Data Analytics for Energy Efficiency: Opportunities and Challenges

Shyam Boriah and Badri Raghavan
FirstFuel Software
420 Bedford Street, Suite 200
Lexington, MA 02420
<sboriah,braghavan>@firstfuel.com

ABSTRACT

Total carbon dioxide emissions from the consumption of energy have been steadily rising and are now roughly 30 billion metric tons annually. Increasing demand for energy in developing countries is expected to further increase the annual rate of emissions. While there are ongoing efforts to develop alternative sources of energy, the magnitude of the issue requires scalable solutions that can be deployed in the near term. Energy efficiency (driving down energy consumption) in the commercial and industrial (C&I) real estate sector is a particularly favorable solution that can lead to significant and sustained reductions in emissions over time, with the added economic benefit of lowering costs of goods and services. In particular, recent advances in measurement (e.g. smart meters) and data analytics technology have given us new insights into commercial and residential energy consumption. In this paper, we discuss opportunities where data science can have an impact in this area, in the context of specific data analytics challenges and FirstFuel's experiences.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data mining*

General Terms

Machine learning; energy efficiency; commercial real estate.

1. INTRODUCTION

The global demand for energy has grown substantially over the past several decades. While energy demand is often driven by economic growth, it has also had the detrimental consequence of increasing carbon dioxide emissions. In fact, total carbon dioxide emissions from the consumption of energy have been steadily rising and are now roughly 30 billion metric tons annually (see Figure 1). The increasing demand for energy in developing countries is expected to further increase this rate of emissions. Policymakers and stakeholders at the national and international level are seeking ways to manage and control both energy production and consumption [1, 2].

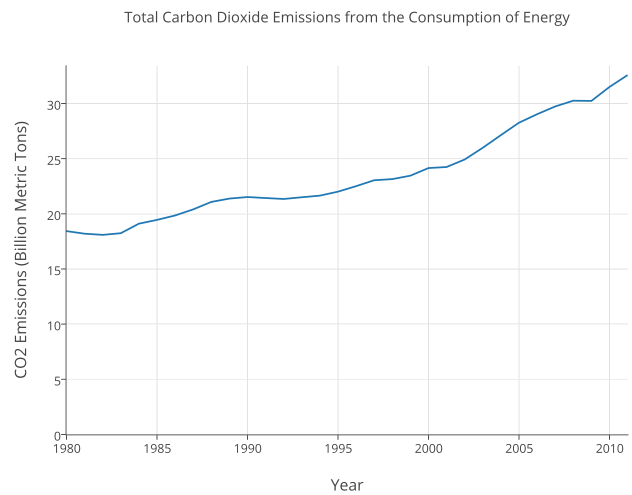

Figure 1: Total global carbon dioxide emissions from the consumption of energy (1980–2011). ***Source:*** **U.S. Energy Information Administration [5].**

While there are ongoing efforts to develop alternative low-carbon sources of energy, the magnitude of the issue requires scalable solutions that can be deployed in the near term. Energy efficiency (driving down energy consumption) in the commercial and industrial (C&I) real estate sector is a particularly favorable solution that can lead to significant and sustained reductions in emissions over time, with the added economic benefit of lowering the cost of goods and services (see Figure 2 for a breakdown of total energy use by sector in the United States). In particular, recent advances in measurement (e.g. smart meters) and data analytics technology [4] have given us new insights into commercial and industrial energy consumption.

Interval meters are electronic devices that measure the consumption of energy (and often other variables) at regular intervals (see Figure 3 for an example). These measurements are either stored on the device (to be retrieved at a later time), or directly communicated to a central repository. The combination of interval meter data and supple-

KDD'14 Workshop on Data Science for Social Good, August 24–27, 2014, New York, NY, USA.
ACM 978-1-4503-2956-9/14/08.

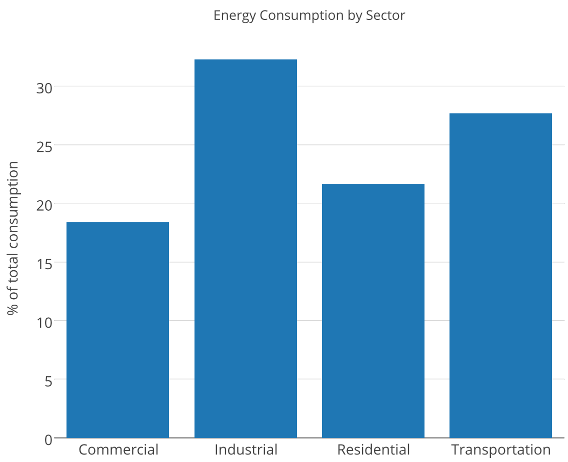

Figure 2: Energy consumption by sector in the United States. The commercial and industrial (C&I) sector account for approximately 50% of total consumption. *Source:* U.S. Energy Information Administration [6].

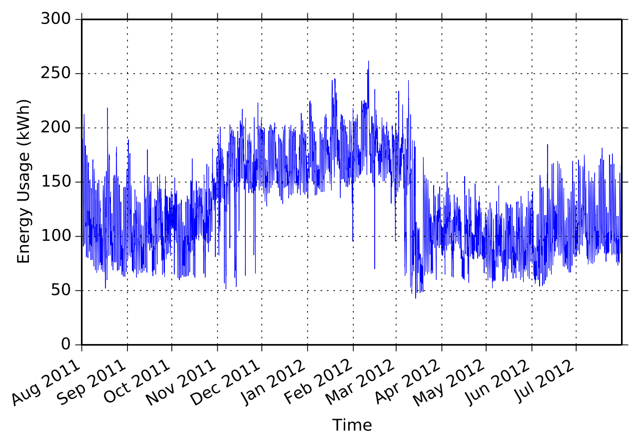

Figure 3: Smart meter usage data (in kWh) for a small suburban office building in the midwest U.S. This meter reads usage data every 60 minutes, yielding 8760 observations per year for the building. We see that there is increased usage during the winter months.

mentary sources of information—including structured public databases, social media, and unstructured web data—offers new opportunities where analytics-driven insights can offer powerful and scalable solutions. The term *smart* meter refers to meters that are more sophisticated than interval meters, often capable of two-way communication, etc. In the remainder of this paper, we use the terms smart meter data and interval data interchangeably.

Advances in data analytics technology have given us new insights into commercial energy efficiency, both at the individual building and at the portfolio level. These insights have real implications for how energy utilities can deliver cost-effective large scale efficiency in the coming years. For example, FirstFuel's Remote Building Analytics (RBA) platform identifies, targets, and monitors energy efficiency savings in commercial real estate portfolios using utility meter data. Since 2010, the company has remotely analyzed and tracked thousands of commercial buildings for utility and government customers across all major segments and geographies in North America. In this paper, we discuss opportunities for data science to impact the area of energy efficiency, particularly in the C&I sector. Note that while our goal is to capture the rich variety of emerging data science problems in energy efficiency, the set of problems presented in not intended to be comprehensive.

2. OPPORTUNITIES FOR DATA SCIENCE

2.1 Regression & Classification

Regression and classification problems are frequently encountered in the area of energy efficiency, both at the building level and the portfolio (collection of buildings) level. For example, regression is often performed in order to disaggregate a specific building's total hourly energy usage into weather-related and occupancy-related components, using a suitably constructed feature space. Insights derived from this type of regression analysis can then be used to improve (in efficiency terms) how the building is operated in various weather conditions. Other common tasks include using classification or regression to rank/target the top-k buildings in a portfolio, based on potential savings that can be achieved through operational changes (using a historical database of building data for training). Another portfolio level ranking task is to identify building operators that have the highest propensity to act on targeted energy efficiency programs, again based on historical patterns of behavior.

2.2 Unsupervised/Semi-Supervised Learning

There are a wide variety of unsupervised and semi-supervised learning problems that arise in the analysis of energy consumption data. These problems range from clustering to anomaly/change detection to time series forecasting. For example, a common semi-supervised learning problem is automatically classifying the type of business occupying a building based on limited information about the building.

Given a building, identifying similar buildings (peers) within a database that are expected to have similar energy usage is another example of a semi-supervised problem.

Time series forecasting is extensively used with energy usage data; two examples are monitoring and verification (M&V) and peak demand forecasting. M&V refers to the process of *monitoring* a specific building's usage data for unusual/unexpected behavior, and *verifying* (and sometimes quantifying) savings achieved through operational changes or retrofitting.

Demand charges are a component of energy billing that is based on the highest capacity required by the end user. Thus, both C&I building operators and utilities seek to forecast periods of "peak demand" so that measures can be taken ahead of time to minimize energy use during the period.

Buildings sometimes undergo major changes; these could be events such as a tenant moving out, or anomalies such as a malfunctioning piece of equipment (some of these changes can be seen in Figure 4). Thus, there can be both gradual and abrupt changes in the meter data. Detecting and characterizing these types of changes based on interval data is a

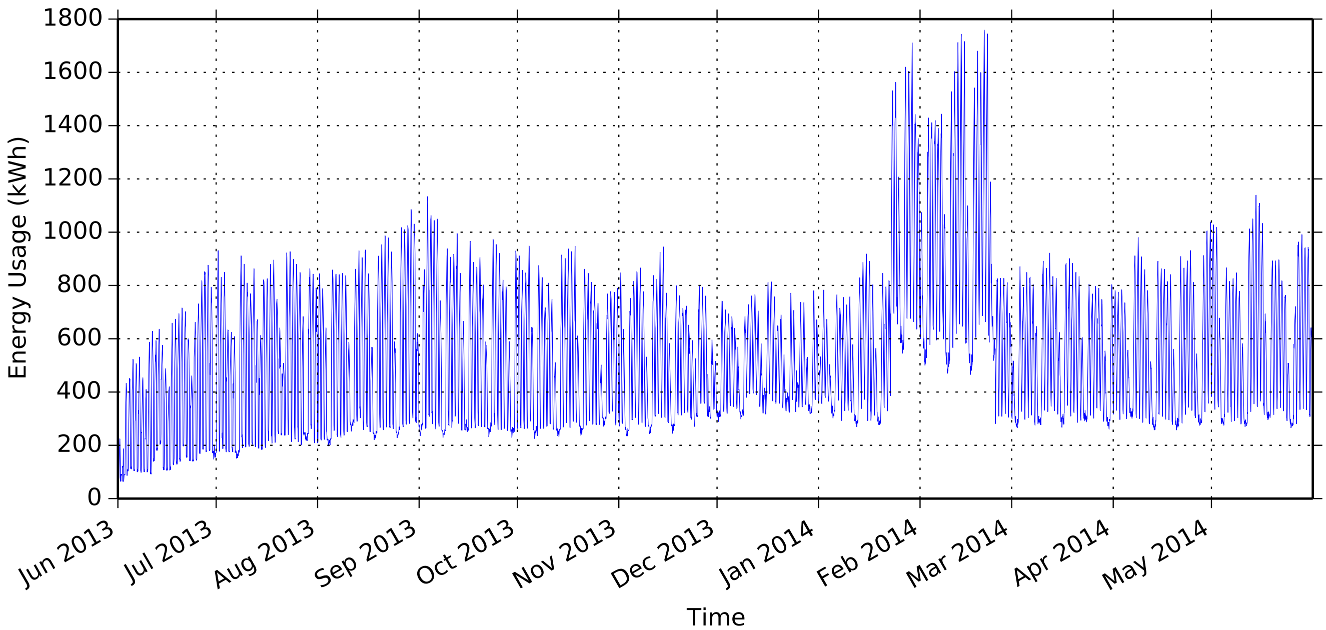

Figure 4: Smart meter usage data (in kWh) for a mid-size mixed office/light industrial building in Southern California. Note the gradual ramp-up in usage at the beginning, and the abrupt shift in usage during February 2014.

very important problem in the domain. This problem can be posed as the time series change detection problem (also called breakpoint detection and regime change detection).

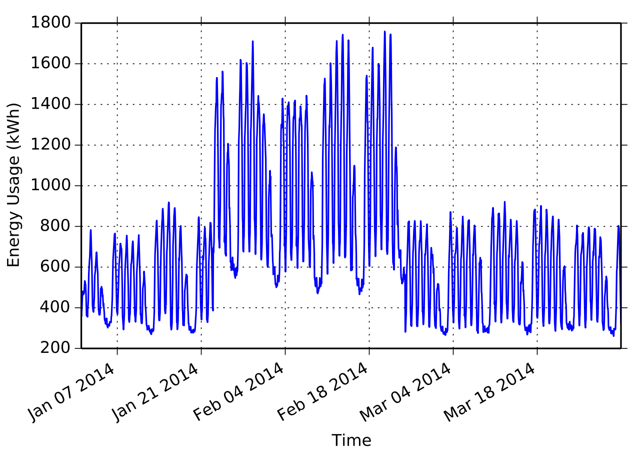

Figure 5: A portion of usage data for the building shown in Figure 4.

3. CHALLENGES

The previous section illustrates the wide variety of machine learning problems encountered in energy efficiency. As in all real-world application areas, there are numerous challenges that must be overcome in order for data analysis to be effective. For example, while interval data is arguably the most critical piece of information on a building, its quality is not always uniformly high. Interval data can be corrupted by *noise*, *outliers*, and *missing* values (this can be seen in Figure 3); the intervals may also not be evenly spaced. Furthermore, as we can see in Figures 3 and 4, these time series are often *non-stationary*, which poses challenges for predictive modeling.

The population of C&I buildings is tremendously *heterogeneous*, i.e. buildings that share similar physical characteristics are often used and operated completely differently. Conversely, it is possible for two similar interval time series data sets to come from two very different buildings (in terms of size, year of construction, etc.). Heterogeneity has become an area of interest in machine learning in recent years [3, 7, 8], and these new developments may have potential applicability in the energy efficiency domain.

Finally, as in other areas involving the analysis of individually identifiable data, *security* and *privacy* concerns must be taken into account when developing machine learning algorithms for energy data analytics.

4. CONCLUDING REMARKS

In this paper, we presented a variety of interesting and challenging machine learning problems that arise in the area of energy efficiency, particularly in the commercial and industrial sector. These problems provide an opportunity for the data science community to impact this application area, with significant societal, environmental and economic benefits.

References

[1] H. Geller, P. Harrington, A. H. Rosenfeld, S. Tanishima, and F. Unander. Polices for increasing energy efficiency: Thirty years of experience in OECD countries. *Energy Policy*, 34(5):556–573, 2006.

[2] K. Gillingham, R. G. Newell, and K. Palmer. Energy efficiency economics and policy. Working Paper 15031, National Bureau of Economic Research, 2009.

[3] P. Gong, J. Ye, and C. Zhang. Robust multi-task feature learning. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, pages 895–903, New York, NY, USA, 2012. ACM.

[4] J. Z. Kolter and J. Ferreira. A large-scale study on predicting and contextualizing building energy usage. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2011.

[5] U.S. Energy Information Administration. International energy statistics, . URL `http://www.eia.gov/cfapps/ipdbproject/IEDIndex3.cfm`.

[6] U.S. Energy Information Administration. Monthly energy review, . URL `http://www.eia.gov/totalenergy/data/monthly/index.cfm`.

[7] X. Yang, S. Kim, and E. P. Xing. Heterogeneous multitask learning with joint sparsity constraints. In *Advances in Neural Information Processing Systems*, pages 2151–2159, 2009.

[8] D. Zhang and D. Shen. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease. *NeuroImage*, 59(2): 895–907, 2012.

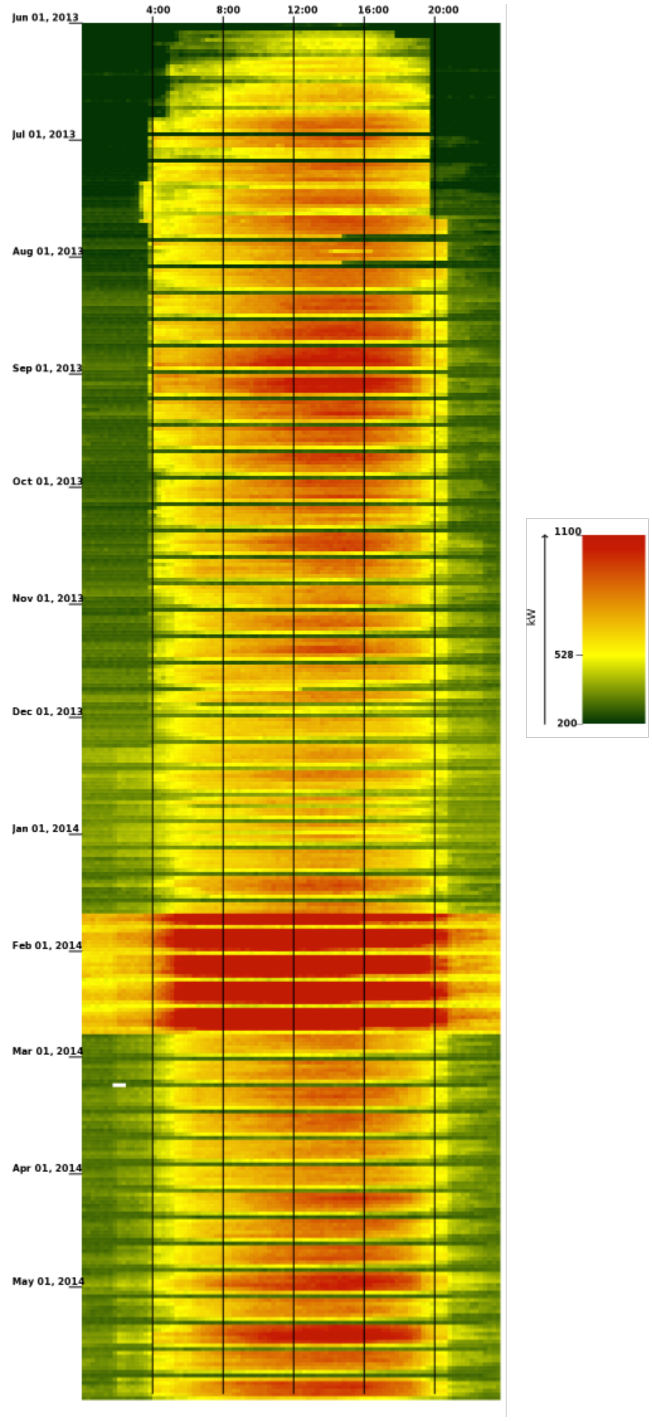

Figure 6: Annual demand intensity chart for the building shown in Figure 4. Each row in this chart depicts the power consumption for every metering interval through a single day. Columns correspond to specific metering intervals. Red indicates intervals of high intensity, green indicates intervals of low intensity, and yellow indicates medium intensity. Note the presence of distinct patterns and shifts across time of day, weekdays and holidays.