Energy 78 (2014) 397–410

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Revealing household characteristics from smart meter data

Christian Beckel [a,*], Leyna Sadamori [a], Thorsten Staake [b], Silvia Santini [c]

[a] Institute for Pervasive Computing, Department of Computer Science, ETH Zurich, Universitätstr. 6, 8092 Zurich, Switzerland
[b] Energy Efficient Systems Group, University of Bamberg, An der Weberei 5, 96047 Bamberg, Germany
[c] Wireless Sensor Networks Lab, TU Darmstadt, Rundeturmstr. 10, 64283 Darmstadt, Germany

ARTICLE INFO

Article history:
Received 20 May 2014
Received in revised form
1 October 2014
Accepted 6 October 2014
Available online 5 November 2014

Keywords:
Data-driven energy efficiency
Domestic electricity consumption
Electricity load profiles
Automated customer segmentation
Supervised machine learning

ABSTRACT

Utilities are currently deploying smart electricity meters in millions of households worldwide to collect fine-grained electricity consumption data. We present an approach to automatically analyzing this data to enable personalized and scalable energy efficiency programs for private households. In particular, we develop and evaluate a system that uses supervised machine learning techniques to automatically estimate specific "characteristics" of a household from its electricity consumption. The characteristics are related to a household's socio-economic status, its dwelling, or its appliance stock. We evaluate our approach by analyzing smart meter data collected from 4232 households in Ireland at a 30-min granularity over a period of 1.5 years. Our analysis shows that revealing characteristics from smart meter data is feasible, as our method achieves an accuracy of more than 70% over all households for many of the characteristics and even exceeds 80% for some of the characteristics. The findings are applicable to all smart metering systems without making changes to the measurement infrastructure. The inferred knowledge paves the way for targeted energy efficiency programs and other services that benefit from improved customer insights. On the basis of these promising results, the paper discusses the potential for utilities as well as policy and privacy implications.

1. Introduction

Customer insights help utilities to optimize their energy efficiency programs in many ways [1,2]. With knowledge of the socio-economic characteristics of individual households, for instance, utilities can automatically tailor savings advice to specific addressees (e.g., to families with children, or to retirees). Further, they can offer consumption feedback that includes references to similar households or consider the financial reach of their customers when suggesting improvements in the appliance stock. Many studies have shown that such specific approaches improve the performance of efficiency campaigns [3–5]. Yet, such targeted measures require detailed information on individual customers, which might be gathered for research studies and local saving campaigns, but which is often not available for large-scale, cost sensitive efficiency programs that are directed to millions of households.

In fact, utilities' knowledge about their customers is often limited to their address and billing information. This is particularly true in Europe, where open information repositories like public tax registers do not exist or cannot be easily accessed. On the other hand, conducting surveys to acquire customer information is typically time-consuming and expensive, and often only a small fraction of customers participate [6]. We argue that utilities can instead utilize the electricity consumption data of a household to reveal customer information that is relevant to optimizing their energy efficiency programs. This is valuable for utilities, because they are already deploying millions of smart electricity meters in private households along with infrastructure to collect, process, and store their electricity consumption data [7–9]. Currently, utilities use this data mainly to improve their meter-to-cash processes, to enable advanced tariff schemes, and to provide customers with detailed information on their electricity consumption. Analyzing smart meter data that is collected anyway can therefore contribute to the value of the metering infrastructure without requiring any changes to the smart meters that have already been deployed.

In this paper, we develop and evaluate a system to automatically infer household characteristics from smart meter data. Examples of such characteristics include the household's socio-economic status, its dwelling properties, and information on the appliance stock. Our analysis takes as input the electricity consumption of a household and estimates the value of several characteristics of interest.

* Corresponding author. Tel.: +41 44 632 7871; fax: +41 44 632 1659.
E-mail addresses: beckel@inf.ethz.ch (C. Beckel), sadamori@inf.ethz.ch (L. Sadamori), thorsten.staake@uni-bamberg.de (T. Staake), santinis@wsn.tu-darmstadt.de (S. Santini).

http://dx.doi.org/10.1016/j.energy.2014.10.025
0360-5442/

Depending on the characteristic, this value is either the class to which the household most likely belongs to (e.g., employment status) or a numerical value (e.g., the number of persons living in the household). To infer the value of household characteristics from consumption data, we extract features from the data itself and pass them as input to a classifier or regression model. An example of such a feature is the average consumption of a household between 10 a.m. and 2 p.m. divided by its daily average consumption. This particular feature helps to reveal household occupancy during lunch time and thus contributes to the estimation of characteristics such as the employment status of the inhabitants. We investigate 18 different characteristics which we have selected because they are relevant to utilities [10]. We have evaluated our system according to these characteristics using smart meter data available at a 30-min granularity from 4232 Irish households over a period of 1.5 years. This data set is publicly available and has been collected in the context of a smart metering trial conducted by the Irish CER (Commission for Energy Regulation).[1] In the following, we refer to this data set as the CER data set. Along with smart meter data, the data set contains information on the characteristics of each household collected through questionnaires before and after the study. This information is crucial for our work, because it represents ground truth data we can use to validate our findings.

The contribution of this paper is a comprehensive system for automatically revealing household characteristics from smart meter data and an elaborate evaluation of our approach. In our previous work [11], we presented a preliminary study to demonstrate the feasibility of revealing household properties from smart meter data. In this paper, we improve upon our previous work in multiple respects: First, we present new components of our system. We extend the feature set, replace the feature selection method, and add a classifier. Second, we perform a detailed analysis to evaluate the applicability of our results. In particular, we advocate and discuss new performance measures (e.g., to handle imbalanced classes), we investigate six additional characteristics that are of interest to utilities, and we propose and evaluate the utilization of the classifier confidence to identify small groups of customers with improved performance. We also propose a regression model in order to estimate characteristics with continuous values (e.g., the number of persons in a household). Finally, we show the stability of the results overall 75 weeks included in the data set, and we show significant performance gains that can be achieved when performing the analysis on the whole measurement period instead of a single week of data only, as it was done in Ref. [11].

The results provided in this paper show that revealing household characteristics from smart meter data is feasible with sufficient accuracy. This holds in particular for characteristics related to the number of persons living in a household and for characteristics related to the occupancy of the household (which also includes information on the employment status of the chief income earner). We show that it is possible to infer 8 of the 18 characteristics with an accuracy between 72% and 82%. Overall, our approach performs roughly 30 percentage points better than assigning characteristics to the households at random. Some applications require identifying households that feature a specific characteristic with high accuracy. This is for instance necessary when a group of households (e.g., those inhabited by a single person) are the target of a marketing campaign. Here, reducing the number of false positives (i.e., of the cases in which a household is erroneously estimated to belong to the target class) is crucial. We show that by exploiting the confidence of the estimation obtained from the classifiers, it is possible to reduce the number of false positives significantly.

According to the results reported in this paper, utilities can reliably estimate household characteristics from smart meter data. Thus, they will be able to improve their energy efficiency campaigns and make them applicable to the mass market as they scale to thousands or millions of customers with little additional effort. Ultimately, creating these services to help their customers use energy more efficiently is crucial for utilities' attempts to comply with regulatory targets [12]. In addition, the system provided in this paper allows utilities to improve customer retention, which is becoming more relevant in a liberalized energy market [13]. To the best of our knowledge, this is the first study that provides a quantitative analysis of the possibility of revealing household characteristics from electricity consumption data on such a large data set and at such a high accuracy.

The remainder of this paper is structured as follows. Section 2 reviews related work. We then present the data set we use in our study in section 3 and our methodology in section 4. Next, we describe our evaluation setup and performance measures in section 5. Section 6 presents the results of the analysis followed by a discussion of the results in section 7. Finally, section 8 concludes the paper and gives an outlook on future work.

2. Related work

Over the past years, an increasing number of researchers have applied machine learning and data mining techniques to model and analyze residential electricity consumption data. This has been made possible thanks to the increasing availability of electricity consumption data. A popular line of research in this context is one focusing on NILM (*non-intrusive load monitoring*). Using aggregated electricity consumption data of individual households (e.g., measured at 1 reading per second or millisecond), researchers have tackled the problem of disaggregating the consumption of individual appliances. This information allows in turn to provide detailed consumption feedback to the households [14–16]. The work we present in this paper is considerably different from NILM, because we aim to infer high-level household characteristics from the electricity consumption instead of disaggregating it into its individual end use.

Other authors have focused on the analysis of coarse-grained consumption data (i.e., data sampled at a granularity of several minutes or higher). Here, we distinguish between (1) analyzing consumption data only and (2) relating it to side-information such as the geographic location of the dwelling or the socio-economic status of the household. Since the first approach imposes less requirements on the collected data, many authors have investigated unsupervised techniques such as clustering to detect patterns and usage categories in the consumption profiles [17–20]. Chicco, for instance, provides an overview of clustering techniques used to group residential or commercial customers according to their electricity consumption pattern [19]. Grouping consumers by their load profile enables utilities to formulate tariffs for specific customer categories, check the effect of tariff modifications, and ultimately optimize their supply management. Using similar techniques, both Kwac et al. [18] and Cao et al. [17] have focused identifying the "right" customers for demand-side management campaigns. Whereas Kwac et al. aim at detecting stable profiles over a certain time period, Cao et al. focus on identifying households with a similar time of peak usage. Finally, De Silva et al. aim at predicting future electricity usage of private households using a data mining framework and an incremental learning algorithm [20]. In contrast to all these approaches, our work goes beyond detecting consumption patterns or usage categories. We utilize such patterns to estimate specific characteristics of the socio-economic status, the dwelling, or appliance stock of the households.

[1] www.ucd.ie/issda/data/commissionforenergyregulationcer/.