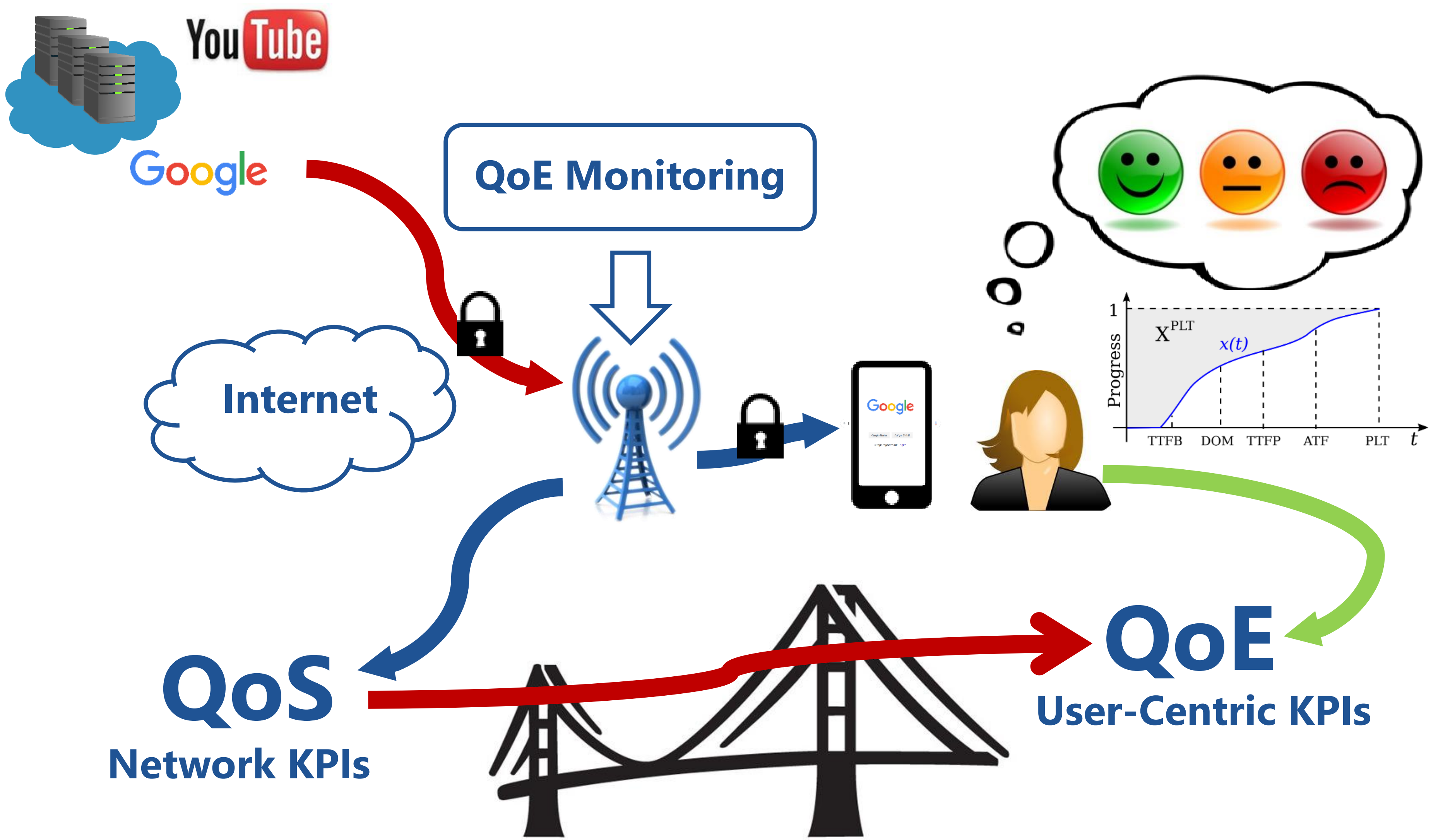
Institute of Computer Science
Chair of Communication Networks
Prof. Dr. Tobias Hoßfeld

Improving Web QoE Monitoring for Encrypted Network Traffic Through Time Series Modeling

Nikolas Wehner, Michael Seufert, Joshua Schüler, Sarah Wassermann, Pedro Casas, Tobias Hoßfeld

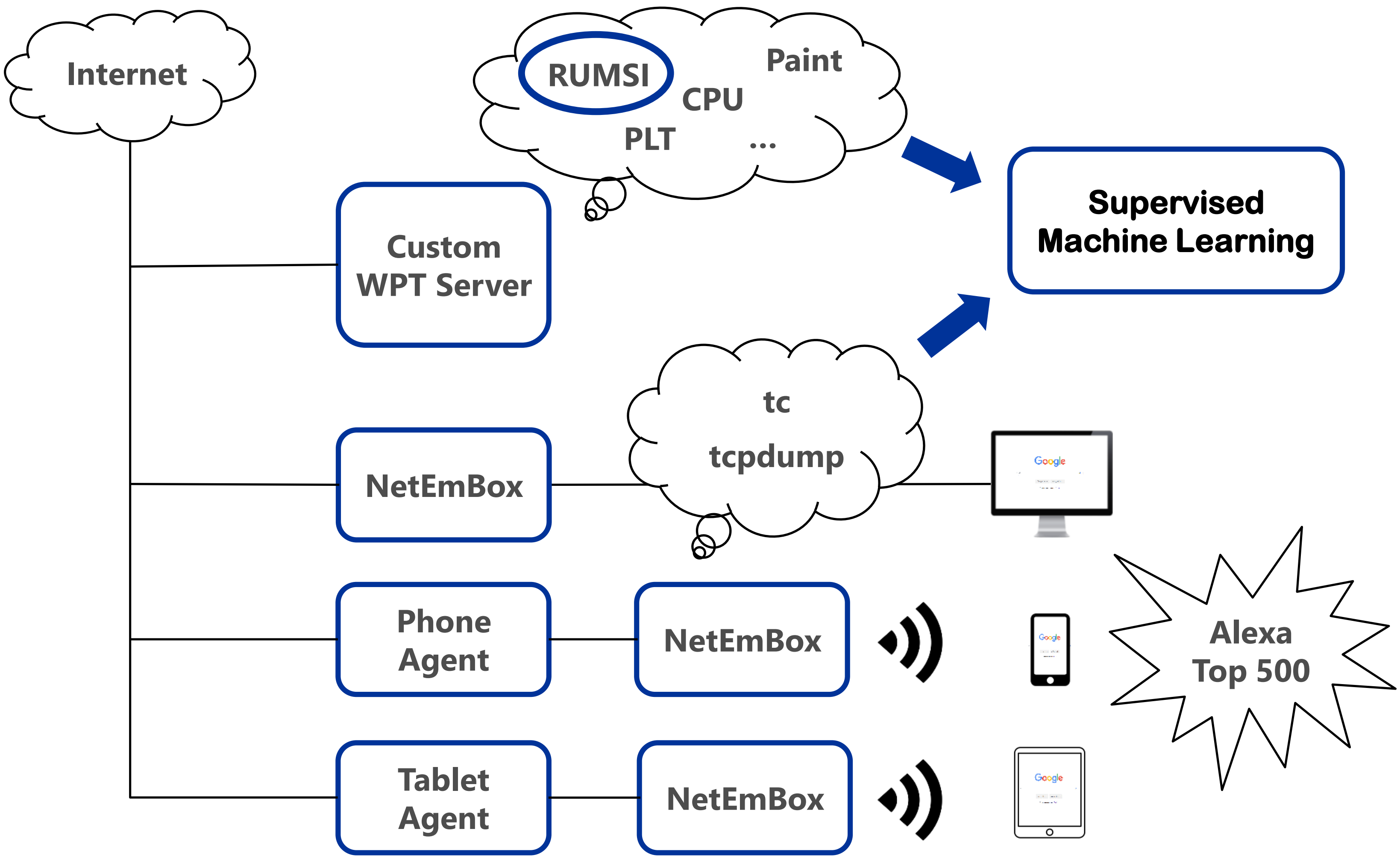
comnet.informatik.uni-wuerzburg.de

Motivation



YouTube
Google
QoE Monitoring
Internet
Google
1
Progress
X^PLT
x(t)
TTFB
DOM
TTFP
ATF
PLT
t
QoS
Network KPIs
QoE
User-Centric KPIs

Methodology



Considered Features for Machine Learning

- Real-time **temporal granularity**, e.g., **predictions in 100ms windows**

- Feature set includes
  - **Count-based** features, e.g., amount of transferred bytes
  - **Time-based** features, e.g., packet inter-arrival times
  - **Distribution-based** features
  - Features are computed for **total**, **downlink**, and **uplink** traffic each
  - Features are computed for **top three IP addresses** each
  - Features are computed for **varying subnets** of top three (24, 16, 8)

- **For every window** the following feature sets are computed
  - Features extracted from **current window (window features)**
  - Features extracted from **all previous windows (session features)**

  → **1979 features** in total

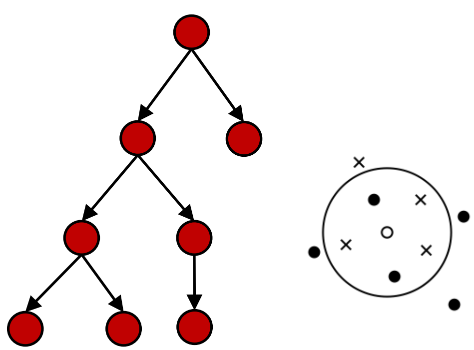
- Around **60000 web page loads** during November 2019
  (only a subset of all data)

Regression Models

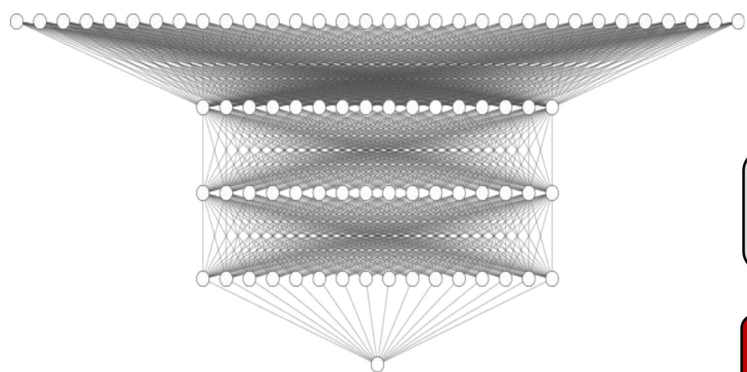
- Several regression models are tested
  - ML regression models with scikit-learn (CART, KNN, etc.)
  - Neural network regressor (DNN)
  - Recurrent neural network regressor (LSTM/GRU layers)

- Either session-based or time-series input features

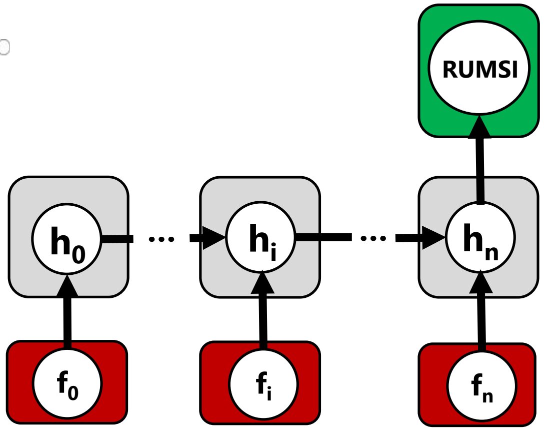
Scikit-Learn (CART, KNN)



DNN



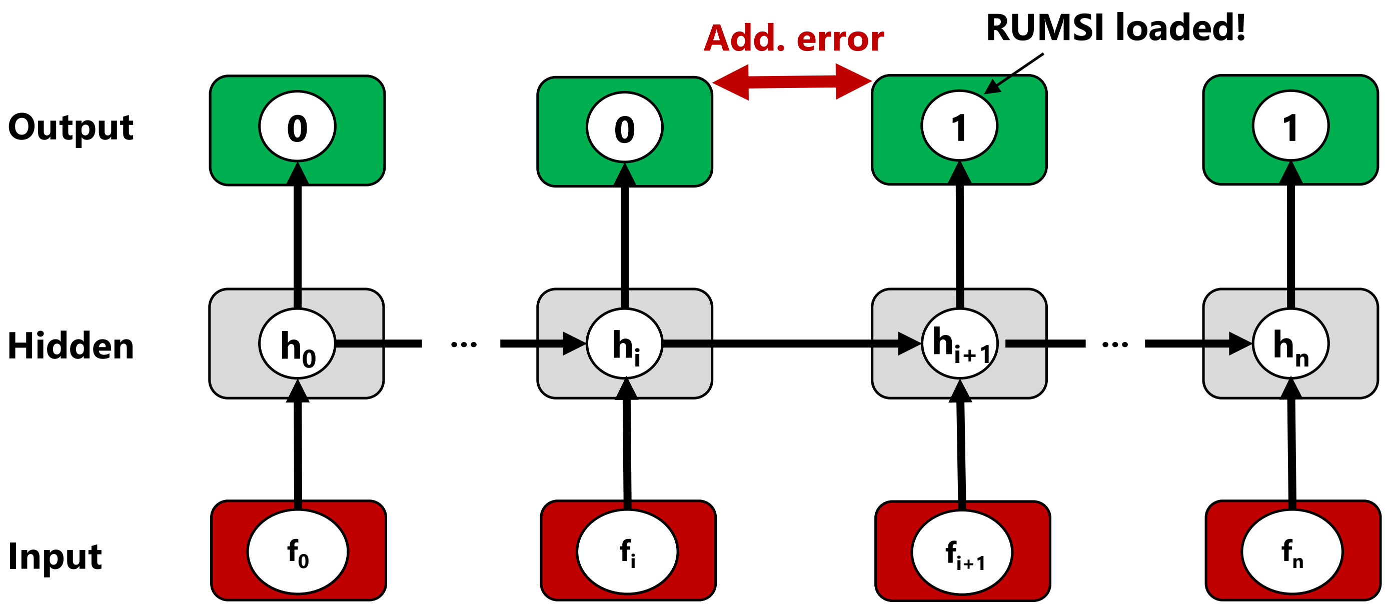
Recurrent Neural Network



Classification Models

- Estimate RUMSI with recurrent binary classification per window
  - Feed **window-based features** into recurrent neural network
  - Map RUMSI to **binary variable** and use as target for training
  - Predict binary RUMSI for 100ms windows **n times**

- **Aggregate binary window predictions** to approximate real RUMSI



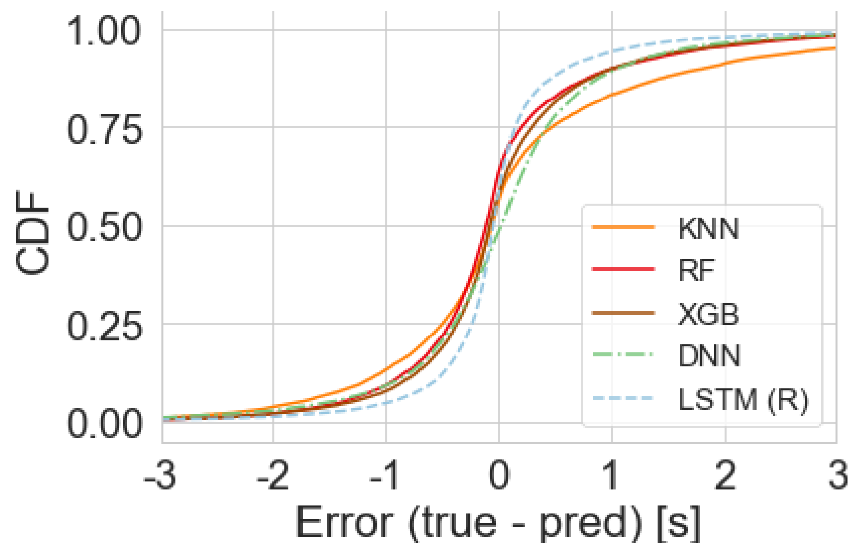
Evaluation of Regression Models

- Stratified website split for training, validation, and test data

- Ensemble methods perform much better than KNN and DNN

→ RF and XGB exhibit a median absolute error of 0.34 seconds

- LSTM outperforms all other models clearly

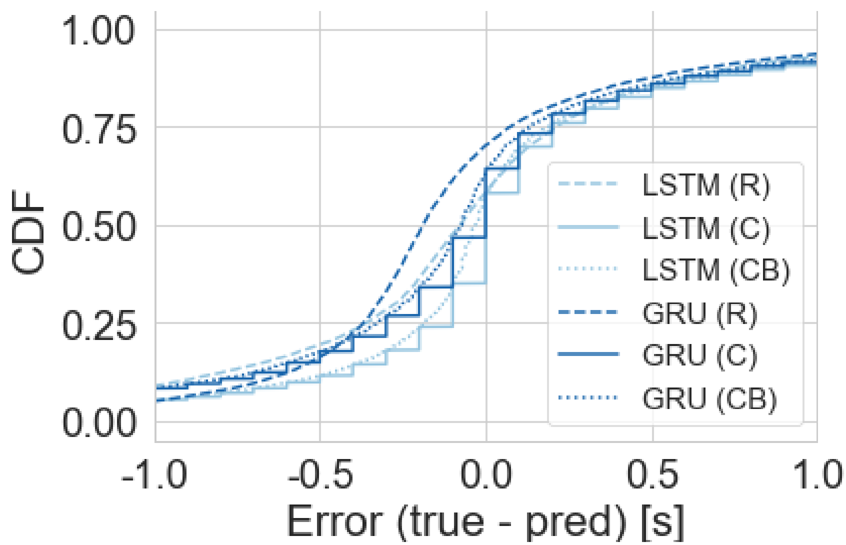


Evaluation of Recurrent Models

- Same evaluation with all recurrent models

- Binary classification (CB) outperforms the regression (R)
  - Additional error (CB-C) does not affect the performance much
  - LSTM layers work slightly better than GRU layers

- *Real-time* prediction is an additional benefit of this approach



Feature Importance Analysis

- Tested several feature subsets for recurrent models
  - A – All, S – Session, W – Window, T/TOP – Incl./Only Top 3
  - No large differences between subsets
  - Top three IP addresses sufficient for good results

	MSE [s]		MAE [s]		mAE [s]	
	GRU	LSTM	GRU	LSTM	GRU	LSTM
A	0.82	0.87	0.52	0.51	0.25	0.20
WT	1.03	0.93	0.58	0.53	0.27	0.24
ST	0.90	0.84	0.50	0.49	0.20	0.19
W	0.94	0.96	0.55	0.54	0.26	0.23
S	1.00	0.87	0.55	0.52	0.24	0.23
TOPW	1.01	0.87	0.61	0.51	0.32	0.22
TOPS	1.00	0.90	0.57	0.53	0.26	0.25

Summary and Outlook

- Testbed for Web QoE measurements on multiple devices

- **Accurate prediction of the RUMSI**
  - On session level with regression approaches
  - On window level with binary classification

- Future works
  - Evaluation on **more diverse data set**
  - Evaluation of multiple Web QoE metrics
  - Reduction of feature dimension
  - Usage of **more sophisticated models**
  - **Clustering of web page families** to enable more fine-grained models for different website families (news, messenger, etc.)

Thank you for your attention!

Questions?