Introduction - Background

- Cities account for more than 70% of global CO₂ 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.

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
  \[ \text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \]
  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.

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.

<table>
<thead>
<tr>
<th>Statistical</th>
<th>Load curve shape</th>
<th>Time related</th>
</tr>
</thead>
<tbody>
<tr>
<td>First order difference</td>
<td>Peaks and valleys</td>
<td>Weekday or weekend</td>
</tr>
<tr>
<td>Daily accumulated mean</td>
<td>Area under curve</td>
<td>Calendar holidays</td>
</tr>
<tr>
<td>Hourly min and max</td>
<td>Change to night mean</td>
<td>Season</td>
</tr>
</tbody>
</table>

Table 1: Examples of features in each family

Methodology

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

Datasets

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

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


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