**3GPP TSG-RAN WG3 Meeting #111electronic R3-210980**

**25 January – 4 February 2021**

**Agenda Item: 30**

**Source: CMCC**

**Title: Summary of offline discussion on EnhDataColl**

**Document for: Discussion and Decision**

# 1 Introduction

**CB: # 17\_Basket\_AI**

**- resolution of FFSs**

**- high-level framework for continuation of SI**

**- new use case: energy saving? (lower prio discussion?)**

**- If consensus, capture any general principles / descriptions / open issues / WA / agreements for upcoming RAN3 work (to be captured in Chair’s Notes)**

**- suggest to avoid discussing details (CRs,TPs, etc.) unless there is full agreement**

(CMCC - moderator)

Summary of offline disc R3-210980

We intend to achieve some high-level agreements during the first phase discussion and proceed with the TPs if possible in the second phase.

# 2 For the Chairman’s Notes

**To be added after email discussion.**

# 3 Discussion

## 3.1 TR 37.817 v0.1.0

The TR [1] is updated based on the agreements on RAN2#110 E-meeting. It has been checked over the email reflector after RAN3 #110e meeting and resubmitted to RAN3 #111e for agreement.

**Proposal 1. Agree the TR 37.817 v0.1.0**

## 3.2 High-level AI framework

Following are open issues left for AI framework:

*Editor's Note: the details for the framework below is FFS including whether Actor and Subject of action should be in one box or separate, whether feedback from action to Model training host is needed, the name in each box is from functionality or from processing point of view, the feedback from Subject of action to the Data sources is Performance feedback or Model performance feedback and other possible refinement.*

Paper [2][3][4] address on these open issues.

### **3.2.1 AI framework from functionality or from processing point of view**

In R3-210917, two alternatives for illustration of the AI functional framework are discussed:



Figure 1 Alternative 1: AI framework from functionality point of view (currently captured in the TR 37.817)



Figure 2 Alternative 2: AI framework from processing point of view

After some comparison, it is found that the two alternatives do not differ two much. It is proposed that if we cannot reach consensus at this stage, we could keep the current alternative as in the TR and refine it at later stage when we found it is not suitable for use case and solution description.

On the other hand, it is pointed out in R3-210617 that, currently in Figure1, each box represents one processing host to enable AI functionality. However, for one AI functionality in one use case, multiple processing hosts may be used. On the other hand, there is a possibility that more than one AI functionality can be supported by the same one processing host. Therefore, it is preferred for each box to represent one processing action to enable AI functionality including data collection, model training, model inference and action.

Companies are invited to provide views on whether to keep the current alternative as in the TR or choose alternative 2:

**Q1: Do you agree to keep the current alternative as in the TR or choose alternative 2?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| ZTE | Prefer AI framework from processing point of view.  (Alternative 2 is partly OK) | The framework we proposed as follow:    The framework aims to show the entire the ML operation process for RAN intelligence. The framework should be separated from the RAN architecture (including RAN logical node). We prefer the framework comprising four boxes including data collection, model inference, model training and action.  The difference between alternative 2 and framework:  - Performance feedback from action to model training.  - Clarify that model training includes the online training. |
|  |  |  |
|  |  |  |
|  |  |  |

### **3.2.2 Whether Actor and Subject of action should be in one box or separate?**

In R3-210917, it is observed that for the purpose of easy identifying the signalling exchange between network nodes, the separation of actor and subject of action seems to be beneficial.

In R3-210785, it is pointed that one box is enough to reflect the function of action. In this framework, the process of the AI should be clearly defined, and the place where model training, model inference, data collection and action (involve actor and subject of action) needs to be discussed case by case.

in R3-210617, it is proposed that to make the Functional Framework simple and straightforward, the separate “Actor” and “Subject of action” boxes can be merged into one block (e.g. Action). The “Action” box does not restrict that only one node or interface is involved for one action, if clear explanation is necessary, one note can be added that “one or more Subjects of Action(s) may act over at least one interface”.

Companies are invited to provide views on whether actor and subject of action should be in one box or separate:

**Q2: Whether actor and subject of action should be in one box or separate?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| ZTE | Yes | The framework aims to show the entire the ML operation process for RAN intelligence. The framework should be separated from the RAN architecture (including RAN logical node).Model inference executes the trained model to get the output based on the inference data. Then, specific policy needs to be generated based on the output. One box is enough to reflect the function of action. |
|  |  |  |
|  |  |  |
|  |  |  |

### **3.2.3 Whether feedback from action to data sources is performance feedback or model performance feedback**

Both R3-210917 and R3-210785 think the name “Model performance feedback” is appropriate. Since ML inference is a process of using a trained ML model to make a prediction or guide the decision based on collected inference data and ML model. The output can be feedback to the model training host to verify the performance of the ML model and in turn help the model training host to improve or re-select the ML model.

Companies are invited to provide views on whether feedback from action to data sources is performance feedback or model performance feedback:

**Q3: Whether feedback from action to data sources is performance feedback or model performance feedback?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| ZTE | Feedback from action to data sources is performance feedback.  Feedback from inference to training is model performance feedback. | Feedback from action to data sources is performance feedback.  For the feedback between Model inference and Model training, since Model inference is one component which execute the trained model to get the output based on the inference data, this performance reflects the ML model performance is good or not. If this model performance feedback is not good, ML model needs to be re-selected or re-trained in the Model training component. So feedback between Model training and Model inference is called “Model performance feedback”. Taking AI-based energy saving as an example, the output of the model inference may be the predicted load of the cell.  For the feedback from Action to data collection, this feedback reflects the network performance after the Action. After the policy is adopted , the network performance may be optimized/maintained/degraded. Taking AI-based energy saving as an example, the feedback from Action is the KPI of the network. |
|  |  |  |
|  |  |  |
|  |  |  |

### **3.2.4 Whether feedback from action to model training host is needed**

Some company deems that performance feedback from action to data source is enough [2], but others find that the feedback from action to model training is needed for re-training or reinforcement learning [3].

Companies are invited to provide views on whether feedback from action to model training host is needed:

**Q4: Whether feedback from action to model training host is needed?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| ZTE | Yes | Model training is one component that train the ML model offline or online according to different use case, so re-training or updating model is also one part of model training. Model training may be triggered to retrain the ML model when the model performance feedback is not good. Similarly, if the network performance feedback from Action is not good, Model training should also be triggered to retrain the ML model or update the ML model.  On the other hand, reinforcement learning, as one of the ML techniques (online training), needs to interact with the network environment during training phase. During training phase, the agent in reinforcement learning needs to get the reward from the environment through performance feedback from action. If the framework should support the reinforcement learning, the performance feedback to the model training is essential. |
|  |  |  |
|  |  |  |
|  |  |  |

### **3.2.5 Other open issues for AI framework**

In R3-210617, some other open issues are discussed and following proposals are proposed:

**Proposal 3: RAN3 supports the case that one ML model demands input from other ML models.**

**Proposal 4: For the sake of discussion, RAN3 further distinguishes the decision-oriented ML model and the prediction-oriented ML model when it comes to ML model feedback provision and ML model retraining/update.**

**Proposal 5: The data collection shall provide the training data according to the demand of model training regarding what/when/how to provide. It is regarded as model performance feedback when the training data provision is triggered due to bad model performance.**

**Proposal 6: The data collection also collects the output from the model inference.**

Since above issues have not been touched in last RAN3 meeting, companies are invited to provide views on whether to discuss these open issues in this offline discussion:

**Q5: Whether to discuss above open issues proposed in R3-210617 in this offline discussion?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| ZTE |  | P3, P4 may discuss case by case later..  For P5, we think there is no need to add the data provision policy from model training to data collection into the current AI framework. The framework aims to show the entire the ML operation process for RAN intelligence rather than signaling design  For P6, not needed, model training can get the prediction result through the model performance feedback from model inference to model training. |
|  |  |  |
|  |  |  |
|  |  |  |

## 3.3 Use case

### **3.3.1 Use case priority**

In R3-210073, it is observed that most interested use cases are: energy saving, traffic/load prediction, trajectory prediction, traffic steering and load balancing. Moreover, some of them can be used as a tool box by other use cases, e.g., load prediction, UE trajectory prediction. Therefore, it is proposed to study the below use cases in R17 with high priority:

1. Tool box use cases:

- traffic/load prediction

- trajectory prediction

1. System level use cases:

- energy saving

Companies are invited to provide views on whether to study above use cases in R17 with high priority:

**Q6: Whether to study above use cases in R17 with high priority?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| ZTE | Yes | The most interested use cases are: energy saving, traffic/load prediction, trajectory prediction, traffic steering and load balancing.  Some of them can be regarded as system level use case. Take energy saving as an example, model training can be deployed outside RAN, e.g., OAM system, the whole system solution including three main functions, AI based scenario classification, AI based Load prediction, and AI based ES performance analysis.  While some of them can be used as a tool box by other use cases, e.g., load prediction, UE trajectory prediction. For example, load prediction can contribute to use cases like energy saving, load balancing, mobility optimization. UE trajectory prediction can also contribute to use cases like mobility optimization, energy saving. For those tool box use cases which can be achieved purely on the RAN side for RAN purpose. |
|  |  |  |
|  |  |  |
|  |  |  |

### **3.3.2 Use case description**

R3-210918, AI based Energy Saving is used as the starting point and intend to work out an example on how to describe the use case and solutions. Companies are invited to provide views on the description if any:

**Q7: Do you agree to use Energy Saving as the starting point and what’s your comments about the description in R3-210918 if any?**

|  |  |  |
| --- | --- | --- |
| **Company** | **Yes/No** | **Reasons/Comments/Suggestions** |
| ZTE | The detail use case description can be further discussed in the next meeting. | As we analyzed above, there are three main functions for AI based energy saving (AI based scenario classification, AI based Load prediction, and AI based ES performance analysis), for each function, how to deploy the AI framework into the current system architecture, and the detail inputs and outputs are different, which needs to be further discussed. |
|  |  |  |
|  |  |  |
|  |  |  |

# 4 Conclusion, Recommendations

To be edited, if needed**.**

# 5 Reference

1. R3-210935, TR 37.817 v0.1.0 (CMCC)
2. R3-210917, Further discussion on high-level AI framework (CMCC)
3. R3-210785, High-level framework and definition for AI RAN (ZTE Corporation)
4. R3-210617, Open issues of framework for AI (Lenovo, Motorola Mobility)
5. R3-210918, AI based Energy Saving (CMCC)
6. R3-210073, Use cases for AI study (ZTE Corporation)