SmartD: Smart Meter Data Analytics Dashboard*

Aylin Jarrah Nezhad, Tri Kurniawan Wijaya, Matteo Vasirani, and Karl Aberer
School of Computer and Communication Sciences
École Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
{aylin.jarrahnezhad, tri-kurniawan.wijaya, matteo.vasirani, karl.aberer}@epfl.ch

ABSTRACT

The ability of smart meters to communicate energy consumption data in (near) real-time enables data analytics for novel applications, such as pervasive demand response, personalized energy feedback, outage management, and theft detection. Smart meter data are characterized by big volume and big velocity, which make processing and analysis very challenging from a computational point of view. In this paper we presented SmartD, a dashboard that enables the data analyst to visualize smart meter data and estimate the typical load profile of new consumers according to different contexts, temporal aggregations and consumer segments.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data mining; G.3 [**Probability and Statistics**]: Time series analysis

Keywords

smart meters; visualization; energy consumption analysis

1. INTRODUCTION

Future ICT-based energy systems will rely on an Advanced Metering Infrastructure (AMI), a system that measures and collects data about energy usage and power quality using smart meters installed at the consumer premises [5]. Smart meter data has an important role in several Smart Grid applications and enables novel data analytics tasks, such as energy consumption behavior analysis, theft detection, outage management, pervasive demand response at residential level, and personalized energy feedback. However, processing and analyzing smart meters data is very challenging, because it is characterized by big volume and big velocity, and how to *extract* useful information from it is still an open question.[1]

*Supported by European Union's Seventh Framework Programme (FP7/2007-2013) 288322, WATTALYST. The authors would like to thank Julien Eberle and Sofiane Sarni for the assistance with GSN.

[1]See Bryan Truex, "Two Opposing Views on Smart Meter Data Analytics", `http://bit.ly/LYADfH`

e-Energy'14, June 11–13, 2014, Cambridge, United Kingdom
ACM 978-1-4503-2819-7/14/06.
http://dx.doi.org/10.1145/2602044.2602046.

In this paper we present SmartD, a dashboard for smart meter data visualization and analysis. SmartD has been built to be (i) seamlessly integrated with existing data collection infrastructures, (ii) intuitive to use, and (iii) easy to extend. To visualize and analyze smart meter data, SmartD supports context selection (e.g., summer, winter, weekend, or weekdays), different temporal aggregations (e.g., hourly, daily, or weekly), and consumer selection (either individual or clusters of consumers). Although this functionality is commonly found in other energy dashboard or time series visualization, SmartD's additional and novel contributions are: (i) estimating the typical hourly load profiles based on demographic information, and (ii) determining the attributes of the demographic profile that are relevant to consumer's energy consumption behavior for a given context. This functionality can be used to predict the typical load profile of new consumers, or to understand the energy consumption behavior of different consumers (e.g., employed vs retired, family vs single).[2]

2. SMARTD

We developed SmartD on top of GSN [2], a widely used middleware for sensor networks deployment. Given that smart meters are essentially sensors, GSN can be seamlessly integrated with an existing smart metering infrastructure, enabling applications running on GSN to receive real-time smart meter readings (push mode) as well as obtaining them from a DB or text files (pull mode). Figure 1 shows the architecture diagram of SmartD.

SmartD needs to be able to (i) retrieve and process smart meter data with big volume and velocity, and (ii) visualize and extract valuable information from that data. While the first capability is provided by GSN, we briefly explain the second in the following sections. We remark that although for demonstration purposes we use the Irish CER dataset [1][3], SmartD can be used with any time series of smart meter data and consumer demographic profiles in the form of <*attribute,value*> tuples.

2.1 Energy Consumption Analysis

For the visualization of energy consumption data (see Figure 2), SmartD has several key features, detailed below.

Temporal aggregations. SmartD supports different time granularities, from half-hourly to monthly. In addition, a set of basic statistical aggregation functions is also provided, such as sum, average, min, and max.[4]

[2]We use the terms *energy consumption* and *load* interchangeably.
[3]This dataset contains measurements of approximately 5,000 consumers for 1.5 years (Jul 2009 - Dec 2010)
[4]More sophisticated aggregation functions can be easily added.

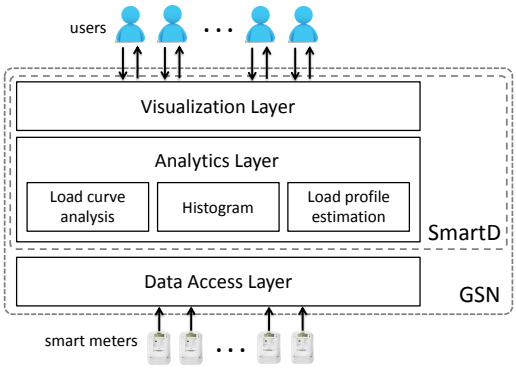

Figure 1: SmartD architecture diagram

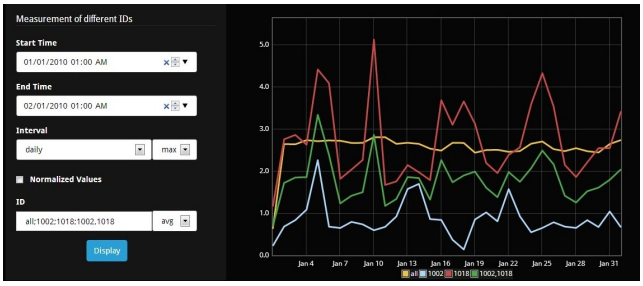

Figure 2: Energy consumption analysis

Consumer aggregations. SmartD supports visualization of energy consumption of a set of individual consumers, clusters of consumers, or a mix of both. We use a simple but flexible grammar $id\big((,id) \mid (;id)\big)^*$ to specify the desired visualization, where id is a consumer identifier, the character ";" separates clusters, and the character "," separates consumers within a cluster. An individual consumer is then expressed as a cluster of one consumer. If a cluster of more than one consumer is specified, the users can choose the functions to aggregate the energy data within the cluster (such as sum, average, min, or max).

Consumer characterization. SmartD provides an option to focus more on consumer demand shape, by plotting z-normalized data. This functionality can be used, e.g., to spot consumers who have morning peak, evening peak, or both.

Histogram. SmartD provides a histogram view with the distribution of energy consumption values, which can be useful to analyze the way people consume energy. For example, we found that the energy consumption of residential consumers follows a log-normal distribution, peaked around their base load, while commercial and industrial consumers follows a normal distribution, peaked around the mean consumption of working hours.

2.2 Energy Consumption Estimation

SmartD supports data analysts by providing insights related to energy consumption behavior. First, it helps to answer questions about consumer load profile given her demographics, such as *what is the difference between load profile of families with and without children?*, or *can we estimate the typical load profile of a new consumer using her socio-demographic information?*. SmartD estimates the load profile of a consumer (see Figure 3), if provided with (a subset of) the consumer demographic information, as well as the context of interest (e.g., weekend, Monday, summer, etc.).

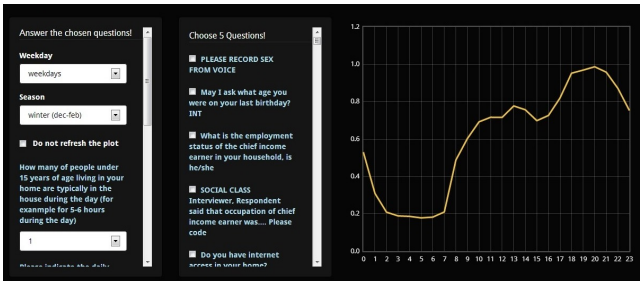

Figure 3: Energy consumption estimation

This is done using *k-nearest neighbor* (*k-NN*), where the best k is determined using *leave-one-out cross-validation*.

Second, SmartD also helps to infer demographic information which significantly influence energy consumption on a specific context, e.g., weekend, Monday, or summer. This is implemented using correlation-based feature selection [4] and *k-NN*, where the features are the demographic information and the target classes are the hourly consumption values. These learning functionalities is developed using the WEKA machine learning library [3].

3. CONCLUSION

In this paper, we presented SmartD, a dashboard for smart meter data analysis and visualization developed on top of GSN. SmartD has been released as an open-source project.[5] SmartD capabilities include visualization of energy consumption data and estimation of the typical load profile of a consumer according to her demographic and contextual information. As future extensions, other functionalities such as customer segmentation [7], interpolation of missing values, and load forecasting [6], could be added.

4. REFERENCES

[1] Smart Metering Trial Data Publication. The Commission for Energy Regulation (CER), 2012.

[2] K. Aberer, M. Hauswirth, and A. Salehi. A Middleware for Fast and Flexible Sensor Network Deployment. In *Proceedings of the 32nd International Conference on Very Large Data Bases*. VLDB Endowment, 2006.

[3] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA Data Mining Software: An Update. *SIGKDD Explor. Newsl.*, 11(1):10–18, Nov. 2009.

[4] M. A. Hall. *Correlation-based Feature Subset Selection for Machine Learning*. PhD thesis, University of Waikato, Hamilton, New Zealand, 1998.

[5] D. Hart. Using AMI to Realize the Smart Grid. In *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, pages 1–2, 2008.

[6] S. Humeau, T. K. Wijaya, M. Vasirani, and K. Aberer. Electricity Load Forecasting for Residential Customers: Exploiting Aggregation and Correlation between Households. In *Sustainable Internet and ICT for Sustainability (SustainIT), 2013*, pages 1–6, Oct 2013.

[7] T. K. Wijaya, T. Ganu, D. Chakraborty, K. Aberer, and D. P. Seetharam. Consumer Segmentation and Knowledge Extraction from Smart Meter and Survey Data. In *SIAM International Conference on Data Mining (SDM14)*, 2014.

[5] See SmartD's source code, demo video, and supplementary material for this paper at `https://github.com/LSIR/smartd`