Supervised model for building occupancy prediction from electrical consumption data

Marina Dorokhova (marina.dorokhova@epfl.ch), Christophe Ballif, Nicolas Wyrsch
EPFL, Photovoltaics Laboratory (PVLab), Neuchâtel

Introduction - Background

- Cities account for more than 70% of global CO_2 emissions and consume over two-thirds of the world's energy. Buildings are responsible for nearly 40% of the total energy consumption. [1]
- Buildings represent huge potential for energy savings, which can be realized through intelligent HVAC control.
- Accurate prediction of buildings' occupancy is an important step towards implementation of efficient automation strategies.
- Usage of electrical consumption data, obtained from widely deployed smart meters, allows to predict occupancy in a non-intrusive manner, thus without compromising privacy and security of the occupants, for both residential and tertiary sectors.
- This research uses supervised machine learning techniques for classification in order to demonstrate the potential of occupancy prediction from electricity smart meter measurements.

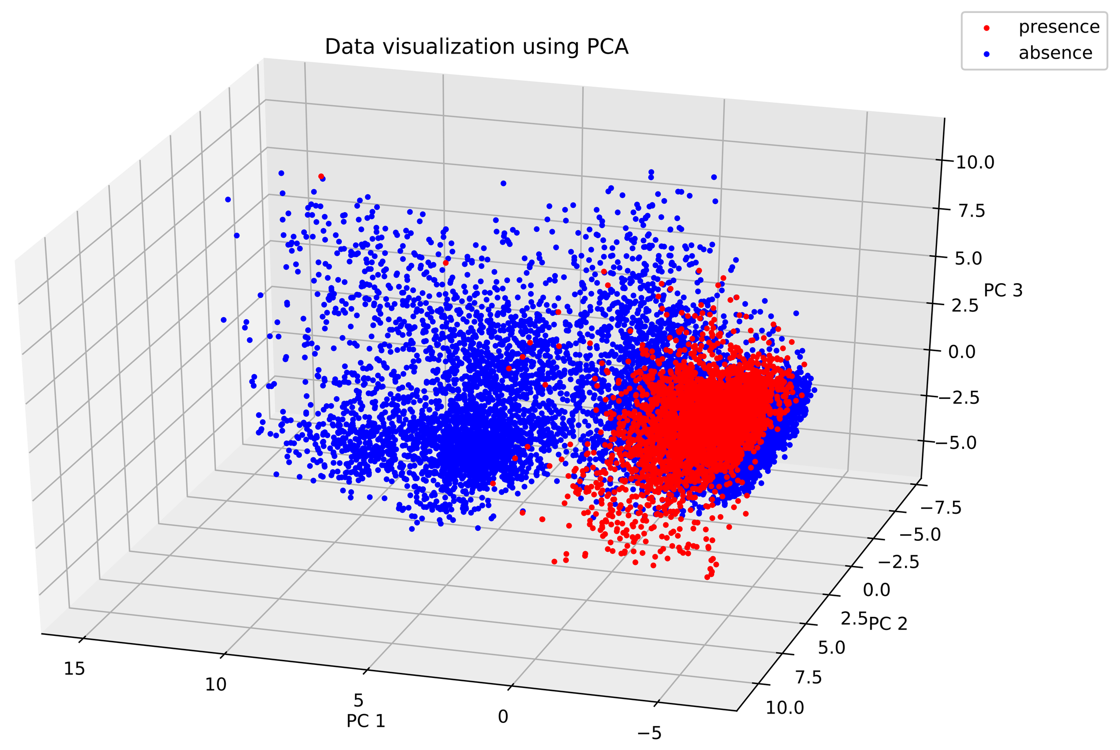

Figure 1: Example of labeled electrical consumption data, the ECO dataset house 2

Methodology

- **Inputs**
  - X: 15-min resolution power measurements from smart meter in [W]
  - y: 15-min resolution binary occupancy data
- **Work pipeline**

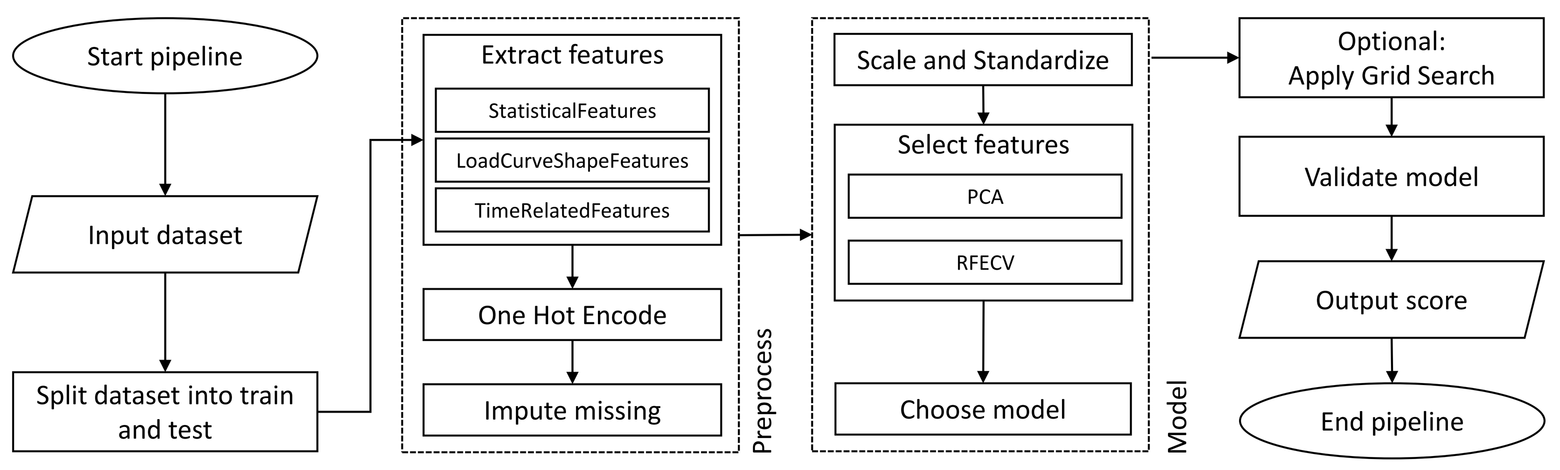

Figure 2: Pipeline for occupancy prediction

- **Datasets**
  - ECO dataset from Switzerland (June 2012 - January 2013) [2]
  - Custom collected dataset from Porto, Portugal (May - June 2018)

Feature Engineering

Three feature families have been manually created (60 features in total):

- **Statistical features**
  Main statistical functions such as min, max, mean, std, median, var, sum and variations of their ratios.
- **Load Curve Shape features**
  Different parameters that are used to describe daily load curve with respect to its shape.
- **Time Related features**
  Information that can be extracted from the measurements' timestamp.

Statistical	Load curve shape	Time related
First order difference	Peaks and valleys	Weekday or weekend
Daily accumulated mean	Area under curve	Calendar holidays
Hourly min and max	Change to night mean	Season

Table 1: Examples of features in each family

Performance Measures

- **Baseline**
  1. Naive predictor, that assumes the building being always occupied
  2. Power variation predictor, that allocates presence, when power goes beyond the threshold (1.25 of daily minimum), and absence otherwise
- **Classification accuracy**

  $$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

  where TP - true positives, TN - true negatives, FP - false positives, FN - false negatives

Results: Implementation and Accuracy Evaluation

- Model output $\hat{y}$ is a binary occupancy vector that spans over the timeframe of the test data.
  - On the graph: presence - green, absence - red.

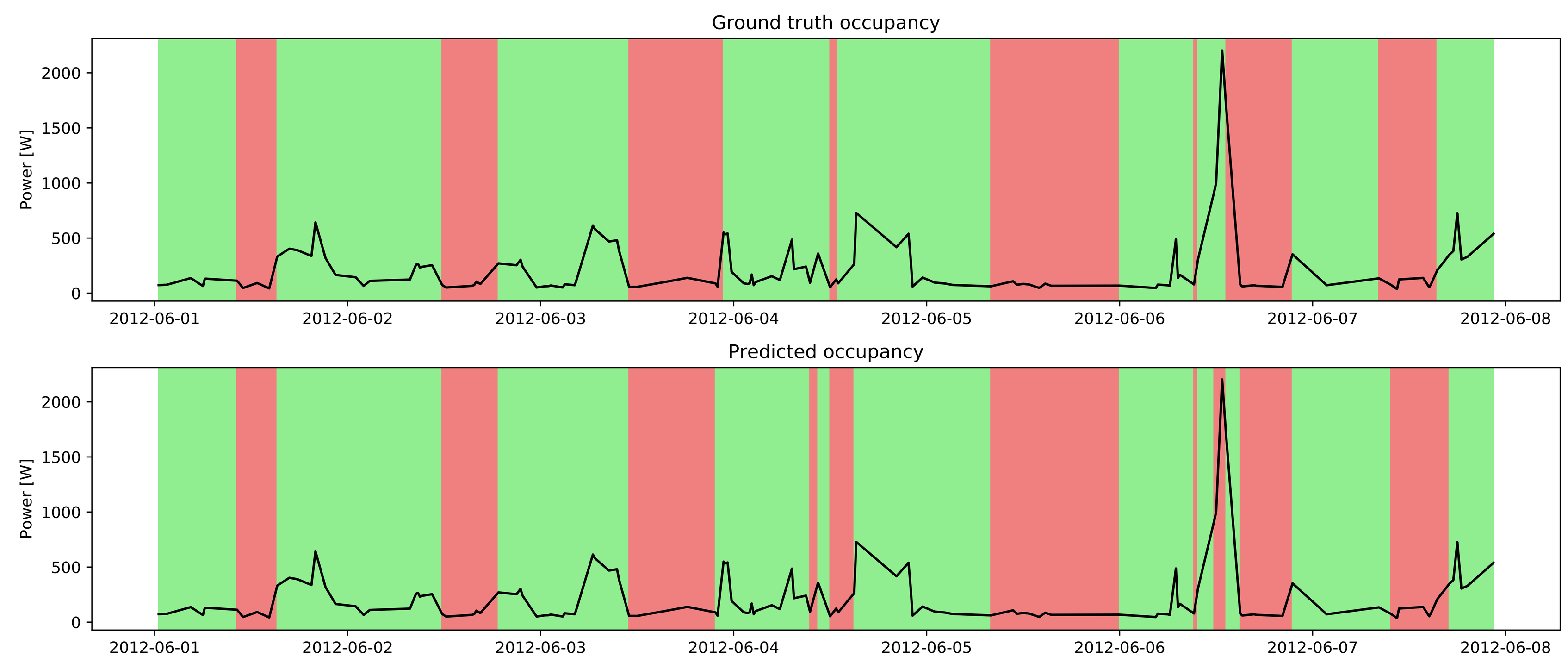

Figure 3: Example of comparison between ground truth and prediction from linear SVM

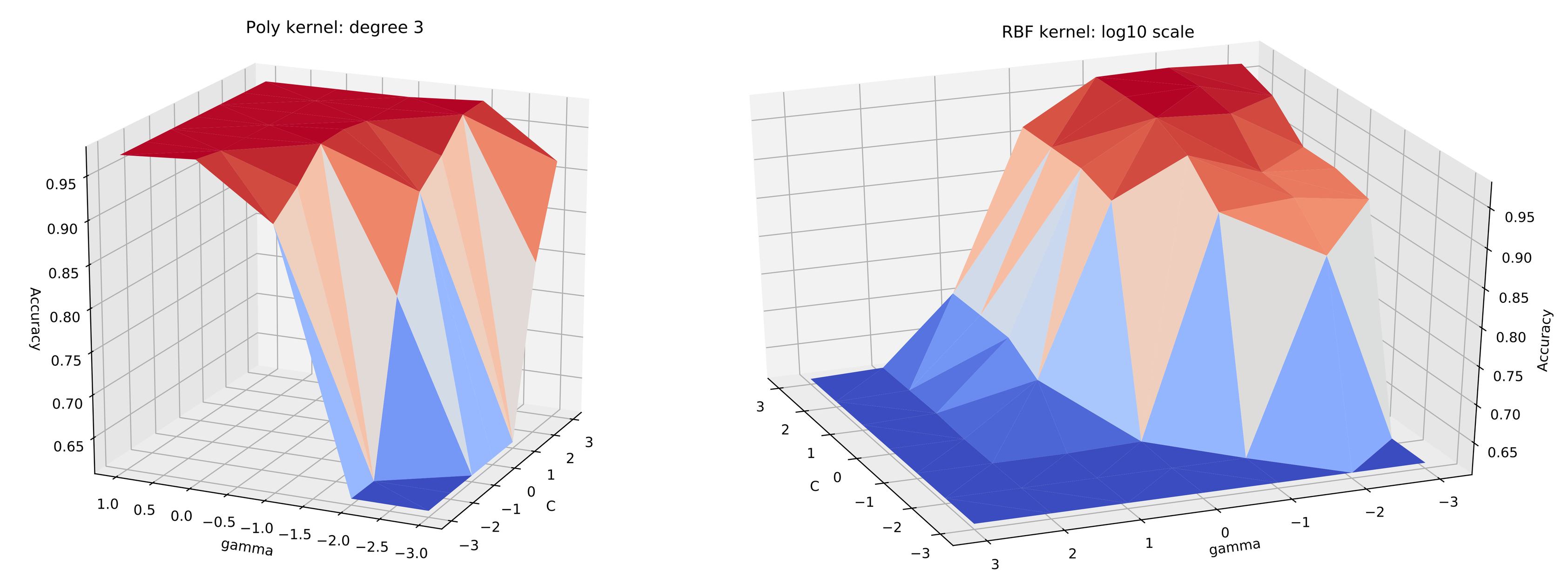

Figure 4: Example of grid search for polynomial and RBF kernels of SVM on Porto dataset

- Evaluation: 5-fold cross-validation and tests on previously unseen data

Model	Cross-validation score		Unseen data score	
	ECO dataset	Porto dataset	ECO dataset	Porto dataset
Baseline 1	73.4%	61.5%	73.5%	61.9%
Baseline 2	73.5%	61.5%	73.1%	61.9%
Linear SVM ($C = 1$)	93.4% (0.51%)	94.8% (0.76%)	93.7%	94.2%
Best SVM on dataset	96.3% (0.38%)	97.5% (0.47%)	96.2%	97.1%
Bagging	95.8% (0.50%)	98.5% (0.72%)	95.6%	98.5%
AdaBoost	90.8% (0.61%)	95.5% (0.64%)	89.6%	95.0%
Feedforward ANN	96.5% (0.68%)	95.8% (0.80%)	96.6%	96.9%
Feedforward ANN with dropout	96.9% (0.64%)	95.8% (1.09%)	96.7%	96.9%

Table 2: Comparison of different models on two datasets

Conclusions

- Implementation of new distinctive features and combined feature selection have been shown to successfully contribute towards better prediction.
- Increased accuracy of occupancy prediction from electrical consumption data has been demonstrated compared to previous works. [1, 3, 4]

References

[1] L. Perez-Lombard, J. Ortiz, and C. Pout. "A review on buildings energy consumption information". In: *Energy and Buildings* 40 (2008), pp. 394–398.

[2] C. Beckel et al. "The ECO Data Set and the Performance of Non-Intrusive Load Monitoring Algorithms". In: *Proceedings of the BuildSys '14* (2014), pp. 80–89.

[3] W Kleiminger, C. Beckel, and S. Santini. "Household Occupancy Monitoring Using Electricity Meters". In: *Proceedings of the UbiComp '15* (2015), pp. 975–986.

[4] T. Vafeiadis et al. "Machine learning based occupancy detection via the use of smart meters". In: *Proceedings of the ISCSIC'17* (June 2017), pp. 6–12.

This research is funded by EU Horizon 2020 Program under the framework of the FEEdBACk project, grant agreement No.768935