Occupancy Inference based on Machine Learning Techniques

T. Vafeiadis, S. Zikos, D. Ioannidis, S. Krinidis, D. Tzovaras, Centre for Research and Technology Hellas/Information Technologies Institute (CERTH/ITI), Thessaloníki, Greece

The knowledge of occupancy in domestic environments is vital for many applications, such as energy management, building management, demand/response, security, etc. But, in the majority of the cases, it is very difficult or even impossible to install sensors for receiving occupancy information.

Thus, the occupancy (absence – presence or two-class classification scenario) should be detected and inferred utilizing machine learning techniques via electricity and water consumption data received from smart meters in a domestic environment.

To this end, the power consumption of crucial electrical appliances (i.e. TV, washing machine, refrigerator, and hair dryer) is monitored every minute. Also, four water consumption sensors have been installed for monitoring the water usage by the occupants in the kitchen as well as the water consumed by the dishwasher and the washing machine. Finally, an occupancy sensor has been installed at the main entrance of the building detecting entries and exits and measuring its occupancy. This information is utilized as ground truth.

After retrieving the raw data of the three systems, a processing step was performed in order to create the final dataset which includes events per 1-minute intervals of all the measured features. The initial aggregated dataset constructed after processing the raw data contains 9 features [Central Power (lights of the domestic environment), Refrigerator, TV, Washing Machine, Dryer, Cold Water - Kitchen, Hot Water - Kitchen, Dishwasher - Water, Washing Machine - Water] denoted hereafter as [CP, R, TV, WM, D, CWK, HWK, DW, WMW] and the target Occupancy, denoted hereafter as [OCCUP]. The dataset contains energy and water consumption data of 1-minute resolution for a time interval of 16 consecutive days during summer-time. Thus, the shape of overall dataset is 23040x9 (without taking into account the target feature) and its sparsity is 74.44%.

In order to rank the influence of each feature to occupancy inference and extract the more useful information, we have used Mutual Information (MI) as the feature selection technique. MI measures how much one random variable provides information. It is a dimensionless quantity and can be thought of as the reduction in uncertainty about one random variable given knowledge of another. Thus, we decide to use only the top-5 ranked features for occupancy inference, meaning Central Power, Cold Water – Kitchen, Washing Machine, Refrigerator and Washing Machine – Water. Under this condition, the shape of overall dataset is 23040x5 (again without taking into account the target feature) and its sparsity is reduced to 70.76% (from 74.44%).

In the sequence, machine learning techniques have been utilized in order to infer the occupancy in a building. The tested machine learning algorithms are:

- Support Vector Machines (SVMs);
- Decision Trees (DT);
- Random Forest (RF);
- Back-Propagation Network (BPN).

All these approaches were also combined with the AdaBoost algorithm for even more accurate performance.

Our main objective is to find the predictive model that is more efficient on occupancy inference based on energy and water consumption data. To that end, our simulation schema is based on the application of all tested classifiers and their boosting versions on both Initial-DS and MI-DS.

Precision, recall, accuracy and f-measure (estimated averages) for 100 monte-carlo iterations with the application of adaboost, on MI-DS.

Classifier	Precision (%)	Recall (%)	Accuracy (%)	F-measure (%)
SVM – POLY	74.79	89.34	79.83	81.42
SVM – RBF	74.35	89.07	80.06	81.04
DT	74.89	91.37	80.94	82.31
RF	73.91	95.17	80.23	83.20
BPN	74.01	92.83	80.21	82.36

Table I presents precision, recall, accuracy and F-measure (on average of 100 Monte-Carlo iterations) with the application of boosting on tested classifiers on the dataset. One can see that the DT with AdaBoost achieves the higher performance compared to the other tested classifiers (see highlighted values) with 80.94% accuracy (82.31% F-measure), while the RF follows closely in accuracy (80.23%), but achieves higher F-measure compared to DT (83.20%).

The overall framework has been thoroughly evaluated in [1].

References:

1. T. Vafeiadis, S. Zikos, G. Stavropoulos, D. Ioannidis, S. Krinidis, D. Tzovaras and K. Moustakas, “Machine Learning Based Occupancy Detection Via The Use of Smart Meters”, International Conference on Energy Science and Electrical Engineering (ICESEE 2017), Budapest, Hungary, 20-22 October, 2017