Qualcomm

3GPP TSG RAN meeting #93e
Electronic meeting, Sept. 13-17, 2021
A.I. 9.0.2

RP-212046

Views on AI/ML for Air Interface for Rel-18

Introduction

- Rel-18 workshop (RWS-210659) identified
  - **AI (Artificial Intelligence)/ML (Machine Learning)**, with the following example areas:
    - Air interface (e.g., Use cases to focus, KPIs and Evaluation methodology, network and UE involvement, etc.)
    - NG-RAN
- Accordingly, in this document, we propose:
  - Project structure for AI/ML related projects in Rel-18
  - Use cases to focus for AI/ML for Air-Interface SI
  - KPIs and Evaluation Methodology for AI/ML for Air-Interface SI
  - Nomenclature for network and UE involvement for AI/ML for Air-Interface SI

Project structure for AI/ML related projects in Rel-18

- We propose to have the following study for Rel-18 RAN AI/ML (#2 below).

	Temporary title	SI/WI	Primary WG	Secondary WGs	Notes
1	AI/ML for NGRAN	WI	RAN3	RAN2	Based on the outcome of RAN3-led Rel-17 SI *FS_NR_ENDC_data_collect*
2	**AI/ML for Air-Interface**	**SI**	**RAN1**	**RAN2, RAN4**	**SI for entire Rel-18 timeframe focusing on limited identified use cases**
3	Additional Use Cases for AI/ML for NG-RAN	SI	RAN3		Study of additional higher layer use cases and further architecture and framework enhancement

Use cases for AI/ML for Air-Interface

- Many companies have internally studied diverse use cases and identified select use cases that they think are good candidates for 3GPP discussion.
- Given the potentially vast scope and workload of this new AI/ML study in RAN1, we believe that, to ensure best progress of the study, it is better to bring a limited number of use cases into Rel-18 SI.
- Therefore, we propose to base further discussion on use cases for candidate Rel-18 SI on AI/ML for Air-Interface on the following set of use cases:
  - CSI feedback (CSI compression, etc.)
  - Beam management (beam selection, beam recovery, predictive mobility, handover, etc.)
  - Positioning
  - RS overhead reduction (channel estimation, etc.)
- We propose that RAN continue to discuss the above use cases for possible further down-selection prior to the start of the Rel-18 SI and detailed scope of them.

Evaluation Methodology and KPIs for AI/ML for Air-Interface SI

- **Proposal**: [slightly modified from last discussed version over R18 NWM discussion]
  - Base the evaluation methodology in AI/ML for Air-Interface study on existing 3GPP framework for evaluations, i.e., *statistical* channel models (*from* TR 38.901 and TR 38.857 [positioning]), link and system level simulations, etc.
    - Additionally, companies may consider using field data to *further* assess the performance *and robustness* in real-world environments. *How the field data set is obtained (per company vs. common data set) needs to be further discussed.*
  - KPIs are broadly understood to be use case specific but a number of common metrics prevail, e.g., complexity and performance vs. proper (non-AI/ML) *state-of-the-art* baseline. Similarly, overhead associated with enabling respective AI/ML scheme should be well documented and accounted for.
  - Evaluated AI/ML schemes should be based on offline training for evaluation.

Nomenclature for network and UE involvement (1/3)

- There are different levels of collaboration and signaling exchange between network and UE for AI/ML operation:
  - For **inference**:
    - Signaling, assistance information, and/or measurement to aid AI/ML operation, such as input to AI/ML model
    - Signaling and/or feedback of output from AI/ML inference
  - For **training**:
    - Signaling and procedures to provide training data for online training
  - For both **inference and training**:
    - Model upload and download
- Importance of agreeing on the **nomenclature** of different types of inter-node collaboration regimes
  - to have a common understanding and to facilitate the discussion by establishing a common language

Nomenclature for network and UE involvement (2/3)

- We propose to define the following categories:
- For **inference**:
  - InfCatNone: No collaboration is needed for inference
    - InfCatNoneA: AI/ML inference purely implementation based and not requiring air-interface changes.
    - InfCatNoneB: AI/ML inference purely implementation based and facilitated by air-Interface modification (such as a new RS pattern)
  - InfCatIn: AI/ML inference requiring assistance input over-the-air, where the assistance input is directly related to the inference result.
  - InfCatOut: AI/ML inference whose output is transmitted over-the-air
  - InfCatJoint: AI/ML whose inference occurs jointly between network and UE
  - Note that a given AI/ML algorithm may belong to multiple categories.
- For **training**
  - TrainCatNone: No collaboration framework is needed for training. Examples include offline training, or an online training that does not need any inter-node assistance.
  - TrainCatData: Inter-node assistance for training AI/ML algorithm. Examples include online training with training data provided by the other node.
  - TrainCatModel: Training requiring AI/ML model exchange between network nodes during training. Examples include federated learning, or potentially online training of AI/ML model belonging to InfCatJoint.

Nomenclature for network and UE involvement (3/3)

- So, as an example, TrainCatNone_InfCatData refers to
  - AI/ML algorithm framework needing assistance info for inference but no assistance data for training,
- and TrainCatData_InfCatNone
  - for AI/ML algorithm framework needing no assistance info for inference but utilizing assistance training data for online training.

- **Note**: the categorization is mainly to facilitate discussion by establishing a common language and is NOT meant to limit interference/training behaviors discussed in R18 in any way.

AI/ML procedure for AI/ML for Air-Interface SI (RAN2/RAN3)

- R18 should study architecture and procedure enhancements for AI/ML, including
  - Network entities and interfaces for model training/inference, data management, model management
  - Procedures for online training and federated learning
  - Procedures for inference configuration in UE and network
  - Procedures for data management and model management
  - Mobility support for AI/ML based solutions
  - UE assistance for AI/ML configuration.

Qualcomm

Thank you!

Follow us on:

For more information, visit us at:

www.qualcomm.com & www.qualcomm.com/blog

Nothing in these materials is an offer to sell any of the components or devices referenced herein.

©2018 Qualcomm Technologies, Inc. and/or its affiliated companies. All Rights Reserved.

Qualcomm is a trademark of Qualcomm Incorporated, registered in the United States and other countries. Other products and brand names may be trademarks or registered trademarks of their respective owners.

References in this presentation to “Qualcomm” may mean Qualcomm Incorporated, Qualcomm Technologies, Inc., and/or other subsidiaries or business units within the Qualcomm corporate structure, as applicable. Qualcomm Incorporated includes Qualcomm’s licensing business, QTL, and the vast majority of its patent portfolio. Qualcomm Technologies, Inc., a wholly-owned subsidiary of Qualcomm Incorporated, operates, along with its subsidiaries, substantially all of Qualcomm’s engineering, research and development functions, and substantially all of its product and services businesses, including its semiconductor business, QCT.