IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) | 978-1-6654-0926-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/INFOCOMWKSHPS54753.2022.9798094

Demonstration of Policy-Induced Unsupervised Feature Selection in a 5G network

Jalil Taghia*, Farnaz Moradi*, Hannes Larsson*, Xiaoyu Lan*, Masoumeh Ebrahimi*‡, and Andreas Johnsson*†
* Ericsson Research, Sweden, Email: *{jalil.taghia, farnaz.moradi, hannes.larsson, xiaoyu.lan, andreas.a.johnsson}@ericsson.com*
† Uppsala University, Department of Information Technology, Sweden, Email: *andreas.johnsson@it.uu.se*
‡ KTH, Division of Electronics and Embedded Systems, Sweden, Email: *mebr@kth.se*

***Abstract*—A key enabler for integration of machine-learning models in network management is timely access to reliable data, in terms of features, which require pervasive measurement points throughout the network infrastructure. However, excessive measurements and monitoring is associated with network overhead. The demonstrator described in this paper shows key aspects of feature selection using a novel method based on unsupervised feature selection that provides a structured approach in incorporation of network-management domain knowledge in terms of policies. The demonstrator showcases the benefits of the approach in a 5G-mmWave network scenario where the model is trained to predict round-trip time as experienced by a user.**

***Index Terms*—Network management, feature selection, machine learning, 5G.**

I. Background and Concepts

A key enabler for machine-learning models for network management is timely access to reliable data for model training and real-time inference, thus requiring pervasive measurement points (MP) throughout the network infrastructure. Unfortunately, excessive data collection and transfer of data from MPs to its consumer come with overhead costs in terms of for example network utilization. Further, processing of many features increase the ML model complexity which can lead to additional compute resources requirements. Furthermore, the process of training machine learning models may also be negatively affected by an excessive amount of features, leading to a reduced model performance as well as increased model complexity imposing challenging requirements on data availability and compute power on nodes in the infrastructure that may not have such capabilities.

Feature selection, both supervised and unsupervised, have been extensively studied in the literature and can potentially mitigate some of the challenges mentioned above [1]. In our recent work we further enhance the concept of unsupervised feature selection [2] in networking, and argue that the effective inclusion of domain knowledge can provide improved and more generalizable feature sets. Domain knowledge can be seen in the form of policies that guide the learning towards selection of an improved feature set. We proposed and evaluated a method called Policy-Induced Concrete Autoencoder (PI-CAE) which takes a set of policies dictating the features that must be monitored in a network, named must-have features. It then selects a set of latent features that are complementary to the policies. The selected features are then communicated

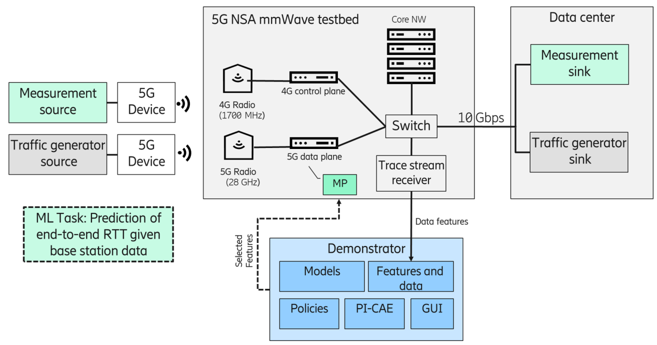

Fig. 1. 5G-mmWave testbed and demonstrator interaction. Dashed line indicating control of MP is not implemented.

for configuration of the MPs across a network infrastructure resulting in lowered monitoring overhead. The must-have features can for example be selected based on experiences acquired by a network operations center and its personnel.

Fundamental to this demonstration is the concept of unsupervised feature selection using operator-specified policies guiding the selection of additional latent features across a network infrastructure. We select features given a budget on monitored features to arrive at a feature set that is capable of modeling RTT as experienced by a user equipment (UE) using baseband features extracted from a 5G-mmWave network. The resulting feature sets achieve reduced overhead in terms of monitoring costs, while at the same time maintaining the predictive capabilities of the ML models.

II. Scenarios and data traces

In the demonstration we focus on an experimental scenario using an in-house 5G-mmWave testbed, illustrated in Figure 1, generating a set of traces for further analysis. The equipment corresponds to a 5G non-standalone (NSA) system, where the control plane is served through a 4G LTE base station (eNB) and the user-plane is served through a 5G NR base station (gNB). The 4G LTE eNB operates on B3, 1800 MHz, with 5 MHz bandwidth. The 5G NR gNB operates on n257, 28 GHz, with 100 MHz bandwidth. The spectrum is time-shared between downlink (DL) and uplink (UL) using a 4:1 TDD pattern (*DDDSU* [3]). In this testbed there are two UEs,

TABLE I
TRACE CONFIGURATIONS AND SPECIFICATIONS FOR THE 5G TRACES.

Trace Name	Traffic Load	RTT Averaging Interval	Number of Tasks
NoTraffic-RTT1000	None	1000 (ms)	2
NoTraffic-RTT100	None	100 (ms)	2
ConstantTraffic-RTT1000	Constant Bit Rate	1000 (ms)	2
ConstantTraffic-RTT100	Constant Bit Rate	100 (ms)	2

TABLE II
SUMMARY OF THE ML TASKS FOR THE 5G TRACES.

Acronym	Description
Low Threshold	A low RTT threshold value
High Threshold	A high RTT threshold value

one for performance measurements (measurement source), and one for generation of traffic load (traffic generator source).

The traces contain RTT values as experienced by a UE in the network, and approximately 200 metrics and events (extracted from the MP in Figure 1) related to the analogue beamforming function, the UEs connected to the base station, and UL and DL events [4]–[6].

Two different experiments were performed for this demonstration, corresponding to *NoTraffic* and *ConstantTraffic* trace categories. In both experiments, the UEs were moving in a test area for a duration of 10 minutes. In one of the experiments, a constant bit rate traffic load (UL) was generated, while in the other experiment, no traffic load was added. The RTT was measured every 10ms using ICMP with a measurement packet size of 1400 bytes. Further, as a pre-processing step, the RTT and the base station metrics are averaged over different time intervals (e.g., every 100ms, 1000ms). In total, 636 features were generated from around 200 metrics based on different averaging intervals. The traces are summarized in Table I. The trace name is encoded according to traffic load (constant traffic/no traffic) and the averaging intervals (100ms/1000ms).

We consider scenarios where the data traces are used to predict RTT as experienced by the UE, based on traces from the 5G-mmWave base station. The prediction is formalized as a binary classification problem based on different service-level metrics (low, high) corresponding to RTT thresholds constituting service violations. The ML tasks using these traces are predicting service-level metrics. Table II summarizes the ML tasks considered in the experiments.

III. DEMONSTRATION

We demonstrate the PI-CAE method for prediction of RTT as experienced by a UE in a 5G network, where the policy (of features) can be interactively changed. The key components of the demonstrator are shown in Figure 1; containing a GUI, features and data, policies, an implementation of the method PI-CAE, and pre-computed models. Features and data have been pre-fetched from the *trace stream receiver* in the testbed. Further, we also indicate the concept of a MP configuration control channel from the demonstrator to the MP where the selected features would be communicated.

The main demo screen, illustrated in Figure 2 and corresponding to the *GUI* in Figure 1, is divided into three sections. In the

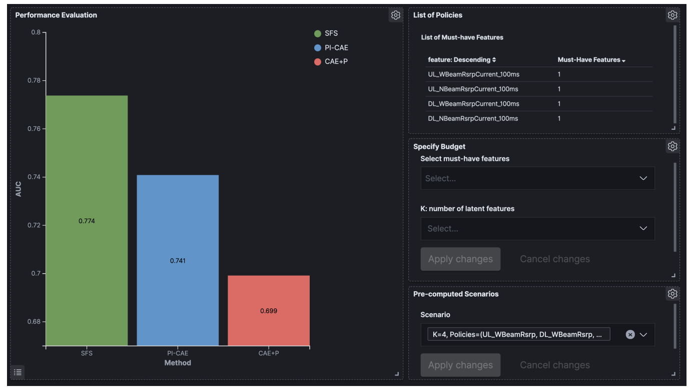

Fig. 2. Main PI-CAE demo screen.

upper right different policies can be chosen, that is selection of must-have features from the list of available features in the base station. Further, the number of latent features needs to be specified. Together, the total number of wanted features corresponds to the monitoring budget.

In the lower right a set of pre-trained ML models (with different feature sets) for RTT prediction is available for further investigation with respect to model performance. The RTT model is a classification task and thus the model performance is measured in terms of Area Under Curve (AUC). The model performance is compared with two other feature-selection approaches serving as baselines; (1) supervised feature selection which constitutes an empirical upper bound, and (2) unsupervised feature selection with the policy added afterwards. This is illustrated in the left part of Figure 2.

The method, PI-CAE is implemented in PyTorch, and the demonstrator is a Kibana dashboard which visualizes the results of the pre-trained models stored in an Elasticsearch database.

Additional demo screens will show further aspects of the PI-CAE method, however they are omitted from the paper due to space limitations.

ACKNOWLEDGMENT

This research has been supported by the Swedish Governmental Agency for Innovation Systems, VINNOVA, through projects ITEA3 AutoDC and Celtic ANIARA, and Swedish Foundation for Strategic Research, SSF Strategic Mobility.

REFERENCES

[1] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, pp. 1–45, 2017.

[2] J. Taghia, F. Moradi, H. Larsson, X. Lan, M. Ebrahimi, and A. Johnsson, "Policy-induced unsupervised feature selection: A networking case study," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022.

[3] 3GPP, "NR; Physical layer procedures for control," 3rd Generation Partnership Project (3GPP), Technical Specification (TS), 2021, version 16.5.0.

[4] ——, "NR; Physical layer procedures for data," 3rd Generation Partnership Project (3GPP), Technical Specification (TS), 2021, version 16.5.0.

[5] ——, "NR; Physical layer measurements," 3rd Generation Partnership Project (3GPP), Technical Specification (TS), 2021, version 16.5.0.

[6] ——, "NR; Medium Access Control (MAC) protocol specification ," 3rd Generation Partnership Project (3GPP), Technical Specification (TS), 2021, version 16.5.0.