|  |  |  |
| --- | --- | --- |
|  | INTERNATIONAL TELECOMMUNICATION UNION**TELECOMMUNICATIONSTANDARDIZATION SECTOR**STUDY PERIOD 2017-2020 | **Focus Group on Machine Learning for Future Networks including 5G** |
| **ML5G-I-223-R1** |
| **Original: English** |
| **Question(s):** | N/A | 8th meeting 2020 |
| **INPUT DOCUMENT** |
| **Source:** | FG ML5G |
| **Title:** | A compilation of problem statements and resources for ITU Global Challenge on AI/ML in 5G networks |
| **Contact:** | Yang MengjiaChina MobileP.R.China | Email: yangmengjia@chinamobile.comTel: +86 13811206654  |
| **Contact:** | Xie YuxuanChina MobileP.R.China | Email: xieyuxuan@chinamobile.com  |
| **Contact:** | Jia ZihanChina MobileP.R.China | Email: jiazihan@cmdi.chinamobile.com Tel: +86 13810024426  |
| **Contact:** | Zhu LinChina MobileP.R.China | Email: zhulinyj@chinamobile.com  |
| **Contact:** | Tengfei LiuChina UnicomP.R.China | Tel: + 86 15652955883Fax: +010 68799999Email: liutf24@chinaunicom.cn  |
| **Contact:** | Vishnu Ram | E-mail: AI5GChallenge@itu.int  |
| **Contact:** | Thomas Basikolo | E-mail: AI5GChallenge@itu.int |

|  |  |
| --- | --- |
| **Keywords:** | AI, Challenge, ML, Sandbox, Data, Resources |
| **Abstract:** | This contribution compiles the list of problem statements and resources contributed by the Focus Group members and partners towards the ITU AI/ML5G Global Challenge. The resources are intended to be a reference list to be used for pointer towards data, toolsets and partners to setup sandboxes for the ITU AI/ML5G Challenge. The problem statements are intended to be analysed, short-listed and used for the challenge to be solved by participants. |

## References

[ITU-T ML5G-I-217-R1] ITU Global Challenge on AI/ML in 5G networks

# 1. Introduction

[ITU-T ML5G-I-217-R1] described the proposal for ITU Global Challenge on AI/ML in 5G networks.

Problem statements which are relevant to ITU and IMT-2020 networks are the backbone of the challenge. They should be aligned with the theme/tracks of the challenge and should provide enough intellectual challenge while being practical within the time period of the challenge. They should address short term pain points for industry while pointing to long term research directions for academia. In addition, many of them may need quality data to solve them. This contribution collates the problem statements from our partners in a standard format. Future steps for these problem statements are:

* analyse the submitted problem statements from our partners and colleagues,
* present them for selection by the challenge management team
* host the selected problem statements on the challenge website.

While discussing and disseminating the challenge with our partners, an important and frequent question posed to us is about the relevant resources. This document contains a collection of resources pointed to us by our members and partners in the context of ITU ML5G global challenge. This is an attempt to compile and classify them so that it is useful to all our partners. We invite our members and partners to add pointers to private as well as public resources which may be of relevance to the Challenge.

# 2. Problem statements

NOTE 1- the structure of the list below is derived from the many discussions that we had with partners across the globe.

NOTE 2- this list is in no specific order.

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-TEMPLATE |
| Title | Do not modify this particular table, this serves as a template, use the one below. |
| Description | NOTE 3- include a brief overview followed by a description about the problem, its importance to IMT-2020 networks and ITU, highlight any specific research or industry problem under consideration. |
| Challenge Track | NOTE 4- include a brief note on why it belongs in this track |
| Evaluation criteria | NOTE 5- this should include the expected submission format e.g. video, comma separated value (CSV) file, etc.NOTE 6- this should include any currently available benchmarks. e.g. accuracy. |
| Data source | NOTE 7- e.g. description of private data which may be available only under certain conditions to certain participants, pointers to open data, pointers to simulated data. |
| Resources | NOTE 7- e.g. simulators, APIs, lab setups, tools, algorithms, add a link in clause 2. |
| Any controls or restrictions | NOTE 8- e.g. this problem statement is open only to students or academia, data is under export control, employees of XYZ corporation cannot participate in this problem statement, any other rules applicable for this problem, specific IPR conditions, etc. |
| Specification/Paper reference | NOTE 9- e.g. arxiv link, ITU-T link to specifications, etc. |
| Contact | NOTE 10- email id or social media contact of the person who can answer questions about this problem statement. |

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-001 |
| Title | 5G+AI+AR |
| Description | **Background**: Remote collaboration has been an important tool to fight the recent COVID-19 outbreak. Effectiveness of such tools could be augmented using the support of AR/VR over IMT-2020 networks. Similar applications of AR/VR over IMT-2020 are emerging in sports, medicine, public welfare, socializing, and entertainment. eMBB specifications of 5G NR can address the needs of rich media needs of AR/VR. Device ecosystem is maturing with examples like Google Glass and Microsoft HoloLens. Infrastructure support with edge computing is already standardised. However certain specific areas needs to be further addressed using AI/ML.**Problems**: <<This requires further work>> Mobile AR/VR applications require low-latency to overcome motion sickness and alignment problems of head movements. Predictive content management and rendering could be a studied under this challenge. Mobility when combined with coverage or interference can lead to connectivity problems for AR/VR applications, which are especially sensitive even to short interruptions. Line of sight requirements when using certain frequency bands can add to this problem. An environment based inference on mobility (indoor and outdoor) could benefit AR/VR experience by end-users as well as adaptive options for application developers.<<TBD: add more>> |
| Challenge Track | Vertical-track (invite participant to make solutions for 5G, AI and AR application in vertical industries) |
| Evaluation criteria | Solution, criteria hasn’t been determined |
| Data source | Training data from existing AR/VR testbeds over IMT-2020 networks, with feedback on connection, quality, responsiveness to head movements, and time-aligned network data. |
| Resources | AR IDE (we are negotiating with partner), SDK which can plugin intelligent agents, simulators like [Unity]. |
| Any controls or restrictions | This problem statement is open to all participants. |
| Specification/Paper reference | [1], [2], [3], [4] from Appendix I. |
| Contact | liutf24@chinaunicom.cn; Tel +86 15652955883; wechat: yudajiangshan |

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-002 |
| Title | Topology optimization based on transmission network data |
| Description | The existing network topology planning does not fully consider the future growth of network traffic, and faces the problem of uneven utilization of loop capacity. Therefore, the existing network topology need to be optimized. By restructuring the sites on the unbalanced loops to achieve the global network fine-grained expansion and to increase the capacity utilization efficiency. So we seek topology optimization solutions for balanced loop utilization.The transmission network data will reflect the transmission network topology, the network's traffic matrix and the network capacity utilization.The task is network topology optimization by using the transmission network data. The evaluation system is the network capacity utilization. The specfic evaluation system will be provided with the detailed data. |
| Challenge Track | Secure-track |
| Evaluation criteria | According to the test set, the prediction result should be saved in a csv file and followed the required format.We will evaluate the result by the network capacity utilization under specfic network conditions. |
|
| Data source | Traning data and test data are all from specfic transmission network area, including the transmission network topology, the network's traffic matrix and the network capacity utilization. |
| Resources | No |
| Any controls or restrictions | Data is under export control and employees of partners cannot participate in this problem  |
| Specification/Paper reference | No |
| Contact | zhulinyj@chinamobile.com  |

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-003 |
| Title | Out of Service(OOS) Alarm Prediction of 4/5G Network Base Station |
| Description | At present, the operation and maintenance of 4/5G BS(base station) follow a passive pattern, repairing orders will not be generated until the out of service(OOS) fault occurs. Once the BS is out of service, users will not be able to connect to the wirelss network, and their regular communication will be affected. In general, there are some secondary alarms before the major alarm(OOS alarm). Therefore, in this challenge, the participants are expected to train an AI model using historical alarm datas with labels of major ones(Perhaps combinated with performance data if its acquirable). By excavating the corelation between alarms, one may use the secondary alarms to predict the probability of the important alarm happenning in a future period, so that the operation and maintenance personnel can solve the fault in advance and avoid network deterioration. Due to the similar operation and maintenance mode of 4G/5G network, after the large scale commercial use of 5G network, the AI model can be smoothly transfered as a pre-trained model. |
| Challenge Track | Secure-trackThis Challenge are aimed to provide applications in 4/5G network operation and maintenance using AI/ML.This track will allow the participants using network data from operators.In this Challenge, participants will design and implement AI algorithms that can help in an end-to end implementation of AI/ML model deployment in a real network.  |
| Evaluation criteria | Submit a comma separated value (CSV) file. The content includes whether the given base station will have an out of service alarm in the next 24 hours (or other period). The accuracy of the current prediction model has reached 78% |
| Data source | 4/5G network fault alarm data from China Mobile (possibly including performance data) |
| Resources | None |
| Any controls or restrictions | This Challenge is open to all participants. |
| Specification/Paper reference | None |
| Contact | jiazihan@cmdi.chinamobile.comTel +86 13810024426 |

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-004 |
| Title | Radio signal coverage analysis and prediction based on UE measurment report |
| Description | Multiple frequency bands are usually deployed in the commercial network to increase the network coverage and capacity. With the increasing number of bands, inter-frequency measurements by UEs may cause amount of signalling overhead and cost huge UE power consumption and severely impact on running service by the data interruption for inter-frequency measurement gap. It takes too long time for UE to choose the proper cell to reside in. This will degrade the network performance and UE experience. So quick inter-frequency measurement is desired. One way to obtain the coverage information of UEs' radio signal quickly is to divide the cell into the grids by serving cell’s and neighbouring cell’s radio signal levels, then locate the UE’s grid and perceive UE’s coverage information based on statistical analysis or directly predict the inter-frequency measurement based on the intra-frequency measurement, which can largely reduce the numbers of UE inter-frequency measurement and benefit for mobility based handover, load balancing, dual connection and carrier aggregation. |
| Challenge Track | Secure-track |
| Evaluation criteria |  Solution, criteria hasn’t been determined |
| Data source | Training data from commercial LTE network with feedback on UE MR data including RSRP,RSRQ,Earfcn,PCI of serving cell and neiboring cells. |
| Resources | No |
| Any controls or restrictions | This problem statement is open to all participants. |
| Specification/Paper reference | No |
| Contact | xieyuxuan@chinamobile.com |

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-005 |
| Title | 5G+AI+Intelligent APP Recognition |
| Description | Background: With the commercialization of 5G, users have higher requirements for service quality. The network needs to recognize ongoing services, conduct differentiated resource scheduling and ensure service guarantee according to the type of service. APP recognition based on manual rules, the maintenance of the rule library is difficult and the cycle is long which consumes a lot of manpower costs, can not meet the needs of rapid business updates. It is necessary to use AI technology to achieve intelligent APP recognition. According to the provided mobile Internet traffic data and APP annotation data, a classification model is constructed to recognize the APP to which the data traffic belongs. The model can be used to support online APP recognition. The effect evaluation method is to predict the code stream of the test set, output the APP type, and evaluate the accuracy of the recognition result.Problem: In the preliminary study, our team has adopted AI technology to recognize types of business flows and achieved some results.It is hoped that the accuracy of the algorithm model will be further improved through the competition, and the expected model accuracy is increased to 95%, so as to improve the application effect of the existing network. |
| Challenge Track | Secure-track |
| Evaluation criteria | Every team needs to predict each sample in the given test data which will be saved in csv files, and then submit csv files. Evaluation criteria are the same as the classification problem. |
|
| Data source | Training data from Mobile Internet traffic capture packet data and APP annotation data. |
| Resources | No |
| Any controls or restrictions | This problem statement is open to all participants. |
| Specification/Paper reference | No |
| Contact | yangmengjia@chinamobile.com,+86 13811206654 |

# 2. Resources

NOTE 1- the structure of the list below is intentionally kept simple for our partners to easily add or change it. The structure is as below:

<<type of resource: 1-line description, link, contact>>

NOTE 2- this list is in no specific order.

[RayMobTime] Data set: Raymobtime is a collection of ray-tracing datasets for wireless communications. <https://www.lasse.ufpa.br/raymobtime/>, aldebaro@ufpa.br

[CUBE-AI] ML marketplace: It is an open source network AI platform developed by China Unicom Network Technology Research Institute, which integrates AI model development, model sharing. <https://github.com/cube-ai/cubeai> , liutf24@chinaunicom.cn

[Adlik] Toolkit: an end-to-end optimizing framework for deep learning models. <https://github.com/Adlik/Adlik> , yuan.liya@zte.com.cn

[KNOW] Challenge platform: a data challenge platform which lists several challenges and competitions. <https://knowledgepit.ml/>

[SE-CAID] Data sets: An open AI research and innovation platform for networks and digital infrastructures for industries, SMEs and academia to share a broad range of telecom data and AI models. <https://se-caid.org/>

[AIIA] Challenge: past competition, led by AIIA in China <https://cloud.tencent.com/developer/contest/AIIA-Unicom>

<http://aiiaorg.cn/AIDC/2019AIDC/index.html>

<https://mp.weixin.qq.com/s?__biz=MzU0MTEwNjg1OA==&mid=2247487451&idx=1&sn=cb4370e9fa9d7f827dc632c79fe41d2d&chksm=fb2fb81ecc583108221592c69fdea3eb226da933859514dbd9fb8c15288c6fcb392c65399ddc&mpshare=1&scene=1&srcid=&sharer_sharetime=1575542631509&sharer_shareid=75fb4d5f665341fa1dafcbc554417e75&key=67a2c7aa29623c33d72ba777f7853d102e6f4db8ac8b23733613e267ce0dae54ca817de36bde651b3cf32c3a0daf055c432e46c3b8f43b088f60edcdef801a54201eea05d0de9051201391ee19fd326f&ascene=1&uin=MjEzNjY3NDQ5Mw%3D%3D&devicetype=Windows+7&version=62070141&lang=en&exportkey=AoB%2BIuWyreUPRCOzxdLg0q0%3D&pass_ticket=fCmC%2FiTFfXlmGxvOLq%2BdVPRElGBj59sZO2eVMyeABxg07Ve7tOfmRWTtKc1rmCRV>

[DuReader] Challenge: past competition, includes data sets, including the largest Chinese public domain reading comprehension dataset, DuReader <https://www.kesci.com/home/competition/5ad56e667238515d80b53704>

[IUDX] Data and challenge: a research project for an open source data exchange software platform, <https://www.iudx.org.in/>

[PUDX] Past challenge, Datathon  to develop innovative solutions based on India Urban Data Exchange ([IUDX](https://www.iudx.org.in/)), <https://cps.iisc.ac.in/pudx/>

[TI-bigdata] Data: a large dataset of 30+ kinds of data (mobile, weather, energy, etc. from Telcom Italia big data challenge. <http://theodi.fbk.eu/openbigdata/>

[TI-phone] Data: The Mobile phone activity dataset is a part of the Telecom Italia Big Data Challenge 2014. <https://www.kaggle.com/ijfezika/mobile-phone-activity-exploratory-analysis>

[MDC] Data: Mobile Data Challenge (MDC) Dataset,  restricted to non-profit organizations, <https://www.idiap.ch/dataset/mdc> (you need to make a request to get a copy)

[MIRAGE] Data: MIRAGE-2019 is a human-generated dataset for mobile traffic analysis with associated ground-truth, <http://traffic.comics.unina.it/mirage/>

[Urban-Air] Data: An air quality dataset that could be useful for verticals <https://www.microsoft.com/en-us/research/project/urban-air/>

[UCR] Data: UCR STAR is built to serve the geospatial community and facilitate the finding of public geospatial datasets to use in research and development. <http://star.cs.ucr.edu/>

[NYU] Data: NYU Metropolitan Mobile Bandwidth Trace, a.k.a. NYU-METS, is a LTE mobile bandwidth dataset that were measured in New York City metropolitian area; <https://github.com/NYU-METS/Main>

[Omnet] Data: Challenge and dataset from comes from Omnet++ network simulator, contains several topologies and thousands of labeled routings, traffic matrices with the corresponding per-flow performance (delay, jitter and losses). <https://bnn.upc.edu/challenge2020>

[GNN] Data: data sets for Unveiling the potential of GNN for network modeling and optimization in SDN. This data set can be divided in two components: (i) the data sets used to train the delay/jitter RoutNet models and (ii) the delay/jitter RouteNet models already trained <https://github.com/knowledgedefinednetworking/Unveiling-the-potential-of-GNN-for-network-modeling-and-optimization-in-SDN/tree/master/datasets>

[Unity] <https://github.com/Unity-Technologies/ml-agents/>

**Appendix I: Academic papers of interest**

[1] ` "Very Long Term Field of View Prediction for 360-degree Video Streaming", Chenge Li, Weixi Zhang, Yong Liu, and Yao Wang, 2019 IEEE Conference on Multimedia Information Processing and Retrieval.

[2] "A Two-Tier System for On-Demand Streaming of 360 Degree Video Over Dynamic Networks", Liyang Sun, Fanyi Duanmu, Yong Liu, Yao Wang, Hang Shi, Yinghua Ye, and David Dai, IEEE Journal on Emerging and Selected Topics in Circuits and Systems (March 2019 )

[3] “Multi-path Multi-tier 360-degree Video Streaming in 5G Networks”, Liyang Sun, Fanyi Duanmu, Yong Liu, Yao Wang, Hang Shi, Yinghua Ye, and David Dai, in the Proceedings of ACM Multimedia Systems 2018 Conference (MMSys 2018),

[4] “Prioritized Buffer Control in Two-tier 360 Video Streaming”, Fanyi Duanmu, Eymen Kurdoglu, S. Amir Hosseini, Yong Liu and Yao Wang, in the Proceedings of ACM SIGCOMM Workshop on Virtual Reality and Augmented Reality Network, August 2017;

\_\_\_\_\_\_\_\_\_\_\_\_\_