Household classification using annual electricity consumption data

Konstantin Hopf[1], Mariya Sodenkamp[1], Ilya Kozlovskiy[1], Thorsten Staake[12]
[1]Energy Efficient Systems Group, University of Bamberg
[2]Department of Management, Technology and Economics, ETH Zurich
{konstantin.hopf, mariya.sodenkamp, ilya.kozlovskiy, thorsten.staake}@uni-bamberg.de

Introduction

The knowledge about household properties (such as number of inhabitants, living area, heating type, etc.) is highly desirable for utility companies to pave the way to targeted energy efficiency programs, products and services.

Raising individual household data via surveys or purchasing it is expensive and time consuming, and often only a small fraction of customers participate.

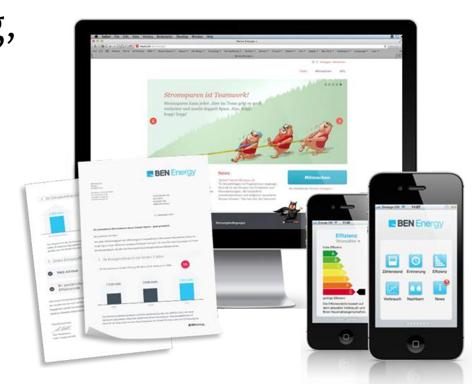

Figure 1: Potential personalized energy-efficiency products and services, e.g. online platforms, apps, direct mailings (Source: BEN Energy AG)

Recently, data mining methods have been developed to automatically infer house-hold characteristics from smart meter consumption data. However, the slow smart metering rollout hampers practical implementation of these methods in many countries. In this work, we present a machine learning approach that reveals household properties from *conventional annual electricity consumption data* currently available at a large scale.

Data

Three real-world datasets containing information about more than 5'500 private dwellings in Germany and Switzerland and are used for algorithm training and validation.

The datasets contain annual electricity consumption over one, three and five years respectively, and the customers' addresses. From this data, we derive three feature categories:

1) Mean consumption (*CPD_mean*)
2) Consumption deviation to the postal code region (*diff_mPLZ*).
3) Consumption development over years: the variance (*CPD_var*) and the deviation from the two-year moving average (*mad_12_3*).

Besides these consumption features, *five household properties* (Table 1) are known.

Table 1: Five household properties that can be recognized by the classification algorithm

Property	Class definition
pHouseholdType	apartment
	house
pLivingArea	≤ 95 m²
	≤ 145 m²
	> 145 m²
pNumResidents	single
	2 persons
	3+ persons
pHeating	electric heating
	not electric heating
pWaterHeating	electric water heating
	not electric water heating

Table 2: Classification settings and feature sets for evaluating the classification transferability

Training with		Classification with: Dataset A	Dataset B	Dataset C
	Dataset A	AA: CPD_mean, CDP_var, mad_1112_13, diff_mPLZ_12	AB: CPD_mean, CPD_var, mad_12_3, diff_mPLZ_11	AC: CPD_mean, diff_mPLZ
	Dataset B	BA: CPD_mean, CPD_var, mad_12_3, diff_mPLZ_11	BB: CPD_mean, CPD_var, mad_0910112_12, diff_mPLZ_11	BC: CPD_mean, diff_mPLZ
	Dataset C	CA: CPD_mean, diff_mPLZ	CB: CPD_mean, diff_mPLZ	CC: CPD_mean, diff_mPLZ

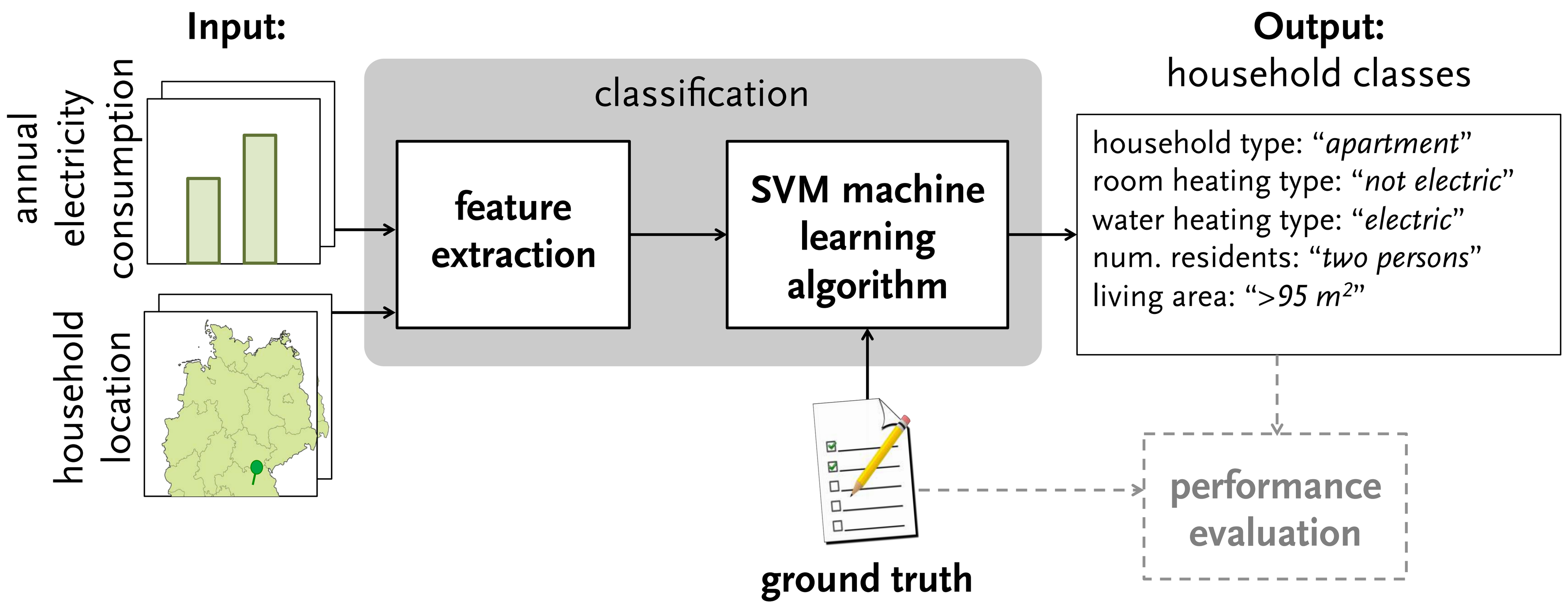

Figure 2: Classification and evaluation methodology

Methods

We propose a supervised machine learning technique for recognition of energy efficiency relevant household properties using annual consumption data and household location information.

The classification procedure applied in this work is schematically illustrated in Figure 2. At first, the input data is prepared with feature extraction methods. The defined features are described in the previous section together with the data.

To reveal household characteristics, the SVM supervised learning algorithm is trained with labeled training instances and is thereafter applied to new data instances for the prediction of household classes.

For higher classification performance, we found optimal parameters for this application empirically.

Evaluation and results

To evaluate the performance, we count the number of correct and misclassified examples in comparing the predicted household classes with ground truth data and calculate the classification *accuracy* as the percentage of correct classified examples in the number of all examples. We answer two research questions that are presented as follows.

Result 1 – Feasibility of household classification based on annual electricity consumption data

The classification results show that supervised machine learning can predict household classes with an accuracy between 47% and 95%. By analyzing the classi-fication results with a single dataset (setting AA, BB, CC), we can conclude the following statements for household classification with annual consumption data:

Classification with only one year of consumption and information about the neighborhood (dataset C), can achieve higher classification accuracy than a biased random guess (with respect to the properties "living area" and "number of residents") when the class sizes within one property are about equally. The results for properties with unbalances classes (one class is more than two times larger than another) show a low accuracy. The class sizes in the property "type of household" are unbalanced, yet the poor result for this property can be related to the selection bias of dataset C, because the dataset contains mainly customers with high consumption.

While *having multiple data points* (dataset A and B), the number of features increases and the classification accuracy is improved. Especially the features describing trends and the variance of consumption have a positive impact.

Result 2 – Transferability of trained household classification models to other datasets

To test the classifier applicability for different datasets (i.e., classifier "transferability"), we train the algorithm using one dataset and check how it performs on two other datasets. As it can be anticipated, the transferred results are lower than classification with the same dataset. However, comparing the transfer between dataset A and B that have multiple years of consumption, classification accuracy of the balanced properties show higher values than the biggest class size, except the unbalanced property "type of heating". Moreover, the recall values of most of the classes are higher than 40% (see Figure 3). This means, for example, for the class "house" that having trained the classifier with A, the algorithm finds >80% of all customers in dataset B who live in a house. Because of the reasons for the lower classification performance in dataset C the transferability from and to a dataset with only one year is limited. We assume that further main influence factors to the transferability results are sample selection effects.

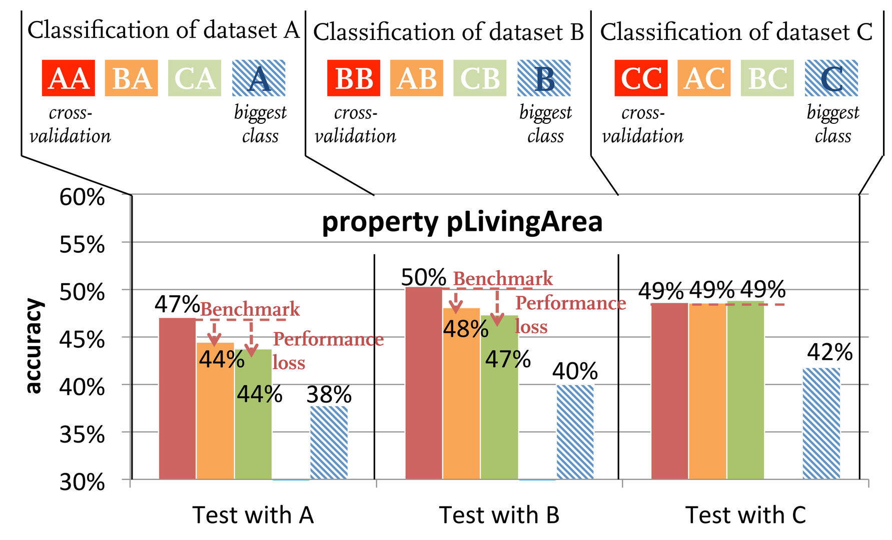

Figure 3a: Classification accuracy and transferability results for the property pLivingArea with legend for accuracy figures

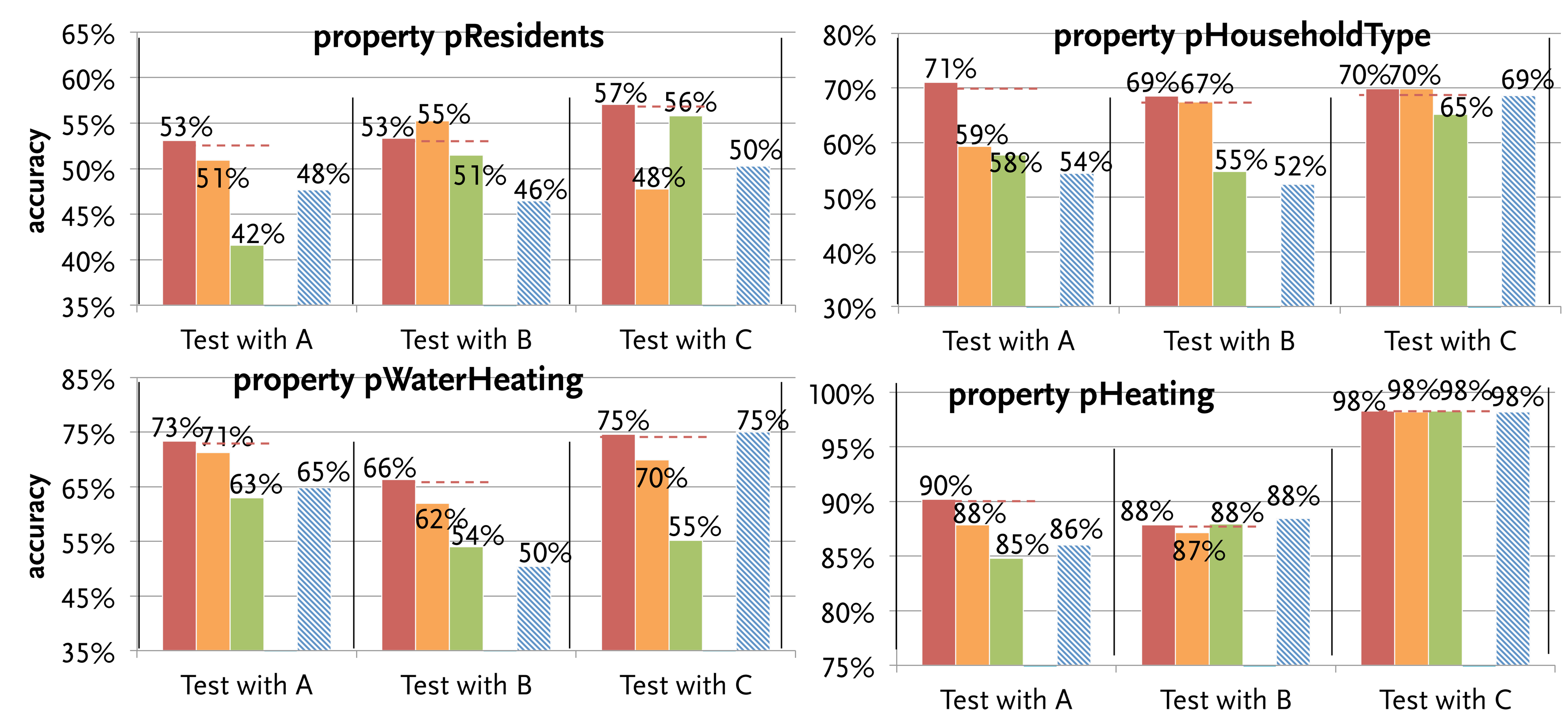

Figure 3b: Classification accuracy and transferability results for the properties pResidents, pHouseholdType, pWaterHeating and pHeating

References

Beckel, C., Sadamori, L., Staake, T., Santini, S.: Revealing household characteristics from smart meter data. Energy. 78, 397–410 (2014).

Fischer, C.: Feedback on household electricity consumption: a tool for saving energy? Energy Efficiency. 1, 79–104 (2008).

Graml, T., Loock, C.-M., Baeriswyl, M., Staake, T.: Improving residential energy consumption at large using persuasive systems. In: ECIS (2011).

Guyon, I., André, Elisseeff: An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182 (2003).

Hopf, K., Sodenkamp, M., Kozlovkiy, I., Staake, T.: Feature extraction and filtering for household classification based on smart electricity meter data. Computer Science-Research and Development. 1–8 (2014).

Sodenkamp, M., Hopf, K., Staake, T.: Using Supervised Machine Learning to Explore Energy Consumption Data in Private Sector Housing. Handbook of Research on Organizational Transformations through Big Data Analytics. 320 (2014).

Vapnik, V.N., Vapnik, V.: Statistical learning theory. Wiley New York (1998).

Vassileva, I., Odlare, M., Wallin, F., Dahlquist, E.: The impact of consumers' feedback preferences on domestic electricity consumption. Applied Energy. 93, 575–582 (2012).

Acknowledgements

The research presented in this poster was financially supported by Swiss Federal Office of Energy (Grant number SI/ 501053-01) and Commission for Technology and Innovation in Switzerland (CTI Grant number 16702.2 PFEN-ES). The anonymized data used in this analysis was provided by BEN Energy AG, Switzerland.

Bundesamt für Energie
Komission für Technologie und Innovation