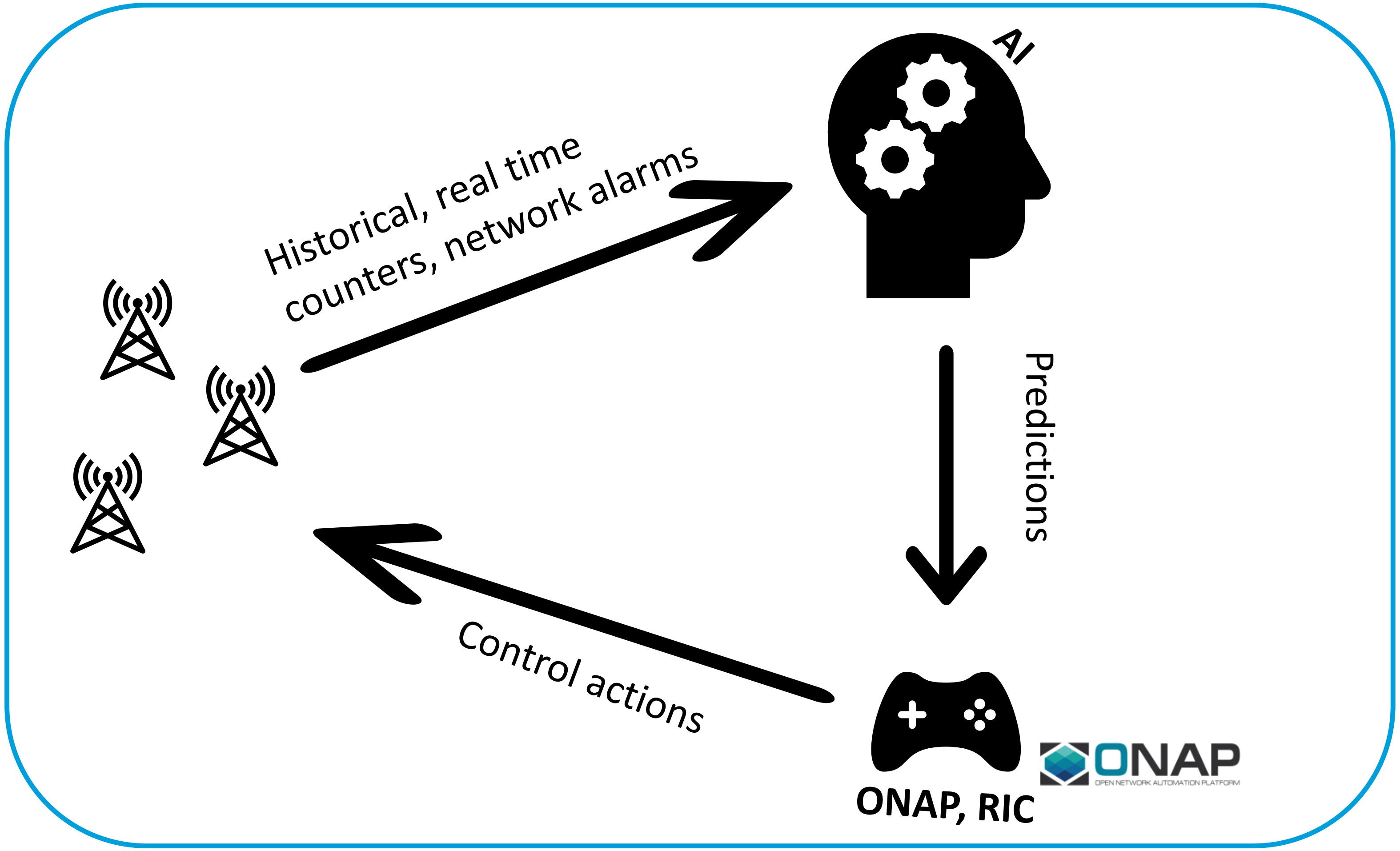
DeepAuto: A Hierarchical Deep Learning Framework for Real Time Prediction in Cellular Networks

Authors: Abhijeet Bhorkar, Ke Zhang, Jin Wang

October 1, 2019

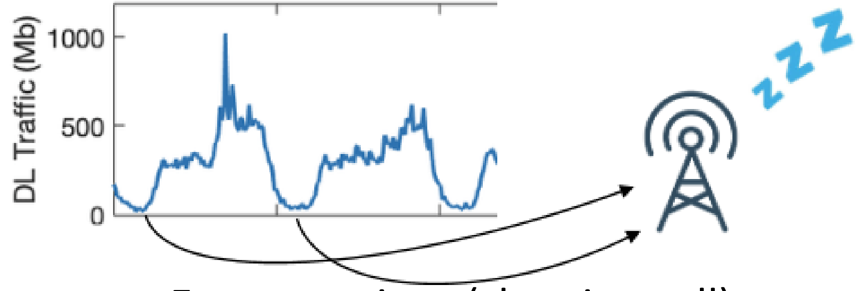
Mobility Network Automation

- Mobile Network optimization and automation is essential to improve network performance, reduce OPEX and keep up with demand, delivery & service.

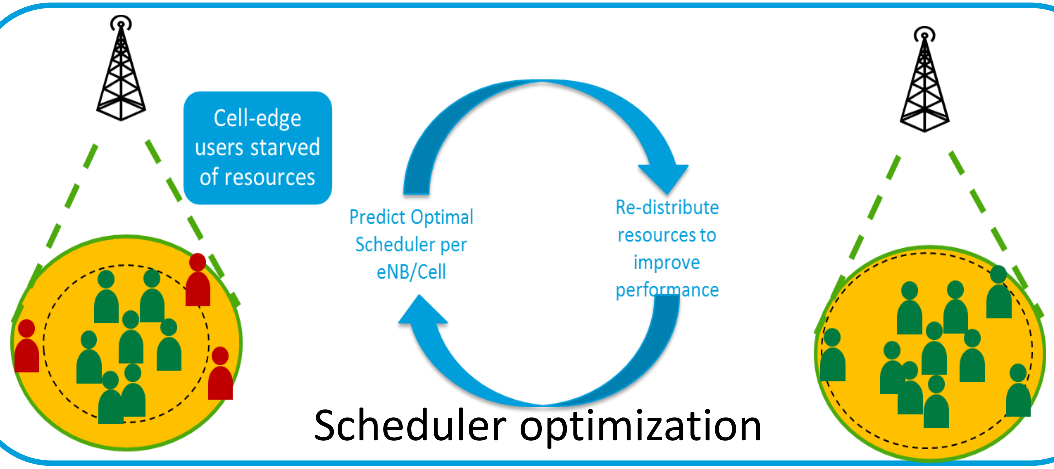


RAN Automation use cases

- Energy savings (sleeping cell)
- Load balancing
- Scheduler optimization
- Adaptive stream saver
- Traffic offload
- Handover optimization



Energy savings (sleeping cell)



Scheduler optimization

Adaptive stream saver

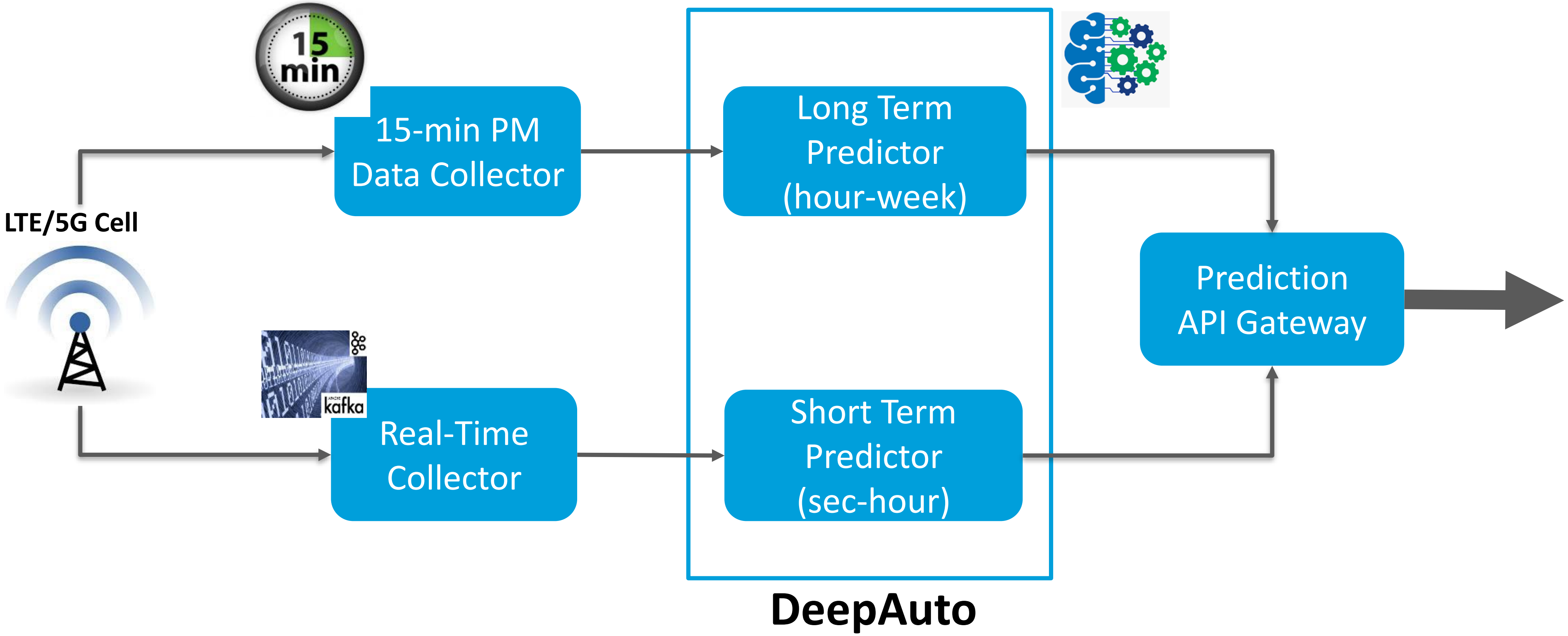
Goals

- Design a scalable reusable accurate* prediction framework to predict RAN KPIs
- Predict KPIs for different history and horizon combinations
- Capture multi-scale temporal and spatial information
- Framework should be able to predict real time
- Include impacts of network configuration changes and external influence

- More accurate estimate results in better adaptation decisions
- Xuan Kelvin et al, “Can Accurate Predictions Improve Video Streaming in Cellular Network”

DeepAuto: Architecture of Prediction System

- Coarse, longer delay data is used for mid term (hour-week) prediction.
- Finer, shorter delay data is used for short term (second-mins) prediction.

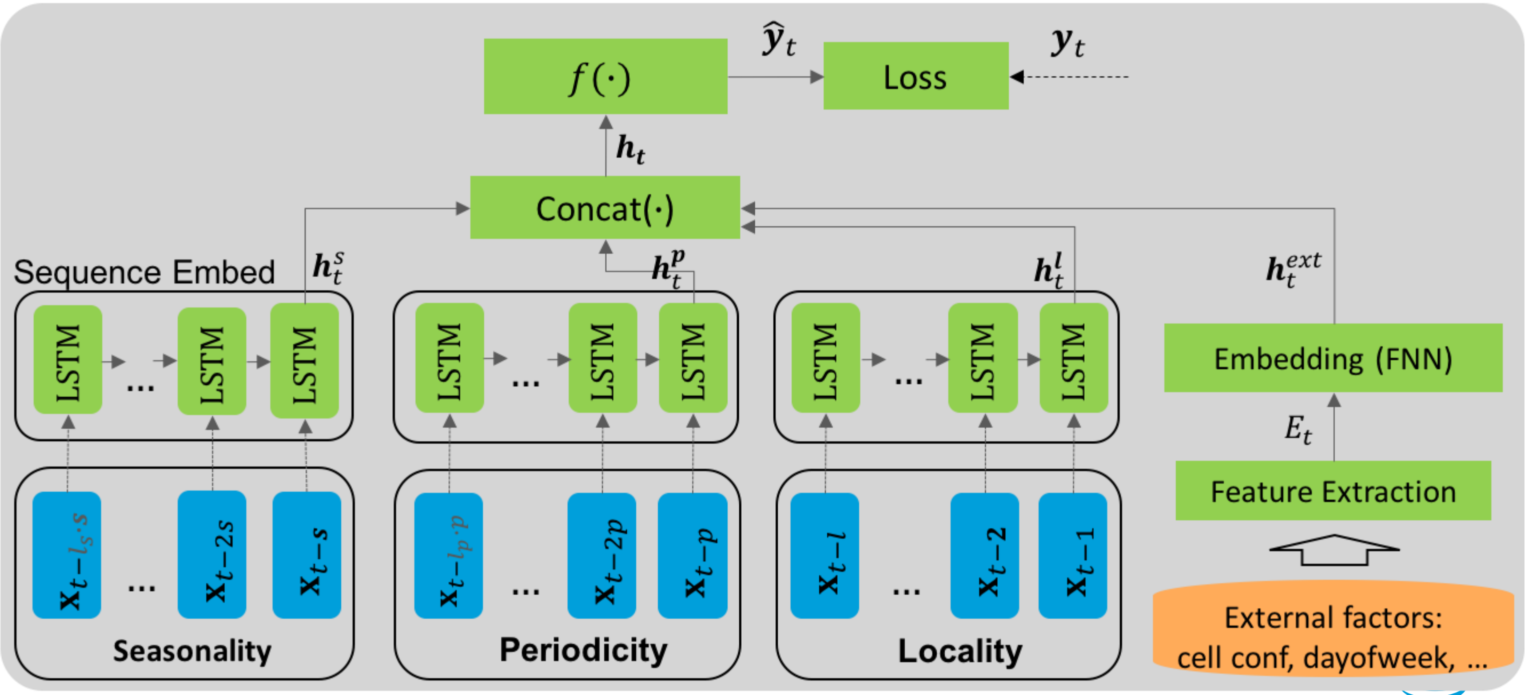


DeepAuto: Reusable prediction engine using deep learning

- The objective of reusable prediction engine is to provide *short-term (seconds to few mins) and long-term (hours, days, weeks)* forecasts of various cell level and UE level KPIs.
- The problem can be framed as time-series prediction and in general using Nonlinear Autoregressive with exogenous inputs (NARX) framework.

$$y_t = F(\underbrace{y_{t-1}, y_{t-2}, y_{t-3}, \ldots,}_{\text{Past values of same series}} \underbrace{u_t, u_{t-1}, u_{t-2}, u_{t-3}, \ldots}_{\text{Current and past values of exogeneous series}}) + \varepsilon_t$$

- We have designed Long Short-Term Memory (LSTM) based framework for the prediction tasks.
- LSTM frameworks have shown benefits over statistical frameworks such as Autoregressive integrated moving average (ARIMA) at the cost of increased complexity [1].
- Traditional machine learning models such as random forest models cannot capture long term dependencies.



Proposed LSTM based architecture for reusable analytics

[1] *LSTM network: a deep learning approach for short-term traffic forecast*

Evaluation of DeepAuto

- We illustrate our work using most important cell level KPIs including cell load prediction and radio channel quality prediction.
  - Key applications such as adaptive scheduler selection and cell load balancing will be enabled using cell load and channel quality KPI prediction
- **Cell load prediction:** The objective is to predict average *Cell PRB utilization in the next 1 min, 15 min and 1 hour* for each cell.
- **Radio channel quality prediction** : The objective is to predict *RSRQ distribution in the next 5 mins* for each cell.

Cell Load Prediction (1)

- Snapshot of the training data for cell load prediction

cell_name	timestamp	cell_prb_util_percent_dl	active_ue_dl	cell_bitrate_kbps_dl	ENODEB	EARFCN_DL	DL_CH_BANDWIDTH	MAX_TX_POWER	LATITUDE	LONGITUDE	ALTITUDE	MARKET_CLUSTER	MARKET	SW_VERSION
CAL00013_2A_1	5/22/2018 0:00	53.13	1.49	5438	CAL00013	2200	10000	490	32.7306694	-117.1596778	84.2	SAN DIEGO/LAS VEGAS	SAN DIEGO	CXP102051_27_R21F24
CAL00013_2A_1	5/22/2018 0:01	46.96	1.25	1153	CAL00013	2200	10000	490	32.7306694	-117.1596778	84.2	SAN DIEGO/LAS VEGAS	SAN DIEGO	CXP102051_27_R21F24
CAL00013_2A_1	5/22/2018 0:02	52.45	1.57	2900	CAL00013	2200	10000	490	32.7306694	-117.1596778	84.2	SAN DIEGO/LAS VEGAS	SAN DIEGO	CXP102051_27_R21F24
CAL00013_2A_1	5/22/2018 0:03	58.47	1.59	3846	CAL00013	2200	10000	490	32.7306694	-117.1596778	84.2	SAN DIEGO/LAS VEGAS	SAN DIEGO	CXP102051_27_R21F24
CAL00013_2A_1	5/22/2018 0:04	69.89	1.83	4394	CAL00013	2200	10000	490	32.7306694	-117.1596778	84.2	SAN DIEGO/LAS VEGAS	SAN DIEGO	CXP102051_27_R21F24

Cell PRB Utilization (target to be predicted)

Additional cell metrics

Cell information

External features:

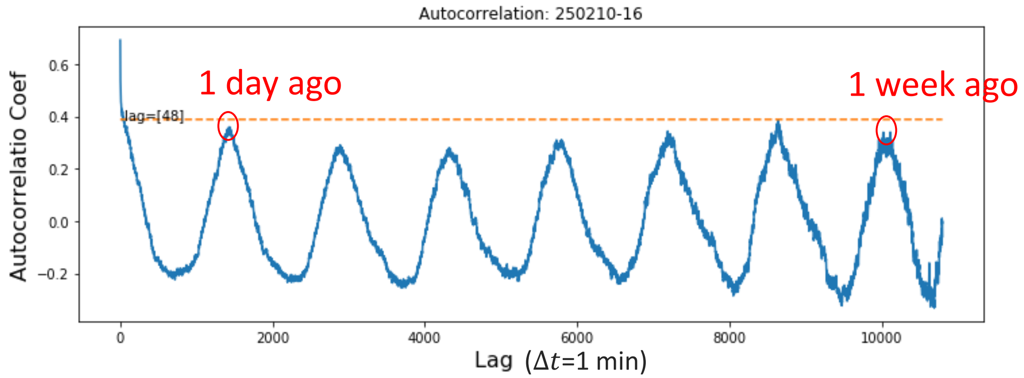
- Active UE count
- Cell throughput
- Avg UE throughput
- Cell configuration: *band, power, bandwidth, …*
- Day of week, Hour of day

Temporal features:

- Instant trend: $x(t-1), x(t-2), \dots, x(t-l)$
- Periodic pattern: $x(t-p), \dots, x(t-l_p \cdot p)$
- Seasonal pattern : $x(t-s), \dots, x(t-l_s \cdot s)$
- p = 1 day, s = 1 week

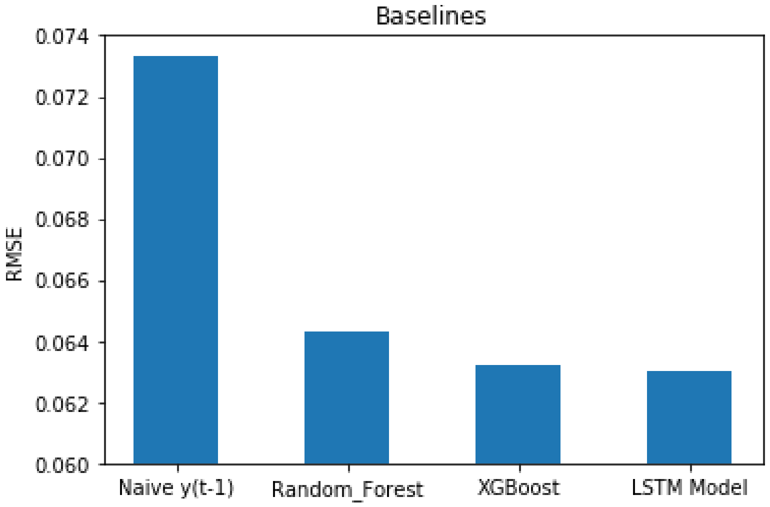
Feature Correlation Matrix

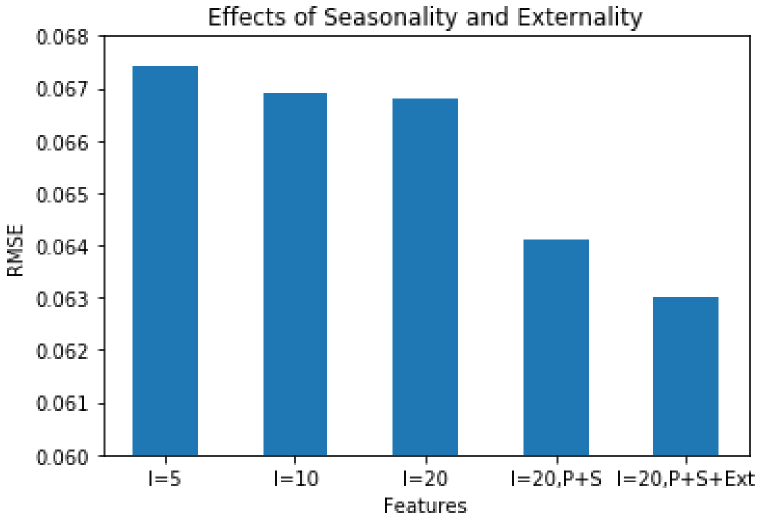
	ue_cnt	prb_util	cell_tput	ue_tput
ue_cnt	1.000000	0.695330	0.566613	-0.001132
prb_util	0.695330	1.000000	0.330548	-0.117647
cell_tput	0.566613	0.330548	1.000000	0.163847
ue_tput	-0.001132	-0.117647	0.163847	1.000000

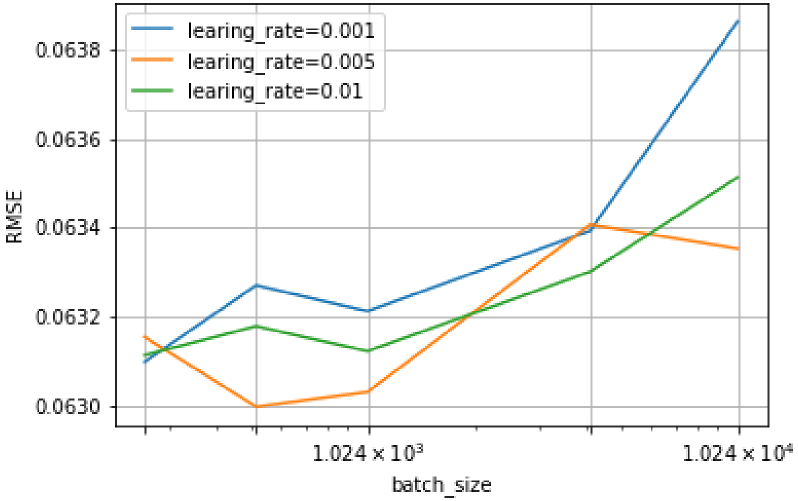


Cell Load Prediction (2): experiment and evaluation

Data Source	Δt	Time span	# instances	Train:val:test	RMSE	MAE	MAPE	MAPE (y>mean)	MAPE (y>mean+std)
PM Counter	15 mins	14 weeks	~ 13.7 M	5:1:1	0.063	0.0425	24.20%	15.30%	12.50%
CTR/STEM	1 min	8 days	~ 14.8 M	4:1:1	0.091	0.0604	35.92%	18.47%	13.73%







With customized loss function, e.g., $MSE(y-\hat{y}) = e^{-\alpha(1-y)} \cdot (y-\hat{y})^2$
the prediction for *heavy load* can be further improved

RF prediction (1)

- The reported RSRQ values are grouped by timestamp (at 5-mins interval) and cell id, and binned into 17 buckets.
- Snapshot of the data for prediction

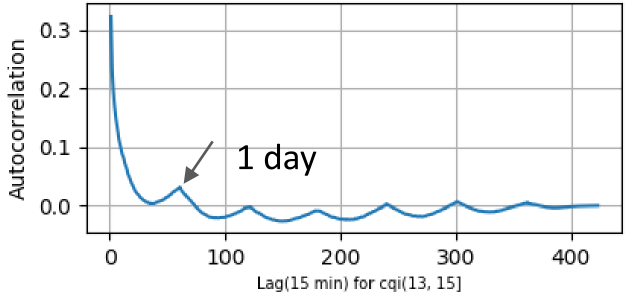
enb	gci	timestamp	(0, 1]	(1, 3]	(3, 5]	(5, 7]	(7, 9]	(9, 11]	(11, 13]	...	(15, 17]	(17, 19]	(19, 21]	(21, 23]	(23, 25]	(25, 27]	(27, 29]	(29, 31]	(31, 34]	measurements
ARL00140C	250140-15	2018-08-08 23:55:00	0.0000	0.0000	0.0000	0.3043	0.0435	0.0000	0.0870	...	0.0435	0.0870	0.2609	0.1304	0.0000	0.0000	0.0	0.0	0.0	23
ARL00140C	250140-15	2018-08-09 00:00:00	0.0181	0.0722	0.1512	0.1309	0.1061	0.1151	0.0722	...	0.0609	0.0813	0.0564	0.0293	0.0339	0.0045	0.0	0.0	0.0	443
ARL00140C	250140-15	2018-08-09 00:05:00	0.0125	0.0357	0.0784	0.1123	0.1176	0.1373	0.0784	...	0.1016	0.0927	0.0553	0.0285	0.0303	0.0267	0.0	0.0	0.0	561
ARL00140C	250140-15	2018-08-09 00:10:00	0.0000	0.0538	0.0846	0.0769	0.0308	0.1462	0.0923	...	0.1077	0.1077	0.1077	0.0308	0.0077	0.0231	0.0	0.0	0.0	130
ARL00140C	250140-15	2018-08-09 00:15:00	0.0093	0.0761	0.0798	0.1262	0.1466	0.1744	0.1020	...	0.0594	0.0649	0.0686	0.0186	0.0093	0.0000	0.0	0.0	0.0	539

RSRQ PDF :
% of RSRQ measurements in each bin (17 bins in total)

Temporal and external features:

- Histogram of RSRQ
- Day of week, Hour of day
- Minute of day

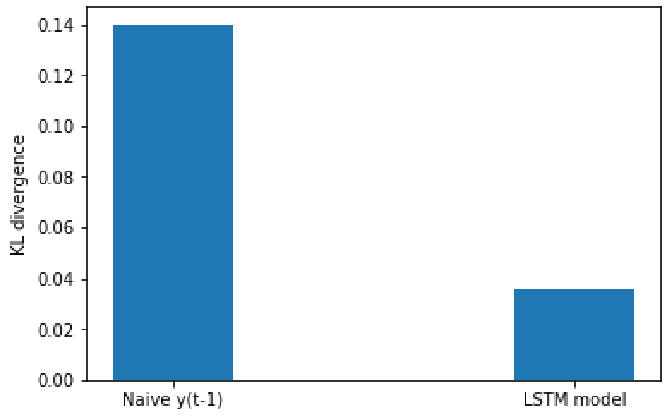
- Instant trend: $x(t-1), x(t-2), \ldots, x(t-l)$
- Periodic pattern: $x(t-p), \ldots, x(t-l_p \cdot p)$
- Seasonal pattern : $x(t-s), \ldots, x(t-l_s \cdot s)$
- l =25, p = 1 day, s = 0 week

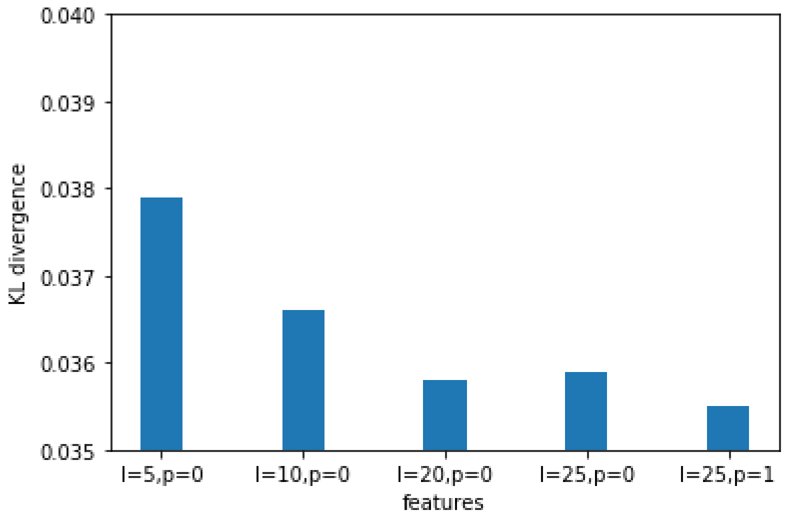


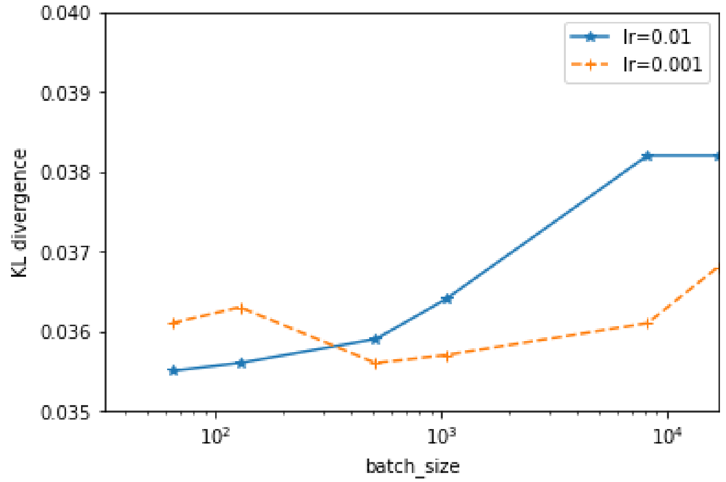
*note: correlation artifact due to non-continuous data

RF prediction (2): experiment and evaluation

- KL divergence is used as a loss metric for comparing true and predicted distribution.
  - $L = \frac{1}{n}\sum_i D_i(P||Q)$ where n is the number of LTE cells in the testing dataset and $D_i(\mathrm{P}||\mathrm{Q})$ is the KL divergence calculated for cell i

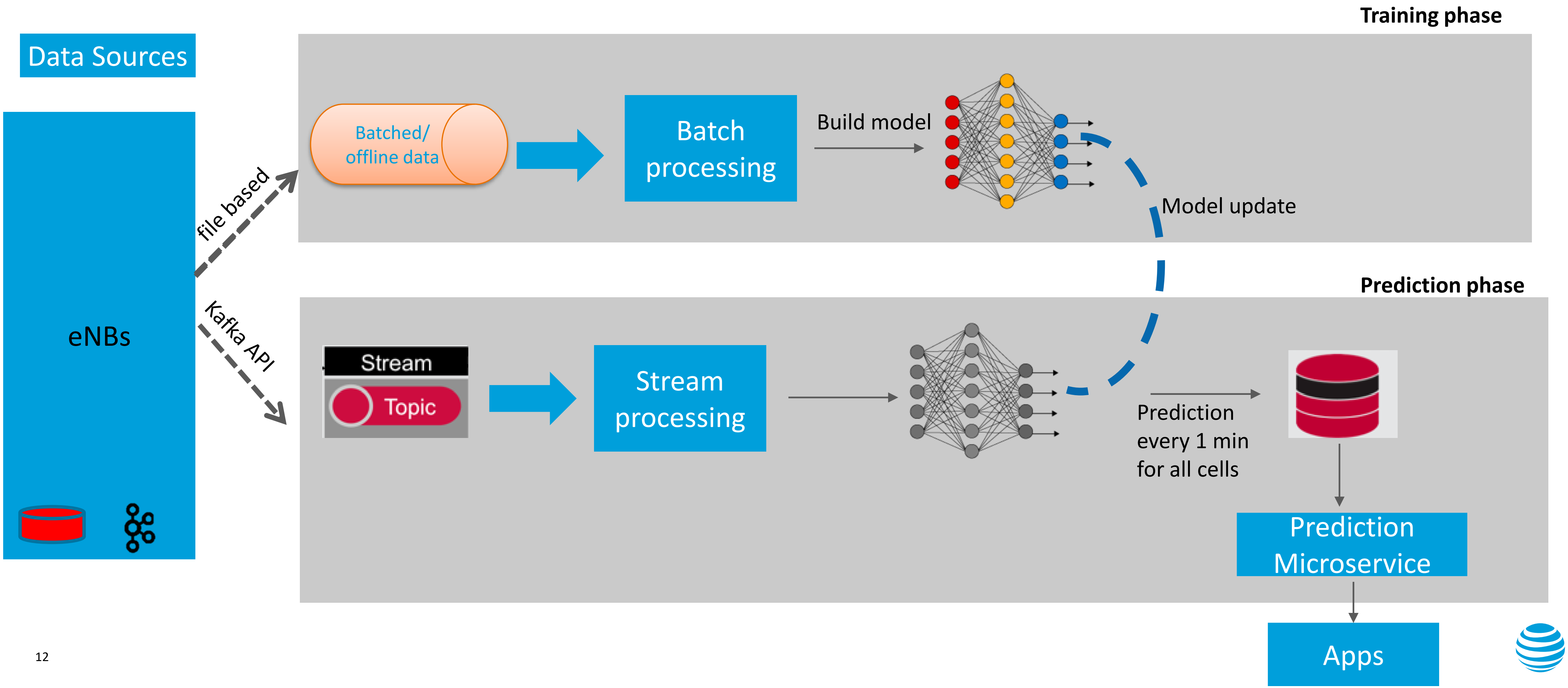




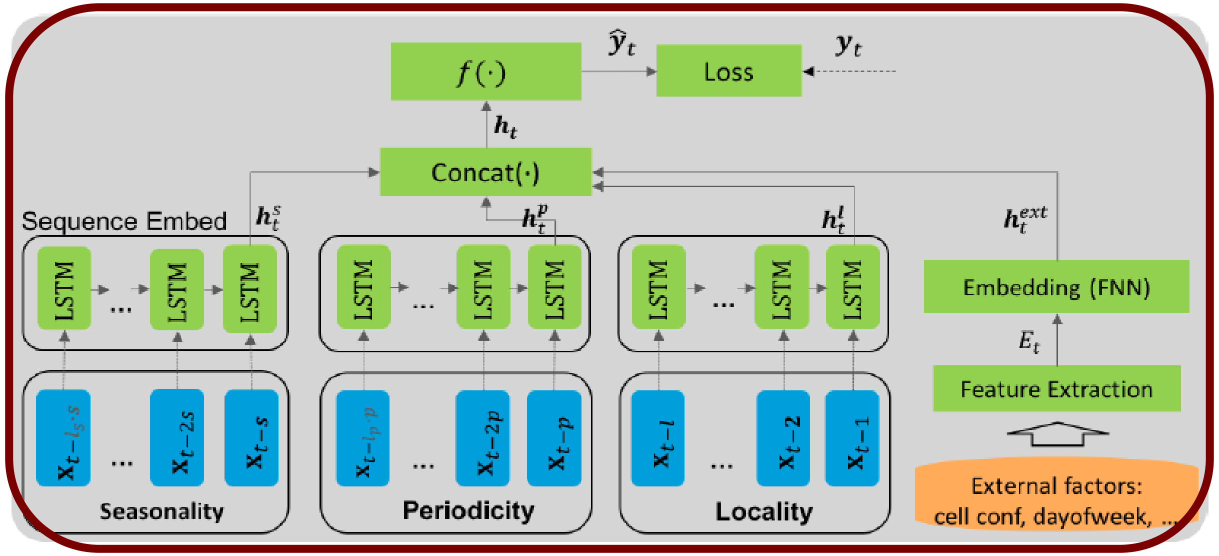


*Traditional statistical/machine learning methods seem unsuitable for this problem. Thus, baseline performance is not provided for other methods

Deployment


Training phase
Data Sources
eNBs
file based
Kafka API
Batched/
offline data
Batch
processing
Build model
Model update
Prediction phase
Stream
Topic
Stream
processing
Prediction
every 1 min
for all cells
Prediction
Microservice
Apps

Concluding remarks



Goal
Design scalable Reusable prediction system

DeepAuto
A deep learning framework to predict RAN KPI with various time horizons

Evaluation
Models show advantage in capturing complex relationships

- Accurate forecasting of RAN KPIs represents an essential part LTE/5G RAN automation. We provided an unified, efficient and effective traffic prediction architecture that predicts various RAN KPIs in real time.
- We presented the prediction model DeepAuto, hierarchical deep learning framework, that constructively captures spatial, temporal and external factors, as well as network configuration changes in a scalable manner.
- Machine learning methods reduce the prediction error by upto 75% compared to the naive method of using recent measurement.