3GPP TSG RAN Meeting #102
Edinburgh, Scotland, December 11-15, 2023

RP-233469

Scope on AI/ML air interface

Agenda: 9.1.1.1
Source: Samsung
Document for: Discussion and decision

SAMSUNG

Status of Rel-18 study | RAN1 (1/2)

RAN1 has not recommended AI/ML-based CSI feedback enhancement for normative work due to trade-off between performance gain and complexity/overhead, and issues related to inter-vendor training collaboration (for two sided-model).

CSI compression

Capture the following as a conclusion in section 8 of the TR

- From RAN1 perspective, there is no consensus on the recommendation of CSI compression for normative work.
- At least the following aspects are the reasons for the lack of RAN1 consensus on the recommendation of CSI compression for normative work.
  - Trade-off between performance and complexity/overhead
  - Issues related to inter-vendor training collaboration
- Other aspects that require further study/conclusion are captured in the summary

CSI prediction

Capture the following conclusion in section 8 of the TR 38.843

- From RAN1 perspective, there is no consensus on the recommendation of CSI prediction for normative work.
- The reason for the lack of RAN1 consensus on the recommendation of CSI prediction for normative work is due to
  - Lack of results on the performance gain over non-AI/ML based approach and associated complexity
- Other aspects that require further study/conclusion are captured in the summary.

SAMSUNG

Status of Rel-18 study | RAN1 (2/2)

RAN1 has recommended AI/ML-based beam management and positioning for normative work

Beam management

For AI-based beam management, from RAN1 perspective, at least the following are recommended for normative work:

- Both BM-Case1 and BM-Case2:
  - BM-Case1: Spatial-domain DL Tx beam prediction for Set A of beams based on measurement results of Set B of beams
  - BM-Case2: Temporal DL Tx beam prediction for Set A of beams based on the historic measurement results of Set B of beams
- DL Tx beam prediction for both UE-sided model and NW-sided model
- Necessary signalling/mechanism(s) to facilitate data collection, model inference, and performance monitoring for both UE-sided model and NW-sided model
- Signalling/mechanism(s) to facilitate necessary LCM operations via 3GPP signalling for UE-sided model

Positioning

It is recommended to specify necessary measurement, signalling and procedure to facilitate training, inference, monitoring and/or other LCM operations for both direct AI/ML positioning and AI/ML assisted positioning, specifically:

- specify necessary signalling of data collection; investigate the necessity of other information for supporting data collection, and if needed, specify during normative work
- investigate on the necessity and signalling details of measurement enhancements, and if needed, specify during normative work
- investigate on the necessity and signalling details of monitoring method(s), and if needed, specify during normative work

A variety of enhancements for measurements (e.g., based on extensions to current positioning measurements or with new measurements) were also identified as potentially beneficial (e.g., trade-off positioning accuracy requirement and signalling overhead) and are recommended to be investigated further and if needed, specified during normative work.

SAMSUNG

Status of Rel-18 study | RAN2 (1/2)

RAN2 studied various solution for operational aspects but not clear outcomes and left to normative phase

Model Identification and meta information

- RAN2 made the following note for the normative phase:
  - Note: Details on the relationship between model IDs and meta information for purposes of model control and management can be addressed during a normative phase.

Model transfer/delivery

- RAN2 studied several solutions for model transfer/delivery and provided their pros and cons in TR 38.843.
  - Solution 1a: gNB can transfer/deliver AI/ML model(s) to UE via RRC signalling
  - Solution 2a: Core Network (except LMF) can transfer/deliver AI/ML model(s) to UE via NAS signalling.
  - Solution 3a: LMF can transfer/deliver AI/ML model(s) to UE via LPP signalling.
  - Solution 1b: gNB can transfer/deliver AI/ML model(s) to UE via UP data.
  - Solution 2b: Core Network (except LMF) can transfer/deliver AI/ML model(s) to UE via User Plane (UP) data.
  - Solution 3b: LMF can transfer/deliver AI/ML model(s) to UE via UP data.
  - Solution 4a: OTT server can transfer/deliver AI/ML model(s) to UE (e.g., transparent to 3GPP).
  - Solution 4b: OAM can transfer/deliver AI/ML model(s) to UE.

RAN2 understanding is: Whether there is a need to consider standardised solutions for transferring / delivering AI/ML model(s) is unclear from the outcome of the present study.
RAN2 can consider a reactive and a proactive approach for initiating a model transfer/delivery in a normative phase.

SAMSUNG

Status of Rel-18 study | RAN2 (2/2)

UE capability Reporting:

- According to TR 38.843:
  - Further discussions concerning UE capability details (e.g., granularity of Feature/FG, content, structure of the related UE capabilities, etc.) can be carried during a normative phase.

Reporting applicability-related information:

- According to TR 38.843:
  - Note: How and whether there is a need to enable UEs to report applicability-related information can be further discussed and defined in a normative phase.
  - Note: Whether necessary signalling from network is needed for proactive UE reporting can be discussed in a normative phase.
  - Note: Whether there is a need for the network to report to the UE applicability-related information of AI/ML models and/or AI/ML functionalities can be discussed in a normative phase.

Data collection:

- RAN2 did not study and analyse any solutions for data collection for UE-side model training in Rel-18 study, hence, no recommendation for normative phase.

SAMSUNG

Status of Rel-18 study | RAN4

RAN4 studied on requirements and testing frameworks to enable AI/ML for NR air interface, including common framework for AI/ML testing, and use-case specific issues, e.g., testing metrics, for three use cases.

For testability and interoperability for two-sided model, RAN4 has studied on options to provide test encoder/decoder, but the feasibility of any of the testing options has not concluded and more study is required.

Common Testing Framework

Two testing goals are identified for AI/ML, and to be selected based on particular test:

(1) to verify whether a specific AI/ML model/functionality can be conducted in a proper way;

(2) to verify whether the minimum performance gain of AI/ML model /functionality/feature can be achieved for a static scenario/configuration.

For the definition of AI/ML requirements,

- principles are provided for cases with or without existing legacy performance, and
- potential requirements (including latency/interruption requirements) can be defined for performance monitoring procedure, functionality/model management procedure.

Reference block diagrams for testing are provided for 1-sided and 2-sided model, respectively.

Candidate methods for data collection/generation for testing are provided.

The necessity and feasibility to verify the generalization of AI/ML are studied, and the baseline of handling of generalization tests are provided.

Testability and Interoperability for 2-Sided Model

Test encoder/decoder for 2-sided model:

- Four options of test encoder/decoder, based on the different sources of test encoder/decoders and training data.

RAN4 compared four options (taking test decoder for CSI compression use case as example), from various aspects including:

- the necessity and feasibility of test decoder verification procedure
- TE requirement to deploy test decoder (e.g., training, complexity and interoperability)
- Specification effort (defining test decoder and requirements)
- Confidentiality/ IP issues in the testing procedure
- Complexity of testing for the ecosystem, etc.

However, RAN4 concluded that the feasibility of any of the testing options has not concluded and more study is required.

SAMSUNG

Way forward for Rel-19 AI/ML air interface

- **We observed from Rel-18 study, “two-sided model” has huge industry impact and is not ready for normative phase mainly due to**
  - Model development feasibility issues (e.g., inter-vendor training collaboration)
  - Not yet concluded feasibility for any of the options to provide test decoder
  - Not yet clarified test ecosystem impact, including deployment complexity, Confidentiality/IP issues in the testing procedure, etc.

- **In Rel-19 WI, the scope should focus only on one-sided model.**

- **However, if there is great interest for two-sided model from the group, instead of continuing study inside the WI, we strongly suggest new and separate SI for two-sided model for the whole Rel-19.**

SAMSUNG

Proposal for Rel-19 WI

Objectives for Rel-19 WI for AI/ML air interface

- For Beam management use case, support DL Tx beam prediction in BM-Case 1 and BM-Case 2 for both UE-sided model and NW-sided model
  - Specify necessary signalling/mechanism(s) to facilitate data collection, model inference, and performance monitoring for both UE-sided model and NW-sided model [RAN 1, RAN 2]
  - Specify necessary signalling/mechanism(s) to facilitate necessary LCM operations via 3GPP signalling for UE-sided model [RAN1, RAN 2]
  - Note: Strive for a common framework design to support both BM-Case 1 and BM-Case 2

- For Positioning use case, all 5 cases (i.e., case 1,2a,2b,3a,3b) are supported
  - Specify necessary measurement, signaling and procedure to facilitate training, inference, monitoring and/or other LCM operations for both direct AI/ML positioning and AI/ML assisted positioning, additionally: [RAN 1, RAN 2, RAN 3]
    - Specify the necessary signaling of data collection and the other necessary information (if any) for supporting data collection
    - Investigate and specify the necessary signaling of necessary measurement enhancements (if any)
    - Investigate and specify the necessary signaling (if any) of monitoring method(s)

- Specify corresponding RAN4 core requirements for above beam management and positioning use cases: [RAN4]
  - For beam management use case:
    - RRM core requirement for necessary signaling/mechanism(s) for data collection, model inference, and performance monitoring;
    - RRM core requirement for necessary LCM operations via 3GPP signaling for UE-sided model.
  - For positioning use case:
    - RRM core requirement for data collection, measurement enhancements and monitoring methods (if any)

- Note1: Only one side model without model transfer is supported
- Note2: Data collection mechanisms for network-side model.

SAMSUNG

Proposal of scope on new Rel-19 SI

Objectives for Rel-19 SI for AI/ML air interface for two-sided model [RAN1/2/4]

- Further study AI/ML-based CSI feedback enhancement on the basis of the outcome of Rel-18 study focusing on [RAN1]
  - Methods to improve trade-off between performance and complexity/overhead including but not limited to new use cases, e.g., joint CSI compression and prediction, and model input/output types, e.g., angular-delay domain.
  - Methods to resolve issues identified in Rel-18 SI for two-sided model training collaboration and ensure inter-operability without requiring offline inter-vendor collaboration, e.g., Type 3 NW-first OTA dataset sharing.
  - Other aspects that require further study/conclusion as per the outcome of Rel-18 SI.
- Further study the testability and interoperability issues in the two-sided model testing, based on the identified aspects for AI/ML-based CSI compression use case in Rel-18 SI [RAN 4]

Note: LCM for two-sided model (e.g., data collection) should be based on the requirements identified from Rel-19 RAN1 study.

SAMSUNG