生成手法:

AI/ML モデルの評価には、テスト用データの収集・生成方法が重要である。例えば、3GPP TR 38.901 に基づくチャネルモデルを用いたデータや、現場計測から得られるフィールドデータ、または RAN4 が定義する仮定やパラメータに基づき試験装置が生成するデータなど、多様なデータソースが考えられる。

・ 汎化検証(Generalization Verification):

汎化とは、トレーニング時には存在しない新たな状況に対しても、モデルが適切に対応できる能力を指す。AI/ML モデルは、標準化された条件下で優れた性能を発揮するように最適化されるため、実際の使用条件下では訓練データへの過剰適合(オーバー

フィット)によって性能が低下する恐れがある。現実のネットワーク条件は実験室環境よりもはるかに多様かつ変動的であり、これをどのようにテストし、要件として組み込むかが大きな課題である。

- 運用中のモデル更新と監視:

AI/ML モデルは、フィールド導入後も性能を維持・改善するため、アップデートが必要となる場合がある。単に出荷時の状態をテストするだけでなく、運用中の性能を監視し、必要に応じて再テストやモデル更新後の再検証を行うことが求められる。このような性能監視手順やモデル管理手順には、機能/モデルの選択・有効化・無効化、切り替え、転送、更新、フォールバック(性能低下時に信頼できる従来モデル・手法へ戻す)といった動的な運用が含まれ、これらに関連する遅延および中断要件を考慮したライフサイクル管理(LCM)としての要件定義が求められている。この LCM 関連のテスト設計においては、フィールドに展開された後でも機能・モデルの更新や切り替えが発生し得ることを想定し、それに対応できるフレームワークの検討が求められている^{9), 10), 17)}。

2.3.4 Release19 以降の動向

Release 19 では、これまでの成果を基に以下の仕様策定および研究が標準化に向けて現在進行している^{11), 18)}。

[主な仕様策定アイテム]

- AI/ML 全般フレームワーク(one-sided モデル):

モデルのライフサイクル管理(LCM)を中心に、モデルの選択、有効化・無効化、切り替え、フォールバックを支援するシグナリングおよびプロトコルの設計。これには、トレーニング、推論、性能監視、データ収集のためのシグナリングや仕組みの設計が含まれる。

- ユースケース別 one-sided モデルの仕様策定:

- ビーム管理:

ダウンリンク送信ビームの空間予測(BM-Case1)および時間予測(BM-Case2)に対応するモデルの仕様策定。また、UE 側での推論における汎化性能を向上させるため、学習時と推論時のネットワーク条件の違いによる影響を最小限に抑える手法や、環境変化に適応する仕組みについての検討。

- 測位精度向上:

UE 側または NW 側モデルを活用した直接的または補助的な AI/ML 測位に関連する仕様策定。これには、必要な測定項目や性能監視の仕組みも含む。また、UE 側での推論における汎化性能向上の検討。

- CSI フィードバック強化(CSI 予測:UE-sided モデル):

CSI 予測に必要なモデル管理や推論における汎化性能向上の検討。

[主な研究アイテム]

- CSI フィードバック強化の CSI 圧縮(two-sided モデル):

以下のポイントに対する更なる研究を継続する。

- 性能と複雑度/オーバーヘッドのトレードオフを改善するための手法の検討(例:空間/周波数圧縮を空間/時間/周波数圧縮へ拡張するアプローチ, CSI 圧縮と予測の統合手法, Rel-18 非 AI/ML アプローチとの比較分析など)。

- ベンダ間トレーニング協調に関する課題(例:データ共有の方法, モデル互換性の確保)を軽減または解決するための方策。

- テスト容易性および相互運用性:

- one-sided モデル:

テストフレームワークと手順の確立。

- two-sided モデル:

以下の観点を含むさまざまなテストオプションの分析。

- 既存要求事項との関係

- 性能監視およびライフサイクル管理(LCM)側面

- 汎化性能の課題

- 静的/動的シナリオおよび条件, さらには CDL(Clustered Delay Line)モデルやフィールドデータなどの伝搬条件

- UE の処理能力および制約

- モデル変更/ドリフトに伴う展開後の検証

3 ビーム管理と CSI フィードバック強化

本章では検討されている 3 つのユースケースのうち、ビーム管理と CSI フィードバック強化の 2 つにおいて AI/ML 技術がどのように既存の課題を解決し、通信品質/効率を改善するかについて説明する。

3.1 ビーム管理

本節ではビーム管理における AI/ML 適用に向けた研究動向の詳細について紹介する。ミリ波を用いた通信では一般的に指向性の高いビームを用いるが、移動局との通信ではビームマネジメント(BM)に関連するシグナリングのオーバーヘッドと遅延が重大なボトルネックとなる。3GPP NR では、次の基本的な BM の 4 つの手順を指定している。ビームスイープ、ビーム測定、ビーム決定、ビーム報告、これら 4 つの手順を周期的に実行し、時間の経過とともに変化する送受信ビームの最適なペアを継続的に更新していく。これら BM の手順は Release16, 17 でさらに強化されたが、BM に由来

するオーバーヘッドと通信遅延、スペクトル利用効率の低下には依然として一定の課題がある。これらの問題に対処するため、AI/MLを適用した BM では特に空間的ビーム予測と時間的ビーム予測という面で注目されている。

(1) SBP(空間領域 DL ビーム予測)

ダウンリンク通信において、ビームセット A を直接すべて測定・走査する代わりに、より少数かつ測定の容易なビームセット B を測定し、その測定結果を基に AI/ML モデルでビームセット A 内の最適ビームを予測する。これにより、システムのオーバーヘッドや遅延を低減しつつ、十分な精度でビーム選択を行うことが可能になる。AI/ML モデルの学習および推論は、gNB 側で行うか、UE 側で行うかを柔軟に選択できる。どちらにおいても、ビーム予測の基本的な考え方は同じだが、システムや端末の状況に応じて利点や実装の仕方が異なる。

・ gNB による SBP:

gNB 側に配置された AI/ML モデルに、UE が測定したビームセット B の L1-RSRP(物理層リファレンスシグナル受信電力)を入力として与え、モデルがビームセット A の中で最適なビームを推定する。推定結果としては、送信ビーム ID や受信ビーム ID などの「ビーム識別子」が好ましい。ここで、図 3 に示すとおり、A と B のビームセットの関係に 2 通りの選択肢がある。B が A の部分集合の場合と、B はワイドビーム(広いビーム)で、A はナロービームの場合である。B が A の部分集合の場合は B として限られた数のビームだけを走査・測定すればよいので、走査の回数や報告のオーバーヘッド、遅延が削減できる。レガシー方式よりも少ないビーム測定で済む点がメリット。ビームセット B はワイドビームで構成され、ビームセット A はナロービームで構成された場合は、ワイドビームを測定することで、大まかな方向性が得られる。そこから対応する 1 本または複数のナロービーム(狭いビーム)の性能を推定し、直接ナロービームを測定しなくても済む。これにより、測定オーバーヘッドや遅延をさらに削減可能。

・ UE による SBP:

UE 側に配置された AI/ML モデルに、UE 自身がビームセット B を測定した結果を入力とし、ビームセット A の中で最適ビームや推定した RSRP 値を、必要に応じて gNB へ報告する。UE 側で推論を行うため、gNB に送る情報をより豊富にして(例: 予測された RSRP 値、ビームインデックスなど)、システム全体の動的なビーム管理精度を高めることができる。ビームセット A とビームセット B の関係は gNB 側で行う場合と同じである。

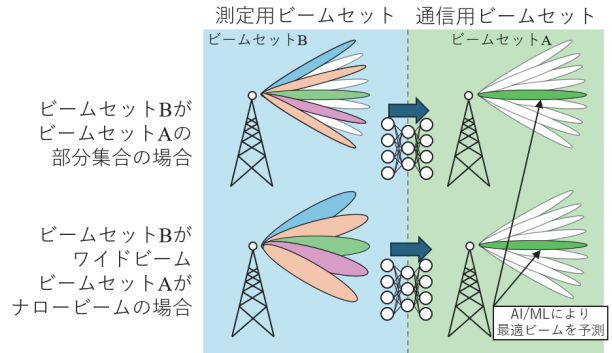


図 3 BM における空間領域 DL ビーム予測

(2) TBP(時間領域 DL ビーム予測)

図 4 に示すようにビームセット A のの中から最も適切な将来のダウンリンクビームを、ビームセット B から得られる過去の測定値(例: L1-RSRP など)を用いて予測する。SBP 同様、AI/ML モデルの学習および推論は gNB 側でも UE 側でも行うことができ、ネットワーク構成やシステム要件に合わせた柔軟な配置が可能。

・ gNB による TBP:

gNB 側にある AI/ML モデルが、ビームセット B の過去 x 周期連続した L1-RSRP 測定値などをもとに、今後の y 周期におけるビームセット A 内の最適ビームを予測する。このプロセスは、測定と予測を繰り返す一連のサイクル($x + y$ サイクル)として運用される。将来のビームを予測して gNB があらかじめビームスイッチや変調符号化方式(MCS)の調整を行うことでリンクの途切れや再送を回避しやすくなる。SBP と同様、すべてのビームを都度走査する必要がなくなるため、ビーム管理の効率化が図れる。

・ UE による TBP:

UE 側に配置された AI/ML モデルが、ビームセット B の過去の測定値(L1-RSRP など)をもとに、将来の複数のタイミングにおけるビームセット A 内のビーム品質を予測する。必要に応じて、予測値(ビームインデックスや予測された RSRP など)を gNB に報告する。基本的なメリットは gNB による TBP と同様であり、事前の適切なビーム選択やリンク適応が可能になる。SBP と同様、gNB からのアシスタンス情報(ビームの最大利得方向、3 dB ビーム幅、アンテナ設定、コードブック ID、端末位置情報など)を活用すれば、UE はより正確な予測を行える。TBP は、SBP が空間的予測をとおりてビーム選択の効率化を目指したのと同様に、時間的な側面を考慮してさらなる効率化を可能にする手法だと言える。

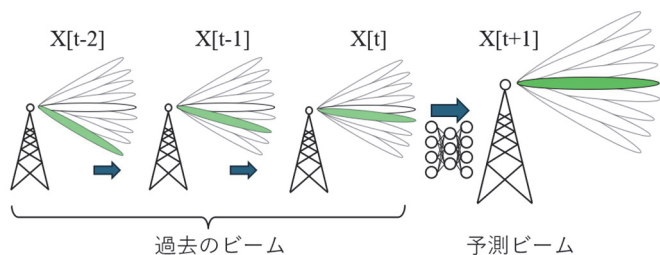


図4 BMにおける時間領域DLビーム予測

3.2 CSI フィードバック強化

本節ではCSIフィードバック強化におけるAI/ML適用に向けた研究動向の詳細について紹介する。CSIフィードバックの強化では、two-sidedモデルによるCSIデータ圧縮とone-sidedモデルによるCSI予測についてAI/ML適用によるパフォーマンスの改善が研究されている。

(1) CSI 圧縮

CSIデータの量は、送信アンテナ素子数、受信アンテナ素子数、およびサブキャリア数の積に比例して増加する。そのため、Massive-MIMOの運用を前提とする5G-Advancedでは、これまでよりもCSIデータの量が増加する。CSIデータは通常PMI(Precoding Matrix)に変換される過程で共有コードブックにより圧縮するが、既存のコードブックによるCSIデータの圧縮では、圧縮率と復元精度のトレードオフの関係により性能が決まっていた。一方で、図5に示す生成AIの一種であるAE(Auto Encoder)を用いたCSIデータの圧縮では、CSIデータの構造に潜在する特長を利用して、UEのエンコーダネットワークにより低次元の特徴空間にCSIデータを圧縮する。特徴空間に圧縮されたCSIデータは、デコーダネットワークにより元のCSIデータに復元される。このユースケースではtwo-sidedモデルであり、前章で述べた互換性・相互運用性の課題が生じる。

AEによるCSI圧縮での近年の関心事項は、主に下記の3つである。

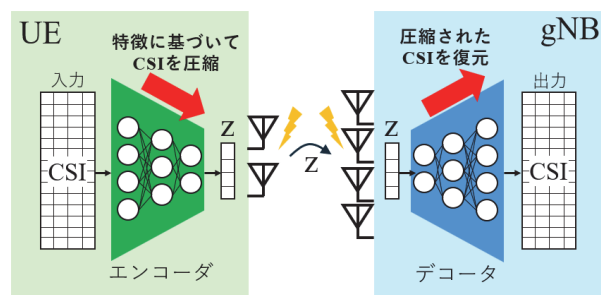
- ・モデルの複雑さと復元精度
- ・モデルアーキテクチャ
- ・モデルのトレーニング

AEにおいてはモデルの複雑度と復元精度の観点で議論がされている。AEはモデルサイズが大きくなるほど復元精度と高い圧縮率が期待できるが、限られたUEのリソースを圧迫するため、モデルサイズとパフォーマンスのトレードオフを考慮して最適化が必要がある。また、複雑度の観点で、モデルのアーキテクチャについても議論がある。CSI圧縮にAEの有効性が示された当初は、AEの

ニューラルネットワーク(NN)にはCNN(畳み込みNN)などの伝統的なネットワークが用いられていたが、近年ではTransformerなどのアーキテクチャが有力視されている。Transformerとは、2018年にGoogleが提唱したNNアーキテクチャで、元は自然言語の翻訳等に向けて開発されたモデルである。Transformerを用いたAEではCNNを用いたAEよりも高い性能が期待できるが、同規模のCNNと比較して計算が複雑であるために、NNの挙動に対する説明可能性に乏しいなどのデメリットも抱えている。また、結果的に計算が複雑になってしまい、CSI圧縮率と精度に対するコストメリットが希薄になるなどの問題もある。これらの懸念事項を踏まえて、CSI予測と統合することで得られる新たなメリットについて議論されている。

(2) CSI 予測

CSIフィードバックにおいて、CSIがUEで取得されてからgNBで使用されるまでには遅延がある。この間にUEの移動や電波伝搬環境の変化があると、使用されるCSIデータはその時点のCSIと乖離する。これはチャネルエージングと呼ばれ、未解決の課題とされていた。その解決策の一つとして、過去に測定されたCSI情報を基に、AI/MLを用いて現在のCSIを推定する「CSI予測」が検討されている。適用が検討されているアーキテクチャはLSTM(Long Short-Term Memory)と呼ばれるネットワークで、時系列データの学習に広く用いられるモデルである。



エンコーダ、デコーダのネットワークにはCNN等が用いられるが、近年の研究ではTransformerが用いられるようになった

図5 Auto EncoderによるCSI圧縮

4 6Gに向けた新たな取り組み

本章では6G通信の実現に向けた全く新しい取り組みの一部を紹介する。

4.1 セマンティック通信

セマンティック通信とは伝送するデータの“ビットの正確性”ではなく、その“意味”や“情報の本質”を正確に伝えることを重視する通信方式である。シャノンの通信理論ではビット誤り率が最小化される

ことを主要な指標とするが、セマンティック通信では受信者側が必要とする“セマンティクス(意味情報)”が正しく伝わるかどうかに焦点を当てる。

通信インフラの発展と共にデータ量が爆発的に増える一方で、すべての生データ(ビット列)を高い忠実度で送信する必要が本当にあるのかについて見直されている。たとえば、IoT センサーが送信する大量の時系列データも、実際には特徴量や傾向のみが重要なケースが多くある。そこで、「意味情報」に着目して不要なデータを削減し、ネットワークリソースを効率的に活用する技術としてセマンティック通信が注目を集めている。例えば近年の研究では、自然言語、画像、音声などの各種メディアにおいて、深層学習(Transformer, CNN, RNNなど)を用いたセマンティックエンコーダ/デコーダが提案されている。さらに画像解析や音声認識、推論などを端末とクラウドのどちらで実行するかという選択も含め、どのタイミングでどのデータを送信するかを最適化する研究が進んでいる。対象タスクの“目的”を達成するために必要な情報のみ送るという考え方である。6G ではAR/VRなど複合現実(XR)サービスが主流になると想定されるため、音声や動画、テキスト、センサー情報等の異なるモードの“意味”を統合的に圧縮・伝送する枠組みも検討されている。

4.2 生成 AI の通信業界への活用について

近年、AI/ML 技術はさまざまな産業分野に進展しているが、通信業界においても例外ではない。

- ・物理層(PHY)・MAC 層設計への応用:

環境や周波数帯、アンテナ構成によって電波の伝搬特性は大きく変化するため、物理層におけるチャネル推定・等化・ビームフォーミングなどが重要である。スコアベース生成モデル(Score-based Generative Models: SGMs)や拡散モデルを用いることで、ノイズを含むチャネル測定データからのチャネルの復元²⁰⁾や、MIMO 設計を最適化する研究が進んでいる^{19), 21)}。また、Large Beamforming Models(LBM)という形でビームフォーミングや RIS(再構成可能な反射板)制御を大規模モデルで扱う試みも見られる²²⁾。

- ・無線資源割り当て・最適化:

Radio Resource Allocation は 5G/6G の要となるテーマであり、これまでは強化学習や数値最適化を用いた手法が主流であった。近年は、拡散モデルを活用した無線資源割り当て²³⁾、あるいは LLM を活用した Mixture of Experts(MoE)による動的ネットワーク最適化²⁴⁾が提案されており、ネットワークの複雑な環境変

化に対して高速かつ高精度な意思決定を実現しようとする動向が伺える。

- ・ネットワーク管理・運用・オーケストレーション:

Intent-Based Networking(IBN)やネットワークスライシング管理において、LLM の自然言語理解力を使って運用者の要求(意図)を動的に翻訳・実行するシステムが提案されている^{25), 26)}。大規模言語モデルにテレコム専用のドメイン知識を学習させる「TelecomGPT」²⁷⁾や「WirelessLLM」²⁸⁾といったコンセプトも登場し、ネットワーク運用プロセスの自動化・効率化が期待されている。さらに自律分散型オーケストレーション²⁹⁾など、クラウドとエッジの協調制御で大規模モデルを適切に分割・配置し、推論や学習を分散実行するアーキテクチャの検討も進んでいる。

- ・エッジインテリジェンスと大規模モデル:

通信と計算の統合により、エッジデバイスやエッジサーバでの推論が高性能化してきたことから、エッジでの LLM 推論が現実的になりつつある^{30), 31)}。エッジ環境では限られたリソース(計算能力、メモリ、通信帯域)が前提となるため、量子化・蒸留・MoEなどを駆使してモデルを軽量化する研究が行われている^{32), 33)}。また、クラウドとエッジの分散推論に加え、フェデレーテッドラーニング(連合学習)などの手法と組み合わせることで、プライバシーを保ちつつ大規模モデルの性能を活用しようとする動きも見られる^{34), 35)}。

5 まとめ

本稿では、次世代移動通信における AI/ML 技術について、3GPP での検討状況を中心に、その中でも特に NR エアインタフェースにおける AI/ML 適用の動向と課題を取り上げた。また、6G に向けた新たな取り組みについても簡単に紹介した。

NR エアインタフェースにおける AI/ML の適用は、複雑化する環境下での通信性能向上と効率化に欠かせない。しかし、その実現にはモデル管理、性能評価、相互運用性、汎化性能向上といった課題があり、AI/ML を持続可能かつ効果的に導入するための重要なテーマとして 3GPP で活発に議論されている。これらの検討の進展は、AI を前提とした次世代通信システムのさらなる発展を可能にするものと期待される。

参考文献

- 1) ITU, Recommendation ITU-R M.2160-0 : “Framework and overall objectives of the future development of IMT for 2030 and beyond,” 2023
- 2) Next G Alliance, “Next G Alliance Report: 6G Radio Technology Part II: Basic Radio Technologies,” 2024
- 3) Xingqin Lin, “Artificial Intelligence in 3GPP 5G-Advanced: A Survey,” arXiv: 2305.05092, 2023.
- 4) Bahador Bakhshi, Malla Reddy Sama, Riccardo Guerzoni, 已之口 淳, “3GPP Release 18 におけるネットワーク自動化および AI/ML の高度化技術,” NTT 技術ジャーナル 2024 年 12 月号 vol.36, 2024
- 5) 3GPP, TS 28.288 V18.8.0 (2024-12), “Architecture enhancements for 5G System (5GS) to support network data analytics services (Release 18)”
- 6) Xingqin Lin, Lopamudra Kundu, Chris Dick, Soma Velayutham, “Embracing AI in 5G-Advanced Towards 6G: A Joint 3GPP and O-RAN Perspective,” arXiv: 2209.04987, 2022.
- 7) 3GPP, TR 37.817 V17.0.0 (2022-04), “Evolved Universal Terrestrial Radio Access (E-UTRA) and NR: Study on enhancement for Data Collection for NR and EN-DC (Release 17)”
- 8) 3GPP, RP-233114 (2023-12), “New SID: Study on enhancements for Artificial Intelligence (AI)/Machine Learning (ML) for NG-RAN,” China Unicom.
- 9) 3GPP, TR 38.843 V18.0.0 (2023-12), “Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR air interface (Release 18)”
- 10) Xingqin Lin, “An Overview of the 3GPP Study on Artificial Intelligence for 5G New Radio,” arXiv: 2308.05315, 2023.
- 11) 5G Americas, “A 5G Americas White Paper : Artificial Intelligence in Cellular Networks,” 2024
- 12) Jiajia Guo, Chao-Kai Wen, Shi Jin, Xiao Li, “AI for CSI Feedback Enhancement in 5G-Advanced,” arXiv: 2206.15132, 2022.
- 13) Ke Ma, Zhaocheng Wang, Wenqiang Tian, Sheng Chen, Lajos Hanzo, “Deep Learning for Beam-Management: State-of-the-Art, Opportunities and Challenges,” arXiv: 2111.11177, 2021.
- 14) Qing Xue, Jiajia Guo, Binggui Zhou, Yongjun Xu, Zhidu Li, Shaodan Ma, “AI/ML for Beam Management in 5G-Advanced: A Standardization Perspective,” arXiv: 2309.10575, 2023.
- 15) Mohammad Alawieh, Georgios Kontes, “5G Positioning Advancements with AI/ML,” arXiv: 2401.02427, 2023.
- 16) 5G Americas, “A 5G Americas White Paper : 3GPP Technology Trends,” 2024
- 17) 3GPP, R4-2320611 (2023-11), “Interoperability and testability aspect of AI/ML for NR air interface,” Samsung.
- 18) 3GPP, RP-242399 (2024-09), “New SID: Revised WID on Artificial Intelligence (AI)/Machine Learning (ML) for NR Air Interface,” Qualcomm.
- 19) Z. Wang et al., Generative AI Agent for Next-Generation MIMO Design: Fundamentals, Challenges, and Vision, arXiv preprint arXiv:2404.08878, 2024.
- 20) T. Wu et al., CDDM: Channel denoising diffusion models for wireless semantic communications, arXiv:2309.08895, 2023.
- 21) L. Zhang et al., Spatial Channel State Information Prediction With Generative AI: Toward Holographic Communication and Digital Radio Twin, IEEE Network, vol. 38, no. 5, 2024.
- 22) T. Zhang et al., Large Beamforming Models for Robust Wireless Transmission in Generalized Scenes of RIS-aided Intelligent IoV Network, IEEE Transactions on Vehicular Technology, 2024.
- 23) T. Liu et al., Generative Diffusion Model (GDM) for Optimization of Wi-Fi Networks, arXiv preprint arXiv:2404.15684, 2024.
- 24) H. Du et al., Mixture of Experts for Network Optimization: A Large Language Model-enabled Approach, arXiv preprint arXiv:2402.09756, 2024.
- 25) Dandoush A et al., Large Language Models meet Network Slicing Management and Orchestration, arXiv preprint arXiv:2403.13721, 2024.
- 26) M. Ameer et al., Leveraging LLMs to eXplain DRL Decisions for Transparent 6G Network Slicing, 2024 IEEE 10th International Conference on Network Softwarization (NetSoft).
- 27) H. Zou et al., TelecomGPT: A framework to build telecom-specific large language models, arXiv:2407.09424, 2024.
- 28) J. Shao et al., WirelessLLM: Empowering large language models towards wireless intelligence, arXiv:2405.17053, 2024.
- 29) J. Wang et al., Toward Scalable Generative AI via Mixture of Experts in Mobile Edge Networks, arXiv preprint arXiv:2402.06942, 2024.
- 30) G. Qu et al., Mobile Edge Intelligence for Large Language Models: A Contemporary Survey, arXiv:2407.18921, 2024.
- 31) Z. Lin et al., Pushing large language models to the 6G edge: Vision, challenges, and opportunities, arXiv:2309.16739, 2023.
- 32) A. Chavan et al., Faster and Lighter LLMs: A Survey on Current Challenges and Way Forward, arXiv preprint arXiv:2402.01799, 2024.
- 33) G Liu et al., Fusion of Mixture of Experts and Generative Artificial Intelligence in Mobile Edge Metaverse, arXiv preprint arXiv:2404.03321, 2024.
- 34) F. Jiang et al., Personalized Wireless Federated Learning for Large Language Models, arXiv preprint arXiv:2404.13238, 2024.

35) X. Chen et al., The role of federated learning in a wireless world with foundation models, IEEE Wireless Communications, vol. 31, no. 3, 2024.

執筆者



赤間 洋祐
先端技術研究所
第3研究室



滝沢 正則
先端技術研究所
技術企画部