vivo

3GPP TSG RAN meeting #100 RP-231020

Taipei, June 15 - 16, 2023

Agenda Item: 9.2.5

Source: vivo

Title: Discussion on Rel-18 AI/ML for air interface progress and handling of AI/ML based CSI prediction spec impact

Document for: Discussion and decision

Content

- **RAN1 Progress assessment**
- **CSI prediction handling**

RAN1 progress assessment (Framework)

- **The following has been finished**
  - Definitions of common AI/ML terminologies
  - Definitions of main aspects of lifecycle management
  - Definition of Collaboration levels : x/y/z
  - Model transfer categorization and boundary clarification
  - Model/functionality identification concept and high level categorization
- **There are still the following issues to be addressed before RAN1 declare finish of the area**
  - Model transfer
    - Observation and feasibility analysis on Cases z1~z5;
  - Functionality/Model identification
    - Procedures of different model identification types;
    - Observation and applicability analysis of different identification procedures;
  - Data collection
    - Dataset categorization to facilitate zone/site/scenario specific operation;
  - Providing inputs for RAN2, RAN4 and SA for further analysis and more detailed study;
    - May need per use case discussion, e.g., data collection content/size, model transfer size and frequency;

RAN1 progress assessment (CSI compression)

- **The following has been finished**
  - Evaluation methodology and initial observations (9.2.2.1) on:
    - Basic AI/ML performance gain for CSI compression (including eventual KPI)
    - Comparison of AI/ML solutions, including input/output types, quantization/dequantization, ground-truth CSI format, Rank>1 solutions, CQI determination, etc.
    - Generalization (scenario, UE distribution, carrier frequency, etc.) and scalability (input/output dimension etc.)
    - Training collaborations (Observation for case 1 in separate training)
  - Specification impact analysis framework and initial observation (9.2.2.2) on:
    - Data collection (assisted information, reporting format, etc.)
    - Monitoring (NW side monitoring and UE side monitoring)
    - Inference related procedure (payload indication, CQI/RI determination, etc.)
- **There are still the following issues to be addressed before RAN1 declare finish of work**
  - Finish making evaluation observations (9.2.2.1) on:
    - Monitoring methods
    - All cases in multi-vendor joint training; Case 2 and case 3 in separate training
  - Finish making spec impact observations (9.2.2.2) on:
    - Training collaboration comparison (type 1, type 2, and type 3)

RAN1 progress assessment (CSI Prediction)

- **The following has been finished**
  - Evaluation methodology and initial observations (9.2.2.1) on:
    - Basic AI/ML performance gain for CSI prediction (including eventual KPI)
    - Comparison of non-AI/ML and AI/ML solutions, e.g., sample hold, AR, and AI-based CSI prediction, etc.
    - Generalization (speed)
- **There are still the following issues to be addressed before RAN1 declare finish of work**
  - Finish making evaluation observations (9.2.2.1) on:
    - Generalization (deployment scenarios, carrier frequency, etc.)
    - Impact of observation window
    - Monitoring methods
  - Finish making spec impact observations (9.2.2.2) on:
    - Data collection procedure, mainly including RS configuration, measurement and report configuration
    - Monitoring procedure and metric
    - Model/functionality selection/switching and finetuning procedure

RAN1 progress assessment (Beam)

- **The following has been finished**
  - Evaluation methodology and initial observations on (9.2.3.1) :
    - Simulation aspects: basic simulation assumption, baseline performance, and KPI definition
    - Generalization consideration: different scenarios/configurations and various Set B of beam
    - General observations: Different Set B for fixed pattern, current quantization scheme, DL Tx beam prediction for BM-Case1
    - General generalization observation
  - Specification impact identified on (9.2.3.2) :
    - Data collection, model inference, general model monitoring for two BM sub-use cases
- **There are still the following issues to be addressed before RAN1 declare finish of work**
  - General observations: beam pair prediction for BM-Case1, different Set B(including Opt 2B/2C/2D), different Rx beam assumption for DL Tx beam prediction, low-precise quantization, quasi-Rx beam, and DL Rx beam prediction and beam pair prediction for BM-Case2.
  - Detail generalization observations: deployment scenarios, ISD, UE distribution, various Set B, UE Rx assumption, various UE parameters, various gNB setting and UE speed(for BM-Case2)
  - Categorize specification impacts for functionality/model-ID based LCM
  - Down-selection on identified performance monitoring metrics

RAN1 progress assessment (Positioning)

- **The following has been finished**
  - Performance observation on direct AI/ML positioning and AI/ML assisted positioning and both can achieve sub-meter @90% positioning accuracy in heavy NLOS scenarios;
  - Generalization observation on AI/ML based positioning show that it may suffer from limited generalization capability for unseen configurations and can be greatly improved by model fine-tuning and dataset enhancement.
  - Multiple types of input and output of AI/ML model are identified.
  - The entities responsible for data collection and the information to be collected are identified.
  - The entities responsible for model monitoring and two model monitoring methods are identified.
- **There are still the following issues to be addressed before RAN1 declare finish of work**
  - Observation on benefits and spec impacts for various types of model input and output;
  - Remaining issues for identification of necessary cooperation among entities for monitoring
  - Remaining issues for identification of necessary cooperation among entities for data collection
  - Functionality/Model identification: The related aspects should be separately identified for case 1~3b of AI/ML based positioning.

Rel-18 AI/ML for air interface progress assessment

- Based on assessment in previous slides, key components are still missing to conclude the study, e.g., to address the following scope in SID:
  - Identify various levels of collaboration between UE and gNB pertinent to the selected use cases, e.g.,
    - No collaboration: implementation-based only AI/ML algorithms without information exchange [for comparison purposes]
    - Various levels of UE/gNB collaboration targeting at separate or joint ML operation.

- **Proposal: To finish the study in RAN1 for Rel-18 AI/ML for air interface, more focus should be put on the following areas in August RAN1 meeting**
  - **Framework:**
    - **Model transfer feasibility observation and conclusions;**
    - **Dataset categorization;**
  - **CSI**
    - **CSI compression: comparison of training collaboration types (type 1, type 2, and type 3);**
    - **CSI prediction spec impact analysis;**

Handling of AI/ML based CSI prediction spec impact

In RAN1 #112 bis-e, the following agreement is achieved:

Agreement

In CSI prediction using UE-side model use case, whether to address the potential spec impact of CSI prediction depends on RAN#100 final conclusion, focusing on the following

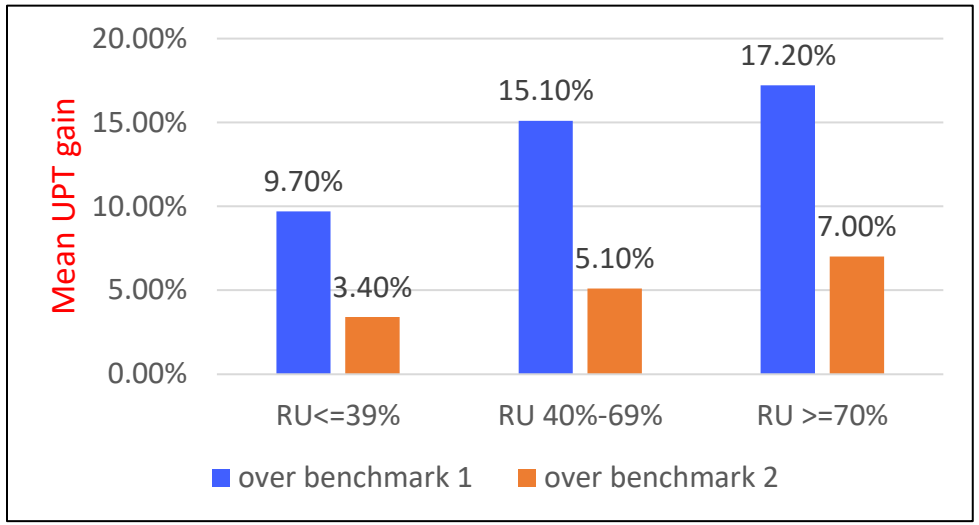
- data collection procedure, mainly including RS configuration, measurement and report configuration, resusing as much as possible what is defined for UE side use cases
- monitoring procedure and metric for AI-based CSI prediction.
- Model/functionality selection/switching and finetuning procedure.
- Note: Discussion on potential specification impact is limited to aspects which would NOT duplicate the work in Rel-18 MIMO WI.
- Note: Minimize LCM related potential specification impact discussion that follow the high-level principle of other one-sided model sub-cases.

Gain of AI/ML based CSI prediction over benchmarks

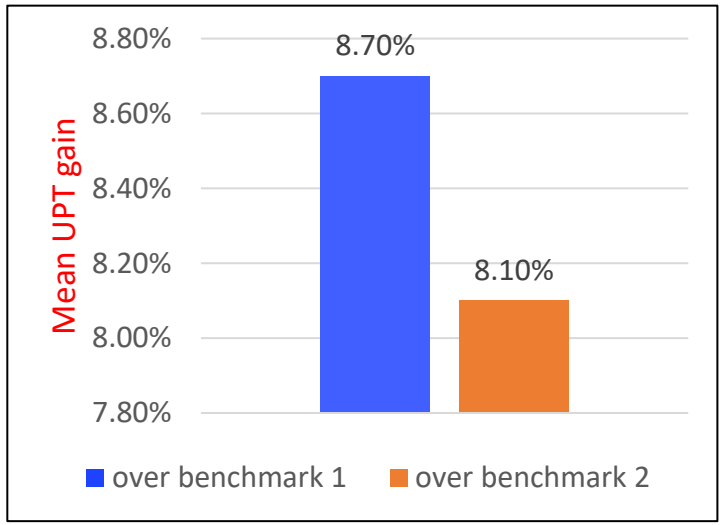
vivo

- ✓ Our simulation results show that the AI based CSI prediction **significantly outperforms** both the benchmark 1 (nearest historical CSI) and benchmark 2 (auto-regression).

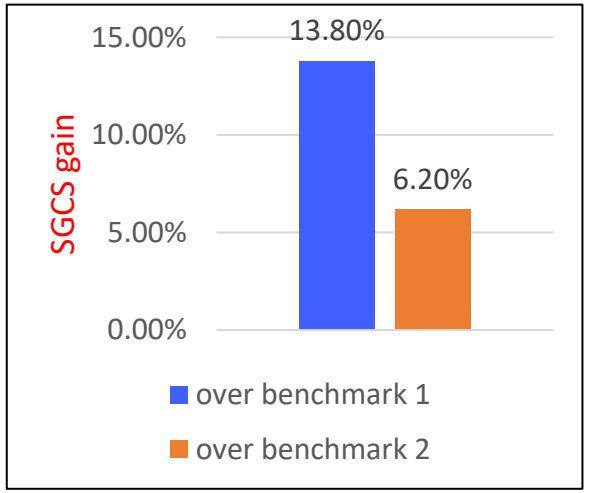
Eventual KPI: UPT, FTP traffic

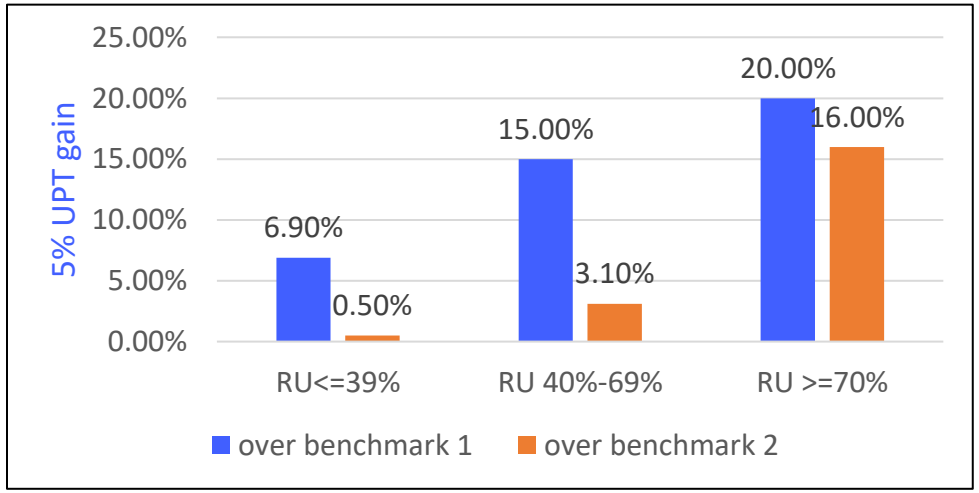


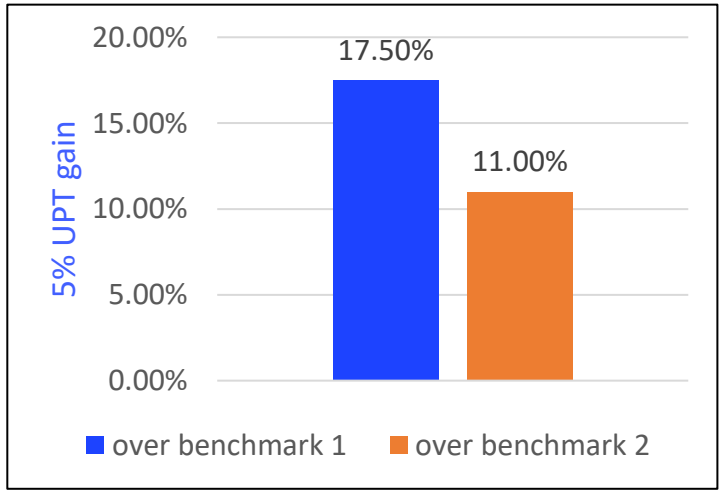
Eventual KPI: UPT, full buffer traffic



Intermediate KPI: SGCS







Simulation parameters: Uma 38.901; 7 cells, 3 sectors for each cell, 10 user for each sector; carrier frequency 2GHz, subcarrier spacing 15KHz, 13 subbands (10MHz, 4RBs/subband), 32 gNB antenna ([Mg Ng M N P; Mp Np] = [1 1 2 8 2; 2 8]), 2 UE antenna ([Mg Ng M N P; Mp Np] = [1 1 1 1 2; 1 1]), 100% outdoor UE, Channel type: Uma, NLOS, Period of CSI-RS: 5ms, UE speed: 60km/h; Input of AI model for CSI prediction: 10 raw historic channels in PRB, the spatial consistency procedure A with 50m decorrelation distance is used where the channel updating periodicity is 1 ms. AI-based CSI compression models: Transformer model with 64 bits payload.

Handling of AI/ML based CSI prediction spec impact

In RAN1 #113, observations are drawn where

- Majority companies found that AI-based CSI prediction has significant SGCS gain (6%~41.75% for 30km/h) and mean UPT gain at high RU (4.2%~48.6%) over benchmark #1.
- 2 companies found that AI-based CSI prediction has moderate mean UPT gain at high RU (3.1%~7.0%) over benchmark #2..

Observation

For the AI/ML based CSI prediction, till the RAN1#113 meeting, compared to the Benchmark#1 of the nearest historical CSI, in terms of SGCS, from UE speed perspective, in general the gain of AI/ML based solution is related with the UE speed:

- For 10km/h UE speed, 4 sources [Fujitsu, Samsung, Xiaomi, InterDigital] observe 1.03%~6% gain, 1 source [CMCC] observes 21.93% gain.
- For 30km/h UE speed, 2 sources [OPPO, ETRI] observes 6%~10.43% gain, 5 sources [ZTE, Fujitsu, Apple, Xiaomi, Spreadtrum] observe 18.72%~31.3% gain, and 3 sources [InterDigital, MediaTek, CMCC] observe 35%~41.75% gain, which are in general larger than 10km/h UE speed.
- For 60km/h UE speed, 2 sources [Fujitsu, InterDigital] observe -3%~5% gain, 4 sources [Huawei, Samsung, vivo, CMCC] observe 11.2%~19.98% gain, which are in general smaller than 30km/h UE speed.
- Note: the above results are based on the following assumptions
  - The observation window considers to start as early as 15ms~50ms.
  - A future 4ms or 5ms instance from the prediction output is considered for calculating the metric.
  - Raw channel matrix is considered as model input
  - The performance metric is SGCS in linear value for layer 1.
  - No post processing is considered.
  - No spatial consistency is considered by 11 sources [Fujitsu, Samsung, Xiaomi, InterDigital, CMCC, OPPO, ETRI, ZTE, Apple, Huawei, Spreadtrum]. 1 source [vivo] provides both results with spatial consistency and results w/o spatial consistency.
  - Note: Results refer to Table 5.1-1 of R1-2306059

Observation

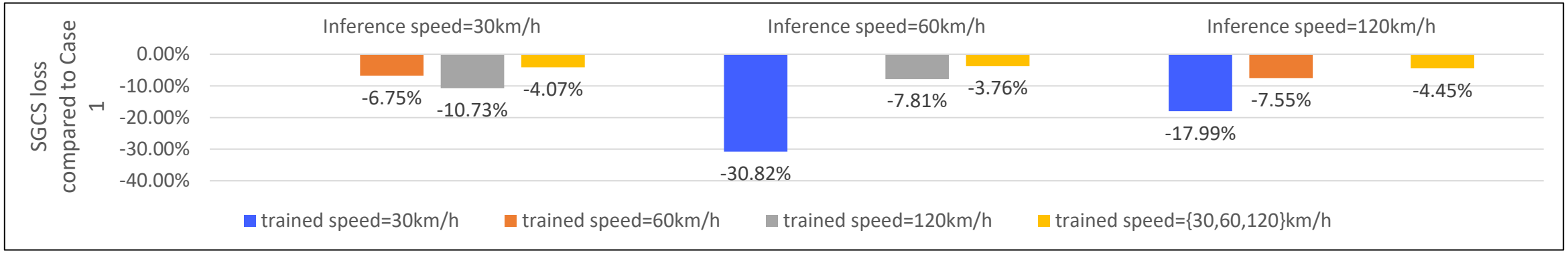
For the AI/ML based CSI prediction, till the RAN1#113 meeting, in terms of mean UPT, gains are observed compared to both Benchmark#1 of the nearest historical CSI and Benchmark#2 of a non-AI/ML based CSI prediction approach:

- Compared to the benchmark of the nearest historical CSI:
  - For FTP traffic:
    - 1 source [Huawei] observes 1.2%~4.2% gain;
    - 1 source [Apple] observes 7.6%~8.5% gain;
    - 1 source [vivo] observes 9.7%~17.2% gain.
    - 1 source [MediaTek] observes 22.6%~48.6% gain.
  - For full buffer traffic:
    - 1 source [Nokia] observes 2%~3% gain;
    - 1 source [vivo] observes 8.7% gain.
    - 1 source [MediaTek] observes 1.01% gain.
- Compared to the benchmark of an auto-regression based CSI prediction:
  - For FTP traffic:
    - 1 source [Huawei] observes 0.7%~3.1% gain;
    - 1 source [vivo] observes 3.4%~7.0% gain.
  - For full buffer traffic:
    - 1 source [vivo] observes 8.1% gain.

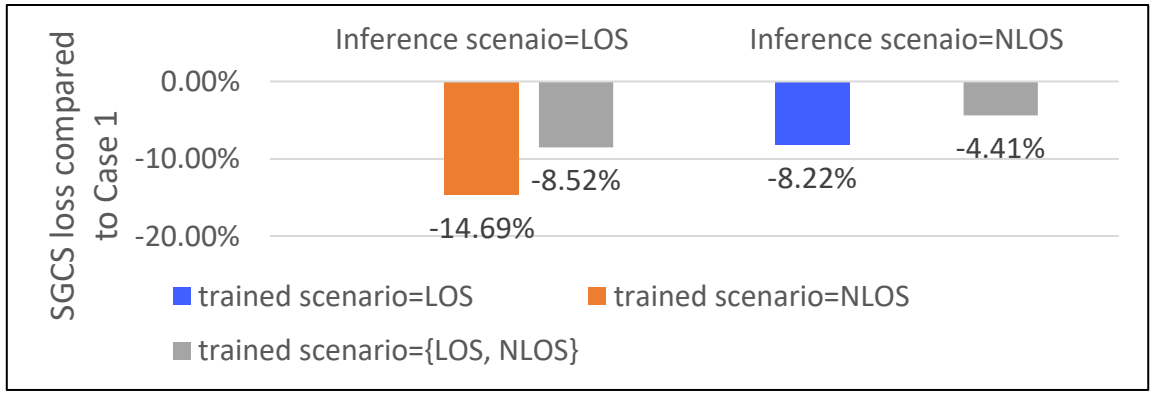
Generalization aspects of AI/ML based CSI prediction

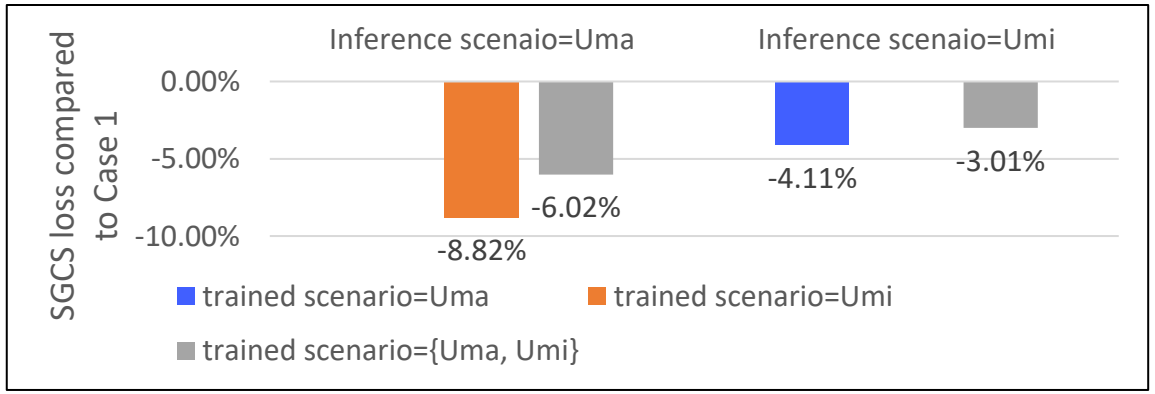
vivo

➢ Generalization over speeds



➢ Generalization over deployment scenarios





- ✓ Generalization of AI-based CSI prediction is not good if the model trained with one speed/scenario is inferred in other speeds/scenarios, SGCS loss is -4.11%~-30.82%.
- ✓ Mixed dataset can improve the generalization performance while the prediction accuracy is still worse than that of speed/scenario specific models, SGCS loss is -3.01%~-4.45%.

Simulation parameters: Uma 38.901; 7 cells, 3 sectors for each cell, 10 user for each sector; carrier frequency 2GHz, subcarrier spacing 15KHz, 13 subbands (10MHz, 4RBs/subband), 32 gNB antenna ([Mg Ng M N P; Mp Np] = [1 1 2 8 2; 2 8]), 2 UE antenna ([Mg Ng M N P; Mp Np] = [1 1 1 1 2; 1 1]), 100% outdoor UE, Period of CSI-RS: 5ms; Input of AI model for CSI prediction: 10 raw historic channels in PRB, the spatial consistency procedure A with 50m decorrelation distance is used where the channel updating periodicity is 1 ms.

Generalization aspects of AI/ML based CSI prediction

➢ Generalization over speeds

In RAN1 #113, one observation is drawn where majority companies found that

- Generalization of AI-based CSI prediction is not good if the model trained with one speed/scenario is inferred in other speeds/scenarios.
- Mixed dataset can improve the generalization performance while the prediction accuracy is still worse than that of speed/scenario specific models.

Observation

For the generalization verification of AI/ML based CSI prediction over various UE speeds, till the RAN1#113 meeting, compared to the generalization Case 1 where the AI/ML model is trained with dataset subject to a certain UE speed#B and applied for inference with a same UE speed#B,

- For generalization Case 2, generalized performance may be achieved for some certain combinations of UE speed#A and UE speed#B but not for others:
  - If UE speed#B is 10 km/h & UE speed#A is 30 km/h, 4 sources [Xiaomi, CATT, Interdigital, Spreadtrum] observe a generalized performance of less than -2% degradation.
  - If UE speed#B is either 30 km/h or 60 km/h or 120 km/h, or if UE speed#B is 10km/h and UE speed#A is either 60km/h or 120km/h, 8 sources [Xiaomi, Samsung, Interdigital, Fujitsu, ZTE, ETRI, vivo, Huawei] observe that moderate/significant performance degradations are suffered:
    - For UE speed#B is 10 km/h & UE speed#A is either 60 km/h or 120 km/h, 1 source [Xiaomi] observes moderate degradation (-2.7% loss), 1 source [Samsung] observes significant degradation (-53%~-61% loss).
    - For UE speed#B is 30 km/h & UE speed#A is either 10 km/h, 60 km/h or 120 km/h, 1 source [Xiaomi] observes moderate degradation (-3% loss), 8 sources [Xiaomi, Interdigital, Fujitsu, vivo, ZTE, Huawei, ETRI, Spreadtrum] observe significant degradation (-6%~-45.6% loss).
    - For UE speed#B is 60 km/h & UE speed#A is either 10 km/h, 30 km/h or 120 km/h, 1 source [ZTE] observes moderate degradation (-3% loss), 7 sources [Samsung, Xiaomi, Fujitsu, ETRI, ZTE, vivo, Spreadtrum] observe significant degradation (-7.8%~-52% loss).
    - For UE speed#B is 120 km/h & UE speed#A is either 30 km/h or 60 km/h, 1 source [ZTE] observes moderate degradation (-3.4% loss), 4 sources [ZTE, ETRI, vivo, Samsung] observe significant degradation (-7.55%~-32.3% loss).
- For generalization Case 3, generalized performance of the AI/ML model can be achieved in general (0%~-4.45% loss) for UE speed#B subject to any of 10 km/h, 30 km/h, 60 km/h and 120 km/h, if the training dataset is constructed with data samples subject to multiple UE speeds including UE speed#B, as observed by 9 sources [Xiaomi, Interdigital, Apple, Huawei, ZTE, Samsung, ETRI, vivo, Spreadtrum].
  - For UE speed#B is 10 km/h, minor loss (-0.6%~-1%) are observed by 3 sources [CATT, Xiaomi, Spreadtrum].
  - For UE speed#B is 30 km/h, minor loss (-0.08%~-1.34%) are observed by 3 sources [Xiaomi, Apple, Huawei], moderate loss (-2.2%~-4.07%) are observed by 3 sources [Interdigital, vivo, Spreadtrum].
  - For UE speed#B is 60 km/h, minor loss (-0.05%~-2%) are observed by 4 sources [ZTE, Apple, Xiaomi, Huawei], moderate loss (-2%~-3.76%) are observed by 2 sources [vivo, Spreadtrum].
  - For UE speed#B is 120 km/h, moderate loss (-2%~-4.45%) are observed by 4 sources [vivo, Samsung, ETRI, ZTE].
  - Note: For generalization Case 3, 5 sources [ETRI, ZTE, Samsung, Interdigital, Fujitsu] observe significant performance degradations (-5%~-26.5% loss) for UE speed#B subject to 10 km/h, 30 km/h, 60 km/h, but compared with generalization Case 2, in general the performance are still improved.

Analysis on AI/ML based CSI prediction

vivo

- Observation: for the eventual KPI gain of AI/ML based CSI prediction over benchmarks
  - ✓ In the high RU, the mean UPT gains over benchmark#1 and benchmark#2 are 8%~17% and 7%~8%, respectively; the 5% UPT gains over benchmark#1 and benchmark#2 are 17%~20% and 11%~16%, respectively.
- Conclusion: mean UPT gain of 7%~17% and 5% UPT gain of 11%~20% at high RU offered by the AI/ML based CSI prediction proves the necessity of AI/ML based CSI-enhancement.

- Observations: for the generalization aspects of AI/ML based CSI prediction
  - ✓ The losses due to the generalization as opposed to the specific model in AI/ML based CSI prediction are -3%~-30%, if over the different speeds/scenarios or their mixture.
- Conclusion: In order to harvest the gain of AI/ML based CSI prediction, the generalization loss should be avoided.

- **Proposal: Study specification impacts related to generalization, monitoring and model adjustment (model selection/switching and finetuning) to harvest the gain of AI/ML based CSI prediction after RAN #100 meeting.**

vivo

THANK YOU.
谢谢。