|  |  |  |
| --- | --- | --- |
|  | INTERNATIONAL TELECOMMUNICATION UNION**TELECOMMUNICATIONSTANDARDIZATION SECTOR**STUDY PERIOD 2017-2020 | **Focus Group on Machine Learning for Future Networks including 5G** |
| **ML5G-I-237** |
| **Original: English** |
| **Question(s):** | N/A | 9th meeting, July 2020 |
| **INPUT DOCUMENT** |
| **Source:** | FG ML5G |
| **Title:** | A compilation of problem statements and resources for ITU Global Challenge on AI/ML in 5G networks (formerly ML5G-I-223) |
| **Contact:** | Xie YuxuanChina MobileP.R.China | Email: xieyuxuan@chinamobile.com  |
| **Contact:** | Jia ZihanChina MobileP.R.China | Tel: +86 13810024426 Email: jiazihan@cmdi.chinamobile.com  |
| **Contact:** | Zhu LinChina MobileP.R.China | Email: zhulinyj@chinamobile.com  |
| **Contact:** | Mostafa EssaVodafone | Email: mostafa.Essa@vodafone.com  |
| **Contact:** | AbdAllah Mahmoud-EissaVodafone, Egypt | Email: AbdAllah.Mahmoud-Eissa@vodafone.com  |
| **Contact:** | Ai MingCICTP.R.China | Email: aiming@catt.cn |
| **Contact:** | Francesc WilhelmiUPF, Spain | Tel: +34 93 5422906Email: francisco.wilhelmi@upf.edu |
| **Contact:** | Aldebaro Klautau UFPABrazil | Tel: +55 91 3201-7181Email: aldebaro@ufpa.br  |
| **Contact:** | Tengfei LiuChina UnicomP.R.China | Tel: + 86 15652955883Fax: +010 68799999Email: liutf24@chinaunicom.cn  |
| **Contact:** | Jiaxin Wei China UnicomP.R.China | Tel: + 86 13126813179Fax: +010 68799999Email: weijx29@chinaunicom.cn  |
| **Contact:** | José Suárez-VarelaBNN-UPCSpain | Email: jsuarezv@ac.upc.edu |
| **Contact:** | Albert Cabellos-AparicioBNN-UPCSpain | Email: acabello@ac.upc.edu  |
| **Contact:** | Pere Barlet-RosBNN-UPCSpain | Email: pbarlet@ac.upc.edu  |
| **Contact:** | Seongbok BaikKT | E-mail: s.baik@kt.com  |
| **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 AI Challenge] ITU AI/ML in 5G Challenge website <https://www.itu.int/en/ITU-T/AI/challenge/2020/Pages/default.aspx>

[[ITU AI/ML Primer​](https://www.itu.int/en/ITU-T/AI/challenge/2020/Documents/ITU%20ML5G%20Global%20Challenge_proposal_v23a.docx)] [ITU AI/ML 5G Challenge: A Primer​](https://www.itu.int/en/ITU-T/AI/challenge/2020/Documents/ITU%20ML5G%20Global%20Challenge_proposal_v23a.docx) (13th March,2020)

[ITU AI/ML Summary] [ITU AI/ML 5G Challenge: Summary Slides](https://www.itu.int/en/ITU-T/AI/challenge/2020/Documents/ITU%20Challenge%20Summary-17-March%202020.pptx)(17th March,2020)

# 1. Introduction

[[ITU AI/ML Primer​](https://www.itu.int/en/ITU-T/AI/challenge/2020/Documents/ITU%20ML5G%20Global%20Challenge_proposal_v23a.docx)] 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 | TBD (Zhejiang Division) |
| Description |  |
| Challenge Track |  |
| Evaluation criteria |  |
| Data source |  |
| Resources |  |
| Any controls or restrictions |  |
| Specification/Paper reference |  |
| Contact |  |

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-002 |
| Title | User churn prediction(Shanghai Division) |
| Description | **Background**: With the arrival of the era of mobile Internet + artificial intelligence, Internet giants have occupied the forefront of AI in the era of AI and IoT. Operators need to think deeply about how to exert their professional advantages, accelerate cross-industry integration, and increase industry value.**Problems**: The number portability prediction is based on the operator's data, with big data + AI as the starting point, to predict the possibility of the user porting the number to other operators, thereby improving the accuracy of the operator to retain users.**Submiting:**Competitors need to submit two parts in the preliminary competition: one is to submit the algorithm model and analysis results (submitted in .csv format); the other is the annotated core code and explanatory documents (separately attached files, submitted in .pdf file format). Finally, all the files are packaged and compressed into a zip file for submission. |
| Challenge Track | Network-track |
| Evaluation criteria | TP(True Positive): 1 for True and 1 for prediction; FN(False Negative): true 0, predicted 1; FP(False Positive): true is 1, prediction is 0; TN(True Negative): 0 for True and 0 for prediction.According to the following formula, the scores of the contestants are calculated. According to the accuracy rate (formula 1) and recall rate (formula 2), f1-score (formula 3) is calculated. Finally, all the contestants are ranked according to f1-score.P = TP/(TP+FP) （1）R = TP/(TP+FN) （2）F1-score = 2\*P\*R/（P+R） （3） |
|
| Data source | Provide users' basic attributes, Internet access, phone calls, SMS and other behavior data, as well as the monthly bill of mobile network XDR data, cang data, resource data (wireless community, broadband community), configuration data (terminal configuration, etc.).In order to protect users' privacy and data security, the data has been sampled and desensitized. There are null values or junk data in the data table, and the participants need to handle it by themselves. |
| 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 | liutf24@chinaunicom.cn; Tel +86 15652955883; wechat: yudajiangshanweijx29@chinaunicom.cn |

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-003 |
| Title | Alarm and prevention for public health emergency based on telecom data(Beijing Division) |
| Description | **Background:** In recent years, the worldwide outbreak of Covid-19, Ebola, MERS and SARS posed grievous and global affects on human beings and seriously challenged WHO as well as the health department of many countries. Apart from the effort of health department, modern informational technologies and data can help in health emergencies. In this problem statement, competitors should use the tracking data of telecom users’ geographical movements and DPI information, technologies including machine learning and big data, to propose comprehensive solutions, product developing or advises on infrastructure for serious public health emergencies. All these works can be considered on aspects of epidemic surveillance, spread monitoring, precise prevention, resource allocation, effect evaluation for health incidents.**Problems:** This topic focuses on epidemic surveillance, spread monitoring, precise prevention, resource allocation, effect evaluation by telecom users’ tracking data and DPI information while the outbreak of Covid-19. Participants should propose related products or solutions by using the data, resources and developing environment provided by the competition organizer. If participants use the data from anywhere else, it should be taken in account that the accessibility and scalability of the data. **Submiting:**Participants do mining and modeling based on the data provided by the organizer, and yield corresponding solutions or products. The final submission should cover the following aspects:Detailed introduction of the solutions or products;The source code of mining and modeling, as well as the completed zip file of applications; The model and explanations;The product prototype, website or APP (optional, plus). |
| Challenge Track | Vertical-track |
| Evaluation criteria | **Full marks 100****Problem analysis (10 marks):** Whether it has a good understanding of the core of the topic and key elements which affect the final results.Application prospects: Whether there are demands, prospects and potentials for the proposed solutions or products.**Solutions (25 marks)**: Whether the solutions are reasonable and feasible, and meet the demand.The use of data: Whether the data provided by organizer is fully used in an effective way.Innovation: Whether the works are innovative and different from matured solutions in current industries, and whether it performs better.**Implementation (25 marks)**: Whether the solutions or products can be implemented or used as a clear pattern in realistic situation, and have prospects in future.Technical foundation: Whether it has a solid technical foundation to carry out the solutions or products, and improve them in future.Social effect: Whether it has social effects and the ability to avoid the risk of data breach.**Completion (40 marks)**: Whether the work is complete within the allotted time and schedule, and meet all the requirements. |
| Data source | The tracking data including geographically locations and time (directional offset) of sampled users (encrypted) in a city, the app use data and the ownership information.Detailed description: The format, parameter, field of the data, etc. More details can be found in the zip file of the topic. |
| Resources | None |
| Any controls or restrictions | Data is under export control and employees of partners cannot participate in this problem |
| Specification/Paper reference | None |
| Contact | liutf24@chinaunicom.cn; Tel +86 15652955883; wechat: yudajiangshanweijx29@chinaunicom.cn |

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-004 |
| Title | Network topology optimization |
| Description | The existing network topology planning does not fully consider the future growth of network traffic, and faces the problem of uneven utilization of link capacity. Therefore, the existing network topology need to be optimized. By restructuring the sites on the unbalanced links to achieve the global network fine-grained expansion and to increase the capacity utilization efficiency. So we seek topology optimization solutions for balanced link capacity utilization. The network information data will reflect the network topology, the network's traffic matrix and the network capacity utilization. The task is network topology optimization by using the network information data. The evaluation system is the network capacity utilization. The specific evaluation system will be provided with the detailed data. |
| Challenge Track | Network-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 specifically by the network capacity utilization balancing value and the ratio of link capacity utilization within the optimization target range. Among them, the smaller the capacity utilization balancing value, the larger the ratio of link capacity utilization within the optimization target range, the better the algorithm optimization result. The capacity utilization balancing E value is the variance of the link capacity utilization values of all links in the network. |
|
| Data source | Training data and test data are all from specific network area, including the network topology, the network's traffic matrix and the network capacity utilization. The network topology data includes the network element number, network element type, network element latitude and longitude, and the connection relationship between network elements. The network element information data includes network element node number, network element type, network element capacity value, network element latitude and longitude, and the daily hourly network's traffic matrix value, etc. |
| 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-005 |
| 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 wireless 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 data with labels of major ones. By excavating the relationship between alarms, one may use the secondary alarms to predict the probability of the important alarm happening 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 transferred as a pre-trained model. |
| Challenge Track | Network-track |
| 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.The data is fault alarm data of several months, including alarm start time, alarm name, base station name, base station ID, vendor name, city, etc. |
| Resources | None |
| Any controls or restrictions | Data is under export control and employees of partners cannot participate in this problem |
| Specification/Paper reference | None |
| Contact | jiazihan@cmdi.chinamobile.comTel +86 13810024426 |

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-006 |
| Title | Radio signal coverage analysis and prediction based on UE measurement 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 neighboring 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-007 |
| Title | UE Moblity Analytics in 5G network |
| Description | **Background:** In 3GPP, the NWDAF is the AI related network function (NF), which collects data from NFs, OAM and to feedback around 9 categories analytics to requested NFs (Please refer to TS23.288). Within the category “UE related analytics”, the UE mobility analytics or predications could be utilized by NFs, e.g. AMF, SMF, EIR for some purposes, such as mobility management parameter adjustment, detect UE been stolen, and etc.The detailed content of “UE Mobility information” collected from 5G network, the output analytics including “UE mobility statics” and “UE mobility predictions” could be found in TS23.288.**Problem:** However, how 5GC NFs utilize aforementioned output analytics in real 5G network would not be standardized in 3GPP now, and has been leave to NF implementation (but how?), and the benefits of such implementation for real network is still not clear.It is very important to find out “how” and demonstrate the benefits. This would help operator to deploy the NWDAF related and make real 5G networks more intelligent. |
| Challenge Track | Operator and vendor -track ? |
| Evaluation criteria | 1. Every team needs to provide output analytics including “UE mobility statics” and “UE mobility predictions”, according to the input “UE Mobility information”.
2. Every team needs to provide the description of their implementation on how to use the output analytics, and corresponding benefits.
 |
|
| Data source | 1. Every team itself needs to provide the “UE Mobility information” from real 5G network or find equivalent from 4G network.
2. Are there operators could possibly kindly provide the “UE Mobility information” all the teams?
 |
| Resources | TBD |
| Any controls or restrictions | This problem statement is open to all participants. |
| Specification/Paper reference | TS23.288 |
| Contact | aiming@catt.cn |

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-008 |
| Title | Intelligent spectrum management for future networks |
| Description | **Background:** Future networks are heterogeneous, e,g, Multi-RAT (5G, 4G, licensed, unlicensed, fixed, mobile), Multiple platforms (edge cloud vs. centralized cloud, VNF vs. PNF, Multiple levels/domains (Access Network vs. Core, network slices with varied KPI demands, various management and orchestration layers). Also there several potential data sources e.g. (Peer-to-peer networks, NF, applications, UEs.**Problem:** In thatcontext**,** spectrum management for future networks is challenging. There is an expectation from end-customer for coexistence and mobility across different networks (see above).Interference management and seamless user experience across different frequency bands used by the network is expected.Power management in the basestation and UE is a challenge in future networks with multi-bands.Current methods for spectrum management has the following disadvantages:* The existing techniques for spectrum management are technology specific, partly standardised + vendor-specific algorithms implemented in scheduler.
* Intra-RAT (radio access technology) standards available (e.g. X2)
* Operator control is lesser, mainly driven by vendor differentiation (scheduler and resource mangament algorithms).
* Suited to less-dynamic network conditions of 4G than to future networks of 5G and beyond.

In future networks, we would like schemes which:1. exploit the upcoming open interfaces and data in RAN and CN
2. flexible to optimize the on-demand spectrum access in tomorrow’s networks.

In this context, the spectrum management for future networks is proposed to be:* Data-driven: Use data from different parts of the network (based on VF contribution to ITU FG ML5G, Supplement 55 to Y.3170 series)
* Federated: Cross-domain exchange of data for ML (based on ITU Y.3172, 3174)
* Self-x: Adaptive, Distributed ML, decisions at the edge (to reduce latency, communication overhead).
* Level 5 intelligent: demand mapping, based on plug-in models from operator ML marketplaces (based on ITU Y.3173).

Advantages of this approach:* Data driven, at the same time, reduces latency, communication overhead
* Based on operator KPIs (e.g. interference reduction)
* Standard (ITU-based) architecture and interfaces for interoperability
* Take advantage of best ML mechanisms - Plugin models from researchers

Challenge problem statement:* Given a set of network bands for various types of future networks, implement intelligent dynamic spectrum management for future networks including IMT-2020 based on data from multiple domains in the network.
* Emphasises self-x strategy of VF.
* Implements pluggable intelligence (AI models).
* An optimal solution should have a model which reduces interference between various networks, uses standard interfaces (e.g. ITU), enables optimal operator KPIs and imposes minimal communication overhead.

[More details, including the VF sandbox setup (lab), will be shared later with interested participants] |
| Challenge Track | Network track (private VF data) |
| Evaluation criteria | In a testbed chosen by VF, shortlisted models and solutions will be evaluated by:1. Comparison with existing benchmarks for operator KPIs
2. Accuracy of models
3. Latency
4. Amount of communication overhead for the model
 |
|
| Data source | Private data from VF (available only to VF approved candidates) |
| Resources | Lab setsup / simulator (available only to VF approved candidates)VF Sandbox will be setup using data and tools from VF. It will be accessible only to selected participants nominated by VF. Data will be hosted in a place of choice by VF. Only the data and tools relevant to the VF problem statement will be hosted in the VF Sandbox. Regular meeting and monitoring of participants having access to the VF Sandbox will be done by ITU. |
| Any controls or restrictions | Data privacy: No data should be moved from the region.Private data from VF (available only to VF approved candidates) |
| Specification/Paper reference | ITU-T Y.3172 and Y.3174 |
| Contact | AbdAllah.Mahmoud-Eissa@vodafone.com |

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-009 |
| Title | ML5G-PHY: Machine Learning Applied to the Physical Layer of Millimeter-Wave MIMO Systems |
| Description | The increasing complexity of configuring cellular networks suggests that machine learning (ML) can effectively improve 5G and future networks. One of the technologies for applications such as vehicular systems is millimeter (mmWave) MIMO, which enables fast exchange of data. A main challenge is that mmWave, as initially envisioned for this application, requires the pointing of narrow beams at both the transmitter and receiver. Taking into account extra information such as out-of-band measurements and vehicles positions can reduce the time needed to find the best beam pair. Beam training is part of standards such as IEEE 802.11ad and 5G, and has also been extensively studied in the context of wireless personal and local area networks. Hence, one of the tasks focuses on beam-selection. Another task is channel estimation, which is challenging due to mobility, strong attenuation in mmWave and other issues. This challenge uses datasets obtained with the Raymobtime methodology. The data consists of millimeter wave (mmWave) multiple-input multiple-output (MIMO) channels, paired with data from sensors such as LIDAR. |
| Challenge Track | Network-track, as the challenge consists of use cases related to signalling or management. |
| Evaluation criteria | Top-K classification for beam selection and normalized mean squared error for channel estimation |
| Data source | Raymobtime datasets - <https://www.lasse.ufpa.br/raymobtime/> |
| Resources | None |
| Any controls or restrictions | This Challenge is open to all participants. |
| Specification/Paper reference | [7] 5G MIMO Data for Machine Learning: Application to Beam-Selection using Deep Learning, 2018 - http://ita.ucsd.edu/workshop/18/files/paper/paper\_3313.pdf[8] MmWave Vehicular Beam Training with Situational Awareness by Machine Learning, 2018 - https://ieeexplore.ieee.org/document/8644288[9] LIDAR Data for Deep Learning-Based mmWave Beam-Selection, 2019 - https://ieeexplore.ieee.org/document/8642397[10] MIMO Channel Estimation with Non-Ideal ADCS: Deep Learning Versus GAMP, 2019 - https://ieeexplore.ieee.org/document/8918799 |
| Contact | Aldebaro Klautau – aldebaro@ufpa.br. Tel: +55 91 3201-7181 |

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-010 |
| Title | Improving the capacity of IEEE 802.11 WLANs through Machine Learning |
| Description | The usage of Machine Learning (ML) is foreseen to be a key enabler to address the challenges podes by future wireless networks. In IEEE 802.11 Wireless Local Area Networks (WLANs), the major challenges will be the user’s density and lack of coordination, which, given the current channel allocation mechanisms, lead to sub-optimal performance. One potential solution is the application of Dynamic Channel Bonding (DCB), whereby an Overlapping Basic Service Set (OBSS) adapts the spectrum to be used so that their performance is maximized. Nevertheless, due to the complexity of massively crowded deployments, choosing the appropriate channel width is not trivial. Moreover, increasing the channel width entails a trade-off between the link capacity and the quality of the link (using more bandwidth entails a lower received signal strength and leads to a higher contention). To address the abovementioned challenges, we propose using Deep Learning (DL) to predict the performance that will be obtained in an OBSS by using different channel bonding strategies. |
| Challenge Track | Network-track (students) |
| Evaluation criteria | Participants should provide a .csv file containing the predicted performance of each BSS (columns) in the different test deployments (rows).The evaluation of the proposed algorithms will be based on the average squared-root error obtained from all the predictions compared to the actual result in each type of deployment. |
| Data source | To be provided |
| Resources | The IEEE 802.11ax-oriented Komondor simulator [3] has been used to generate both training and test datasets. |
| Any controls or restrictions | This Challenge is open to all student participants. |
| Specification/Paper reference | [11] Barrachina-Muñoz, S., Wilhelmi, F., & Bellalta, B. (2019). Dynamic channel bonding in spatially distributed high-density WLANs. *IEEE Transactions on Mobile Computing*.[12] Barrachina-Muñoz, S., Wilhelmi, F., & Bellalta, B. (2019). To overlap or not to overlap: Enabling channel bonding in high-density WLANs. *Computer Networks*, *152*, 40-53.[13] Barrachina-Muñoz, S., Wilhelmi, F., Selinis, I., & Bellalta, B. (2019, April). Komondor: a wireless network simulator for next-generation high-density WLANs. In *2019 Wireless Days (WD)* (pp. 1-8). IEEE. |
| Contact | Francesc Wilhelmi, francisco.wilhelmi@upf.edu (+34 93 5422906) |

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-011 |
| Title |  |
| Description |  |
| Challenge Track | Network-track  |
| Evaluation criteria |  |
| Data source | TBD |
| Resources |  |
| Any controls or restrictions |  |
| Specification/Paper reference |  |
| Contact | stephan.blicker@telekom.de |

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-012 |
| Title | Graph Neural Networking Challenge 2020 |
| Description | Network modelling is essential to construct optimization tools for networking. For instance, an accurate network model enables to predict the resulting performance (e.g., delay, jitter, loss) and helps finding the configuration maximizes the network performance according to a target policy. Currently, network models are either based on packet-level simulators or analytic models. The former are very costly computationally while the latter are fast but not accurate. In this context, Machine Learning (ML) arises as a promising solution to build accurate network models able to operate in real time. Recently, Graph Neural Networks (GNN) have shown a strong potential to be integrated into commercial products for network control and management. Early works using GNN have demonstrated an unprecedented capability to learn from different network characteristics that are fundamentally represented as graphs, such as the topology, the routing configuration, or the traffic that flows along a series of nodes in the network. In contrast to previous ML-based solutions, GNN enables to produce accurate predictions even in networks unseen during the training phase. Nowadays, GNN is a hot topic in the ML field and, as such, we are witnessing significant efforts to leverage its potential in many different fields (e.g., chemistry, physics, social networks). In the networking field, the application of GNN is gaining increasing attention and, as it becomes more mature, is expected to have a major impact in the networking industry.**Problem statement:**The goal of this challenge is to create a neural network model that estimates performance metrics given a network snapshot. Specifically, this model must predict the resulting per-source-destination performance (delay, jitter, loss) given a network topology, a routing configuration, and a source-destination traffic matrix.scheme_challengeAs a baseline, we provide RouteNet [5][6], a GNN architecture recently proposed to model network performance. Participants are encouraged to submit their own neural network architecture or update RouteNet. |
| Challenge Track | Network-track (design, train and test a neural network model for a networking use case) |
| Evaluation criteria | By means of an unlabelled dataset. Participants must label this dataset with their neural network models and send the results in CSV format. For the evaluation we will use a score that combines the Mean Absolute Error (MAE) and the Mean Relative Error (MRE) of the per-source-destination performance predictions produced by the candidate solutions. The MAE indicates the absolute error of the predictions with respect to the ground-truth labels, while the MRE measures the relative distance between them. |
| Data source | Datasets are generated using a discrete packet-accurate network simulator (OMNet++). The dataset contains samples simulated in several topologies and includes hundreds of routing configurations and traffic matrices. The data is divided in three different sets for training, validation and test. The validation and test datasets contain samples with similar distributions.You can find more details about the datasets at <https://bnn.upc.edu/challenge2020>.https://lh5.googleusercontent.com/yKEcuyII1DL9LKN3YCxQLGgxawig2VDX6AX8Kf1YgrnpYHzcP6Flap22tD9IoRYfOEQN7UPrAGfQSB1-zEzAfIDrfl24Gx36lkCB4vAbZtEoZFT4JxBQ=w572 |
| Resources | - Paper, source code and tutorial of RouteNet, a reference GNN model that can be used as a starting point for the challenge [5][6] - User-oriented Python API to easily read and process the datasets - Mailing list for questions and comments about the challenge [[Challenge-KDN mailing list](https://mail.knowledgedefinednetworking.org/cgi-bin/mailman/listinfo/challenge-kdn)]- Website with a more detailed description of the challenge and the resources provided (<https://bnn.upc.edu/challenge2020>) |
| Any controls or restrictions | The following rules must be satisfied to participate in this challenge:* The solutions must be fundamentally based on neural networks
* The proposed solution cannot use network simulation tools.
* Solutions must be trained only with samples included in the training dataset we provide. It is not allowed to use additional data obtained from other datasets or synthetically generated.
* The challenge is open to all participants except members of the organizing team and the research group “Barcelona Neural Networking Center-UPC”.
* It is allowed to participate in teams. All the team members should be announced at the beginning and will be considered to have an equal contribution.

Final submissions must include the code of the neural network solution proposed, the neural network model already trained, and a brief document describing the proposed solution (1-2 pages).Important notice: In the challenge, you may use any existing neural network architecture (e.g., the RouteNet implementation we provide). However, it has to be trained from scratch and it must be clearly cited in the solution description. In the case of RouteNet it should be cited as it is in [5]. |
| Specification/Paper reference | [5] Rusek, K., Suárez-Varela, J., Mestres, A., Barlet-Ros, P., & Cabellos-Aparicio, A, “Unveiling the potential of Graph Neural Networks for network modeling and optimization in SDN,” In Proceedings of ACM SOSR, pp. 140-151, 2019. [6] Source code an tutorial of RouteNet [6] |
| Contact | José Suárez-Varela – jsuarezv@ac.upc.edu |

|  |  |
| --- | --- |
| Id | ITU-ML5G-PS-013 |
| Title | DL-based RCA (Root Cause Analysis) |
| Description | **Background*** It is important for carriers to operate their complex network stably.
* The stable operation includes locating and identifying the root cause by looking at symptoms when some faults occur on their networks.
* Vendors provide a variety of indicators (logical syslogs, or physical LED indicators) to indicate the status of the equipment when they release their equipment.
* When constructing a network with a small number of equipment, it is easy to find the root cause and reasoning the core problems.
* By making this reasoning process into a rule set, it is possible to automate the whole inference logic, only under the condition that the size of the network is moderately large
* However, in a very large and complex environment of the network, the rule-based inference method shows the very limited performance.
* Especially in the 5G network, stability and speed are emphasized to provide the new 5G services. Various brand-new 5G equipment, which is physical and also virtual, is deployed, resulting in the number of management points increased exponentially.
* In this situation, the introduction of DL can be of great help to the operators, because it is almost impossible to set up the rules to pin-point the root causes in such a complex environment.

**Motivation*** For the introduction of DL technology, it is essential to collect the training data
* However, it is almost impossible to acquire the fault situation data much enough for training, because the fault situations do not occur frequently in nature
* A promising alternative is to build a test-bed that simulates 5G network to simulate various fault situations and collect data
* Using this collected data, a DL model for RCA can be developed
* This DL model is developed in the form of a pre-trained model through learning the characteristics of network equipment on a test-bed
* In actual application, the characteristics of operator's network can be fine-tuned to quickly increase accuracy and be applied to the site

**Objectives*** By implementing the following two items, the DL-based RCA system can be implemented for complex 5G network
* 1) Implement a Test-bed simulating 5G network (ML5G test-bed)
	+ Composed of communication equipment common to telecommunications operators providing 5G services
	+ Interworking with DB by adding data collection function at the major management points in the simulated network
	+ Configured to enable the fault scenario settings and labeled data collection according to research needs
* 2) Development of DL model optimized for RCA
	+ General DL model for RCA should be pre-trained on this test-bed
	+ The pre-trained DL model will be fine-tuned to be applied to the commercial environment
	+ Once constructed, the simulation test-bed can be used for various purposes other than RCA
 |
| Challenge Track | Network-track  |
| Evaluation criteria |  |
| Data source | TBD |
| Resources |  |
| Any controls or restrictions |  |
| Specification/Paper reference |  |
| Contact | Seongbok Baik s.baik@kt.com  |

# 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/>

[ETSI ARF] ETSI GS ARF 003 V1.1.1 (2020-03) Augmented Reality Framework (ARF); AR framework architecture <https://www.etsi.org/deliver/etsi_gs/ARF/001_099/003/01.01.01_60/gs_ARF003v010101p.pdf>

**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;

[5] Rusek, K., Suárez-Varela, J., Mestres, A., Barlet-Ros, P., & Cabellos-Aparicio, A, “Unveiling the potential of Graph Neural Networks for network modeling and optimization in SDN,” In Proceedings of ACM SOSR, pp. 140-151, 2019. [[ACM SOSR](https://dl.acm.org/doi/abs/10.1145/3314148.3314357)] [[arXiv](https://arxiv.org/pdf/1901.08113.pdf)]

[6] Source code and tutorial of RouteNet. (URL: <https://github.com/knowledgedefinednetworking/demo-routenet>)

[7] 5G MIMO Data for Machine Learning: Application to Beam-Selection using Deep Learning, 2018 - http://ita.ucsd.edu/workshop/18/files/paper/paper\_3313.pdf

[8] MmWave Vehicular Beam Training with Situational Awareness by Machine Learning, 2018 - https://ieeexplore.ieee.org/document/8644288

[9] LIDAR Data for Deep Learning-Based mmWave Beam-Selection, 2019 - https://ieeexplore.ieee.org/document/8642397

[10] MIMO Channel Estimation with Non-Ideal ADCS: Deep Learning Versus GAMP, 2019 - <https://ieeexplore.ieee.org/document/8918799>

[11] Barrachina-Muñoz, S., Wilhelmi, F., & Bellalta, B. (2019). Dynamic channel bonding in spatially distributed high-density WLANs. *IEEE Transactions on Mobile Computing*.

[12] Barrachina-Muñoz, S., Wilhelmi, F., & Bellalta, B. (2019). To overlap or not to overlap: Enabling channel bonding in high-density WLANs. *Computer Networks*, *152*, 40-53.

[13] Barrachina-Muñoz, S., Wilhelmi, F., Selinis, I., & Bellalta, B. (2019, April). Komondor: a wireless network simulator for next-generation high-density WLANs. In *2019 Wireless Days (WD)* (pp. 1-8). IEEE.

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