

SMAS: A Smart Meter Data Analytics System

Xiufeng Liu, Lukasz Golab and Ihab F. Ilyas

University of Waterloo, Canada

{xiufeng.liu,lgolab,ilyas}@uwaterloo.ca

Abstract—Smart electricity meters are replacing conventional meters worldwide and have enabled a new application domain: smart meter data analytics. In this paper, we introduce SMAS, our smart meter analytics system, which demonstrates the actionable insight that consumers and utilities can obtain from smart meter data. Notably, we implemented SMAS inside a relational database management system using open source tools: PostgreSQL and the MADLib machine learning toolkit. In the proposed demonstration, conference attendees will interact with SMAS as electricity providers, consultants and consumers, and will perform various analyses on real data sets.

I. INTRODUCTION

As conventional power grids turn into smart power grids, traditional electricity meters are being replaced by smart meters. The main operational advantage of smart metering systems is their ability to automatically collect fine-grained (typically hourly) electricity consumption data, which enables accurate billing without sending workers to customer premises to manually read the meters. Additionally, a new application area has emerged: smart meter data analytics. The vast amounts of data collected by smart metering systems can help understand electricity consumption patterns, thereby helping governments and utilities plan for the future and helping consumers reduce their bills. The market for smart grid data analytics is growing rapidly and is expected to reach over \$4 billion by year 2020¹.

Previous Work: Some database vendors currently offer smart meter analytics software, e.g., SAP² and Oracle/Data Raker³. There are also several startups in this area, e.g., C3Energy.com and OPower.com, and a recent research prototype system [5]. Furthermore, some utilities such as California's PG&E⁴ provide on-line portals where customers can view their electricity consumption and compare it with their neighbourhood's average. However, these systems and tools focus on simple aggregation and simple ways of visualizing consumption. On the other hand, a number of advanced smart meter mining techniques have been proposed, mostly in the smart grid and machine learning literature, but the focus has been on the insights that can be obtained from the data rather than algorithm and system implementation.

In this paper, we introduce our smart meter analytics system, SMAS, which is unique in at least two ways:

- 1) In addition to simple visualization and temporal aggregation, SMAS supports state-of-the-art smart meter data mining techniques, including those for extracting daily consumption profiles [2] and understanding the effect of temperature on consumption [3]. The patterns and features identified by these algorithms can be used for clustering similar consumers (which is useful for designing targeted customer engagement programs), forecasting electricity consumption and providing personalized feedback to consumers.
- 2) Rather than using separate systems for data storage and analysis (as was done in [5]), we implemented SMAS inside an open-source relational Database Management System. Specifically, we used PostgreSQL and the MADLib in-database machine learning toolkit [4].

In Section II, we describe the system architecture of SMAS and the supported analytics. In Section III, we outline the functionalities of SMAS that will be demonstrated, which will show the insights that can be obtained from real smart meter data sets. Conference attendees will interact with SMAS in three ways. As electricity providers, they will identify different types of consumers and predict future consumption. As electricity consultants, they will perform virtual audits to understand consumption patterns and find ways to save electricity. Finally, as electricity consumers, they will see how their consumption ranks against that of their neighbours, and they will obtain advice on how to change their consumption habits to lower their bills and reduce their carbon footprint. We will also demonstrate the privacy concerns that arise in smart meter analytics (see, e.g., [1]), such as being able to find out people's daily habits and work schedules.

By demonstrating SMAS, we hope to increase the awareness of the emerging application area of smart meter analytics within the data management community. We also hope to stimulate further research on this topic, including data privacy, and applying novel database technologies such as column-stores, in-database machine learning and distributed data processing to improve the efficiency and effectiveness of smart meter analytics. We plan to make SMAS freely available online for others to use and extend.

II. SYSTEM OVERVIEW

Figure 1 shows the architecture of SMAS. The real-time layer consumes raw smart meter data and runs simple alerting queries, such as those which detect very high consumption readings. The batch layer packages raw data into chunks that are batch-loaded into the analytics database, e.g., once an hour

¹<http://www.greentechmedia.com/research/report/the-soft-grid-2013>

²<http://www.sap.com/pc/tech/in-memory-computing-hana/software/smart-meter-analytics/index.html>

³<http://www.oracle.com/us/products/applications/utilities/meterdata-analytics/index.html>

⁴<http://www.pge.com/myenergy/>

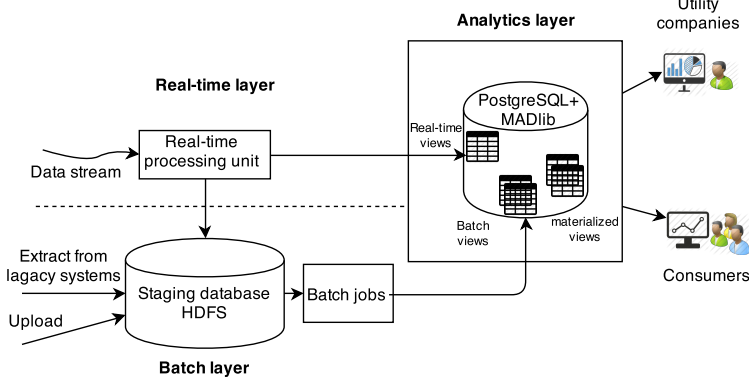


Fig. 1. SMAS system architecture

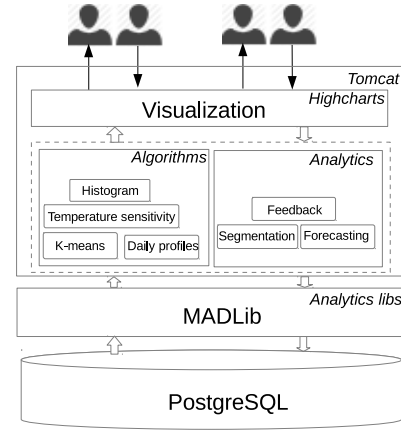


Fig. 2. SMAS analytics layer

or once a day. The batch layer also accepts bulk-uploads, e.g., historical consumption data from legacy systems. The analytics layer, which is the focus of this paper, consists of an underlying database, analytics libraries, and a Web front-end. In addition to the consumption time series generated by smart meters, the database also contains the location of each consumer and hourly temperature data obtained from a local weather station. Thus far, we have implemented the analytics layer; we envision implementing the real-time layer using a data stream engine and the batch layer using a database system or Hadoop/Hive.

Figure 2 shows the analytics layer, which uses PostgreSQL as the database and MADLib as the (in-database) machine learning library. Furthermore, we use Highcharts (www.highcharts.com) as the visualization engine, and Tomcat as the web application server. The main functionalities of SMAS include consumer feedback, segmentation (clustering) analysis and forecasting. Each of these tasks requires a set of features and patterns to be extracted from the data (which are then stored in materialized views). We employ three algorithms for feature extraction, as detailed below.

A. Algorithms

First, for each consumer, we compute a histogram of its hourly consumption to understand its variability. The electricity grid must be provisioned for peak demand, therefore it is important for utilities to identify consumers with highly variable consumption and offer them incentives to smooth out their demand.

Next, we characterize the effect of external temperature on the electricity consumption of each customer. In the winter and summer, consumption rises as temperatures become more extreme due to heating and cooling. Consider the scatter plot shown in Figure 3, with temperature on the X-axis and consumption on the Y-axis. Each point on the scatter plot corresponds to a particular hourly consumption value and the corresponding temperature at that hour (for the same customer). We implemented a recent algorithm [3] that computes a piecewise linear regression model to characterize

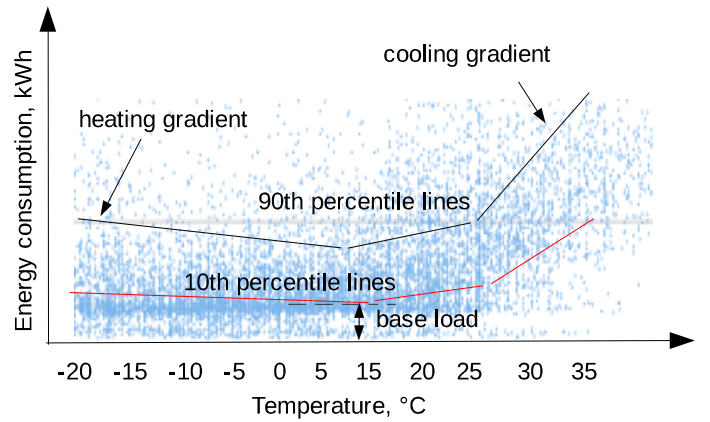


Fig. 3. The piecewise linear regression model

the temperature sensitivity of each customer. The algorithm computes two regression models: one corresponding to the 90th percentile consumption at each temperature and one corresponding to the 10th percentile at each temperature. As shown in Figure 3, these models reveal several interesting features for each customer. For example, the slope of the 90th percentile line corresponding to high temperature is the *cooling gradient*, and the slope of the line corresponding to low temperature is the *heating gradient*. Furthermore, the height of the 10th percentile lines at their lowest point is the *base load*, which corresponds to load due to appliances that are always on, such as a refrigerator.

The third algorithm extracts daily consumption patterns that occur regardless of the external temperature [2], as illustrated in Figure 4. On the left, we show a fragment of the hourly consumption time series for some consumer over a period of several days. Since smart meters report the total electricity consumption of a household, we can only observe the total consumption time series (the upper black curve). The goal of the algorithm is to determine, for each hour, how much load is due to temperature (i.e., heating and cooling), and how much load is due to daily activity independent of temperature (the lower blue curve). Once this is determined, the algorithm fits

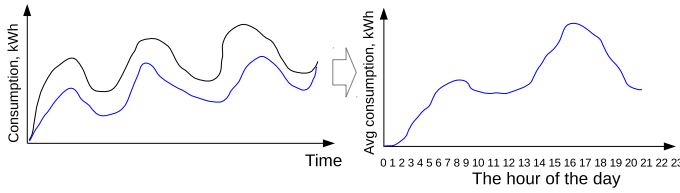


Fig. 4. Extracting daily consumption profiles

a time series auto-regression model and computes the average temperature-independent consumption at each hour of the day, illustrated on the right of Figure 4 (the X-axis is the hour of the day and the Y-axis is the average consumption). Since weekday and weekend activities may differ, it is useful to separately compute the typical weekday and weekend profiles for each consumer.

B. Analytics

We now describe how the features and patterns identified by the above algorithms enable the three functionalities of SMAS: consumption feedback, segmentation and forecasting.

Feedback is important to consumers and is generated from the features extracted by the above algorithms, including variability, temperature sensitivity and typical daily profiles. We implemented a rule-based interface, in which utilities and consultants express rules for generating feedback that consumers will see when they log in. A simple rule may be: if the average highly consumption (regardless of outdoor temperature) is above some threshold, then produce the following recommendation – “Your household appears to have an unusually high nightly consumption. Consider replacing your refrigerator or another appliance that is always on with a more energy-efficient one”. Another straightforward example is: if the cooling gradient is above some threshold, then output “Your electricity consumption increases significantly when it is hot outside. Consider turning up the air conditioning set point on your thermostat and have your air conditioner serviced”. Additionally, for jurisdictions with time-dependent electricity pricing such as the province of Ontario, if the daily profile shows high consumption during peak-pricing times, we can recommend shifting some activities to off-peak times.

Segmentation and forecasting are important to utilities. SMAS supports K-means clustering based on the features we discussed earlier, namely base load, heating and cooling gradients and the daily profiles. For example, clustering on daily profiles can reveal groups of consumers with similar lifestyles, and the utility may design targeted energy-saving programs for each group. Moreover, SMAS can use the temperature dependence and daily profile, along with a temperature forecast, to predict future demand. Additionally, we implemented several time series forecasting algorithms, including ARIMA and Holt-Winters.

In future work, we plan to add more algorithms and analytics to SMAS. However, we believe that the algorithms and analytics we currently support are a representative set and can

illustrate the kinds of insights that can be obtained from smart meter data.

III. DEMONSTRATION SCENARIOS

The conference attendees will interact with SMAS as consultants, utilities and consumers. They will explore the insights that may be obtained from smart meter data and they will be able to examine the algorithm implementations in PostgreSQL/MADLib. Throughout the demonstration, we will also point out various data privacy issues arising from smart meter analytics. We will use two real smart meter data sets: a publicly-available data set from the Irish Social Science Data Archive⁵ and a (anonymized) 10 Gigabyte data set corresponding to nearly 30,000 consumers from several cities in the province of Ontario, Canada.

A. Electricity Provider’s Point of View

An attendee is given a provider account, and logs in to SMAS. In the “Consumption analysis” report, she can select a neighbourhood or a whole city, and plot the total consumption in that area at different time granularities. Next, she clicks on the “Segmentation analysis” tab to view the clustering results; for the demonstration, we will prepare several clustering outputs using different feature sets. These results can be viewed in two ways. First, she can examine the cluster centroids to understand different groups of customers. For instance, if the clustering is on the daily consumption profiles, the centroids correspond to different activity patterns. Second, the customers may be viewed on a map, with different colours denoting different cluster membership, making it easy to identify homogeneous and heterogeneous neighbourhoods. Finally, she clicks on the “Forecasting” tab to predict the energy consumption of an individual consumer, a neighbourhood, or an entire city.

B. Electricity Consultant’s Point of View

An attendee is given a consultant account and logs in to examine a particular customer in detail. In the “Pattern discovery” view, she selects a consumer and views the histogram, temperature regression lines and daily profiles (for weekdays and weekends) of this consumer (see Figure 5; the regression lines are shown on the top-left, the histogram on the top-right and the daily profile at the bottom). This gives the consultant enough information to understand the customer’s habits and formulate suggestions for saving electricity. Some of these suggestions may be elevated to rules, which the consultant can specify in the “Feedback service” screen; after a rule is specified, it is automatically applied to all customers that satisfy it.

C. Electricity Consumer’s Point of View

An attendee is provided a consumer account, and logs in to SMAS as that particular consumer. In the “My consumption” view (see Figure 6), she browses her electricity consumption aggregated at different granularities (hourly, daily, etc.). In

⁵<http://www.ucd.ie/issda/data/commissionforenergyregulationcer/>

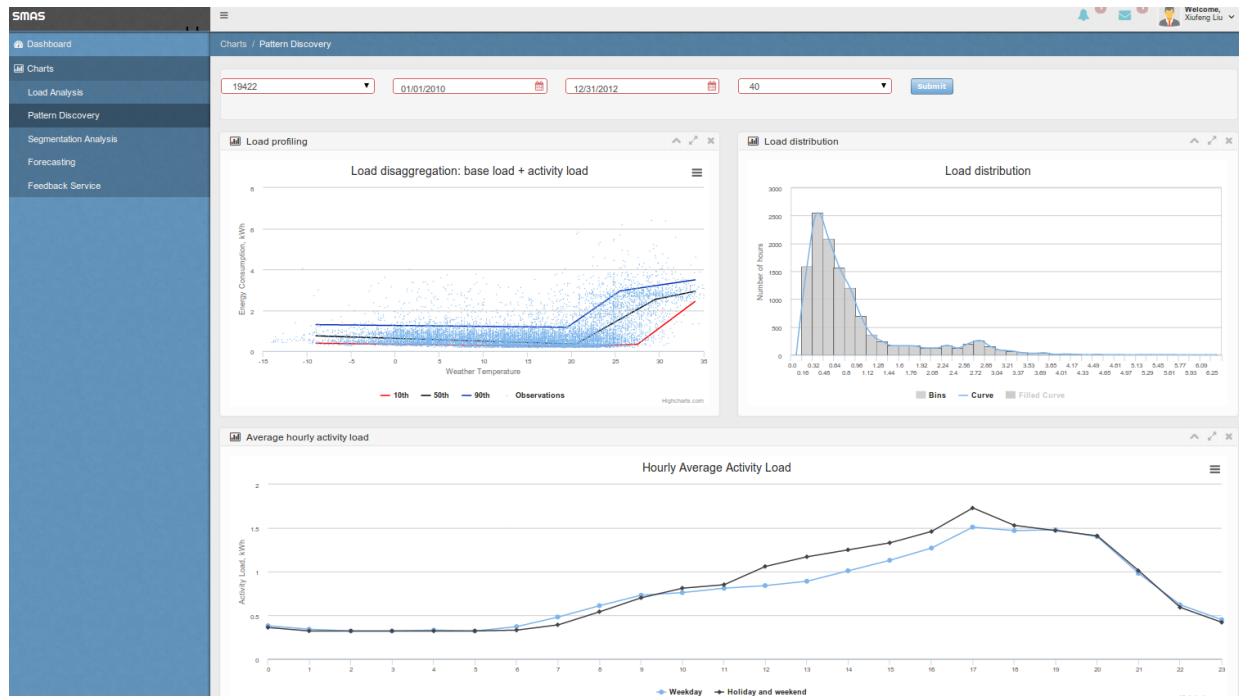


Fig. 5. A screenshot of consumption pattern discovery

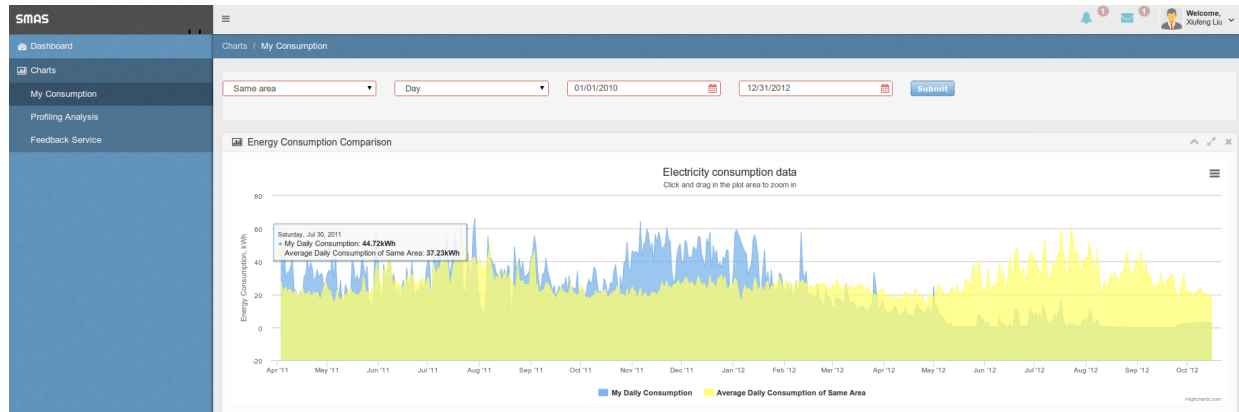


Fig. 6. A screenshot of consumption analysis

addition to her consumption, SMAS can show the *average consumption in her neighbourhood* over the same time interval. Next, she chooses the “Feedback service” view and browses personalized recommendations for saving electricity and reducing her bill; these correspond to all the rules, entered earlier by the utility or a consultant, which apply to her based on her consumption features. SMAS additionally provides comparative feedback via ranking: what is this consumer’s rank within the neighbourhood and within the whole city in terms of overall consumption, base load, heating gradient and cooling gradient? We will pre-select several representative consumers and show the feedback generated by SMAS under various circumstances.

REFERENCES

- [1] G. Acs and C. Castelluccia, I have a DREAM (DiffeRentially private smArt Metering), in Conf. on Information Hiding, 118-132, 2011.
- [2] O. Ardakanian, N. Koochakzadeh, R. P. Singh, L. Golab, and S. Keshav, Computing Electricity Consumption Profiles from Household Smart Meter Data, in Proc. of the EDBT Workshop on Energy Data Management (EnDM), 140-147, 2014.
- [3] B. J. Birt, G. R. Newsham, I. Beausoleil-Morrison, M. M. Armstrong, N. Saldanha, and I. H. Rowlands, Disaggregating Categories of Electrical Energy End-use from Whole-house Hourly Data, Energy and Buildings, 50:93-102, 2012.
- [4] J. M. Hellerstein, C. Ré, F. Schoppmann, D. Wang, E. Fratkin, A. Gorajek, and A. Kumar. The MADlib Analytics Library: or MAD Skills, the SQL. PVLDB, 5(12):1700–1711, 2012.
- [5] A. J. Nezhad, T. K. Wijaya, M. Vasirani, and K. Aberer. SmartD: Smart Meter Data Analytics Dashboard, in Proc. Int. Conf. on Future Energy Systems (ACM e-Energy), 213-214, 2014.