Position statement: Data analytics for energy systems

J. Zico Kolter
School of Computer Science
Carnegie Mellon University

November 1, 2013

The increasing amount of data we have available from the energy systems — growing due to the increasing number of sensors such as PMUs, the ease of transmitting and storing this data, and the perceived potential of "big data" analysis — raises the immediate question about how we can use data analytics to increase the reliability, efficiency, and sustainability of the existing energy infrastructure. Although data has long played a role in certain grid systems (forecasting, markets, etc all use data to build their models), for the most part power systems have focused on a largely "engineering" approach: using first principles laws to develop simulations and traditional control methods to optimally control the resulting systems. My research looks at developing and applying new machine learning techniques that we may use to analyze and process this data, with the ultimate goal of increasing the efficiency of the grid.

As a concrete example, one of my research projects look at understand load end-use through individualized by relatively low-frequency monitoring systems like smart meters (which are installed in over 40 million homes). Load disaggregation has a long history of research in power systems, but most techniques require customized sensors to be installed in homes in order to understand load. In contrast, the work in my group looks at matrix factorization methods for separating the large components of a load using just the available smart meter data, with the ultimate goal of better characterizing the true instantaneous "cost of demand" for a small region or even at a single home level. The ultimate goal is to incorporate real-time smart meter data to build such costs in an online manner, so that system operators can use this data to better understand current grid conditions and make control decisions.

Although this is a fairly narrow example, I believe it points to a larger paradigm for future operation of cyber-physical energy systems. Rather than use fixed control and estimation procedures, the systems can make increasingly more use of large-scale machine learning methods to adaptively update and modify their internal parameters, ultimately leading to a grid that is substantially more adaptive and controllable. A chief obstacle here is how we can guarantee systems stability and performance even with the inclusion of such systems; such considerations have a long history in adaptive control, but the difficulties are compounded in cases of using more "black box" probabilistic machine learning techniques). Solving such problems requires new algorithms at the boundary of control, optimization, and machine learning research, and has the potential to drastically change how we incorporate and use data in energy systems.