Chapter 19

Using Supervised Machine Learning to Explore Energy Consumption Data in Private Sector Housing

Mariya Sodenkamp
University of Bamberg, Germany

Konstantin Hopf
University of Bamberg, Germany

Thorsten Staake
University of Bamberg, Germany & ETH Zurich, Switzerland

ABSTRACT

Smart electricity meters allow capturing consumption load profiles of residential buildings. Besides several other applications, the retrieved data renders it possible to reveal household characteristics including the number of persons per apartment, age of the dwelling, etc., which helps to develop targeted energy conservation services. The goal of this chapter is to develop further related methods of smart meter data analytics that infer such household characteristics using weekly load curves. The contribution of this chapter to the state of the art is threefold. The authors first quadruplicate the number of defined features that describe electricity load curves to preserve relevant structures for classification. Then, they suggest feature filtering techniques to reduce the dimension of the input to a set of a few significant ones. Finally, the authors redefine class labels for some properties. As a result, the classification accuracy is elevated up to 82%, while the runtime complexity is significantly reduced.

DOI: 10.4018/978-1-4666-7272-7.ch019
1. INTRODUCTION

Smart grid technology is one of the main pillars of next-generation sustainable energy management. Detailed information about energy consumption is useful to both energy users and utilities. Energy providers can better forecast demand, improve meter-to-cash-processes, and manage supply, while customers get better control over their own electricity usage, costs, and caused carbon footprint. In the last decade, governments across Europe placed high hopes on smart metering as an indispensable means to improve energy efficiency in the residential sector.

Smart meters typically record energy consumption in 15- to 30-minute intervals, and the information is sent back to the energy supplier. The gained information then can be used to support grid optimization processes and to derive targeted interventions toward customers. Combined with direct consumer feedback, smart metering has been promised to increase energy efficiency by 5-15%, though the current field studies show savings of around 3% (Degen et al., 2013).

Given the lower-than-expected saving effects, European countries are moving from the euphoria regarding its potential. The reasons are high upfront costs, customers’ privacy concerns, and, most importantly, lack of analytical implications from the data for policy and efficiency measures development. Meanwhile, utilization of the information from fine-grained consumption profiles is in its infancy. With current expectations running high – still burdened with a risk of disappointment – developing new analytical approaches of data processing, interpretation, and utilization remains one of the major challenges.

This work is aimed at filling the gap between raw smart metering consumption data and the information required for developing energy policy toward dwellings. We improve previous algorithms for household classification with predicting energy efficiency relevant characteristics (such as type of heating, size/age of house, number of inhabitants/children, etc.) from the consumption traces. In essence, we rely upon the pioneering work of Beckel et al. (Beckel et al., 2013) who showed that household properties can be inferred from fine-grained consumption data using supervised machine learning with accuracy above 60%. In short, we focus on three following aspects to improve the classification accuracy and to reduce runtime complexity of the Beckel et al.’s algorithm:

1. Extending the feature set,
2. Applying feature filtering methods, and
3. Refining property definitions.

For training and test of the proposed model, we relied upon the dataset acquired from the Irish Commission for Energy Regulation (CER) (Irish Social Science Data Archive, 2014). It contains 30-minute smart electricity meter and survey data from about 4200 private households collected during a 76-week period of 2009 and 2010 in Ireland. We analyzed one noise-free week (21-27.09.2009) without public holidays and with average weather conditions. The results show that the classification accuracy of the upgraded algorithm can be significantly improved (up to 82%), while reducing the computational complexity.

This chapter is organized as follows: Section 2 gives an overview of the related work. Section 3 describes our classification system. It is followed by a more detailed description of the underlying dataset, along with the proposed feature/property definition and feature selection methods in Section 4. Finally, Section 5 presents the attained results and main conclusions of this work.