α-OMC: Cost-Aware Deep Learning for Mobile Network Resource Orchestration

Dario Bega¹,², Marco Gramaglia², Marco Fiore³, Albert Banchs¹,² and Xavier Costa-Perez⁴

¹Institute IMDEA Networks, Madrid, Spain
²University Carlos III of Madrid, Madrid, Spain
³CNR-IEIIT, Italy
⁴NEC Laboratories Europe, Heidelberg, Germany
5G mobile networks
CAPACITY FORECASTING
• Traditional approaches deal with **demand** forecasting

**A traffic demand forecasting algorithm aims to minimize the error wrt to the original data, so underestimation is possible.**

**A capacity forecasting algorithm minimizes the amount of resources needed to serve a given demand.**
Cost-aware orchestration - Design

- The loss function is a **key** element in a Deep Learning architecture
  - Measures the error between output and real sample
  - Error values utilized to minimize future errors

\[
\alpha\text{-OMC}(x) = \begin{cases} 
\beta & \text{if } x \leq 0 \\
\gamma \cdot x & \text{if } x > 0. 
\end{cases}
\]

Operator Monetary Cost (OMC)

Overprovisioning | SLA Violations
Cost-restrained orchestration - Implementation

- The loss function is a **key** element in a Deep Learning architecture
  - Measures the error between output and real sample
  - Error values utilized to minimize future errors

![Graph showing Cost-aware orchestration](image)

**Operator Monetary Cost (OMC)**

\[
\alpha\text{-OMC}(x) = \begin{cases} 
\alpha - \varepsilon \cdot x & \text{if } x \leq 0 \\
\alpha - \frac{1}{\varepsilon}x & \text{if } 0 < x \leq \varepsilon\alpha \\
x - \varepsilon\alpha & \text{if } x > \varepsilon\alpha.
\end{cases}
\]

**Overprovisioning**

**SLA Violations**
• Real-world measurement data from a major mobile network operator
  – Single datacenter (470 4G eNodeBs)
  – Orchestration occurs over 5-minutes interval
• Five different fully-connected layers architectures
  – Variable number of hidden layers (from 2 to 20)
  – 16 neurons per layer employing ReLU

• Trained with Adam optimizer

\[ f(s) = \sum_{ij} w_{ij} x_i \]

\[ f(s') : \text{activation function} \]
### Forecast comparison

**MAE**

- Capacity forecast: 187.04 Mbps
- Error

**MSE**

- Capacity forecast: 181.25 Mbps
- Error

**α-OMC**

- Capacity forecast: 29.32 Mbps
- Error

- Accurate forecast for all cases
- Only cost-aware loss minimize penalties
- MAE and MSE result in +500% total cost
• Loss functions behavior over learning time

- MAE and MSE converges to a fixed fee
- $\alpha$-OMC loss function minimizes the operator’s cost
• Loss functions behavior over learning time

- MAE and MSE converges to a fixed fee
- The fee depends by $\alpha$

- $\alpha$-OMC loss function minimizes the operator’s cost
• Loss functions gradients wrt last layer

- All gradients converge to 0
- Minimum (local or global) point in the loss function reached
- Convergence reached after around 50 epochs
Conclusions

- 5G mobile networks requires ad-hoc AI integration
- $\alpha$-OMC represents a very first cost-aware loss function
- It solves the mobile network resource orchestration problem
- Effective in meeting real network requirements compared to SoA loss functions
- Next steps: DeepCog...
Thanks!