A Few Models to Rule Them All: Aggregating Machine Learning Models

Florian Siepe
Phillip Wenig
Thorsten Papenbrock

Objectives

- **Reduce Costs and Complexity:** Transition from using individual models for each heating system to using fewer consolidated models without compromising the prediction accuracy
- **Model Clustering & Consolidation**: Identify groups or clusters of similar heat generator models

CAML

(C)lustering and (A)ggregating (ML) models

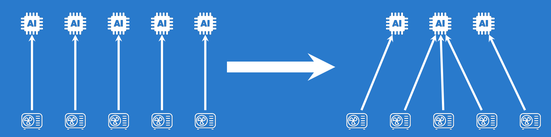
- Hierarchical clustering with custom **Cross Performance** distance function.

  Idea: Calculate the error of any two models m_i, m_j against each other's test sets (x_i, y_i) by applying a loss function l for measuring of how well m_i performs in m_j's environment and vice versa:

$$d(m_i, m_j) = \frac{1}{2}\left(l(m_i(x_j), y_j) + l(m_j(x_i), y_i)\right)$$

- Aggregation of every cluster into a single cluster model by training a new model on the combined training data of all individual models of that cluster

Philipps Universität Marburg



Use Cross Performance distances for aggregating your ML Models

Take a picture to download the full paper

Evaluation

Idea: Measure the ability of a cluster model to generalize

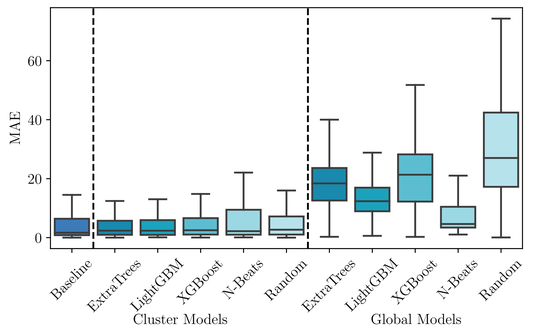
Prediction error of the cluster model on **all test sets** (μ_c) of that cluster vs. the average error of the original models on **their respective test sets** (μ_b)

$$\mu_c(i) = \frac{1}{|C_i|}\sum_{j=1}^{|C_i|} \mathrm{MAE}(M_i(x_{i,j}), y_{i,j})$$

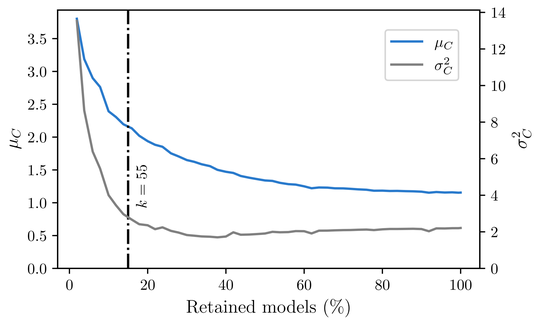
$$\mu_b(i) = \frac{1}{|C_i|}\sum_{j=1}^{|C_i|} \mathrm{MAE}(m_j(x_{i,j}), y_{i,j})$$

$$\mu_C = \frac{1}{N}\sum_{i=1}^{k} |C_i| \cdot \mu_c(i)$$

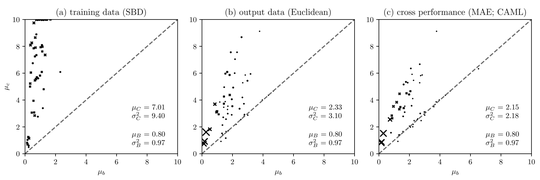
$$\mu_B = \frac{1}{N}\sum_{i=1}^{k} |C_i| \cdot \mu_b(i)$$



Prediction error of cluster models is close to the baseline



With 15% of the models the prediction error μ_C is 2kWh



Cross Performance outperforms clustering on training data and output data