Applications of Energy Data Compression in Smart Cities

Charlotte Gehrs, Class of 2024

As homes and buildings with smart meters become more prolific, utility companies may be overwhelmed by the resulting volume of energy data. Smart meters can collect data as often as multiple times a second from each individual electric device in a building (such as lights, dishwashers, and washing machines). If a single home's smart meter collects energy output data from ten devices once every second, one can imagine how the number of datapoints would be very large over just a single day. Utility companies are interested in datasets from many buildings (even cities) over the course of weeks or months. This huge volume of data is very difficult to store, exchange, or work with. One solution to this problem is to compress this energy data. Developed at Bowdoin by Professor Sean Barker and John Ward '19, Powerstrip is a tool that compresses energy output data by as much as 97%, which allows previously unwieldy data to be stored, shared, and studied. Unlike general compression algorithms, Powerstrip was created specifically to compress energy data. For my Gibbons Fellowship project, I studied how different types of real-world energy output data behave when compressed through Professer Barker's existing Powerstrip tool.

Using real-world datasets of smart buildings, I explored the relative compression rates of different types of devices, relative compression rates of different devices when they are active, and Powerstrip's effectiveness with high-frequency, fractional data. First, we wanted to see which types of devices compressed well compared to those that did not. To do this, I used real-world datasets from smart buildings (namely, the REDD and Dataport datasets). These datasets are grouped by different buildings and the individual devices in each building. I ran the REDD and Dataport datasets through Powerstrip, and found that washing machines tended to compress well, while solar panels compressed poorly (because they can have a negative energy output when they generate energy). Yet, one significant component of Powerstrip's compression is recognizing and compressing inactive periods found in data files. Thus, devices that are off for long periods of time, such as washers and dryers, would naturally have very high compression rates. While it is useful to see the relative compression rates of devices in their typical use, we also wanted to compare the compression rates of devices when they are being used, without inactive periods skewing the compression rates. I created copies of the data files with inactive periods removed, and ran those files through Powerstrip again to generate new statistics. Lightbulbs tended to compress very well when compared to other active-only devices. Finally, I worked with a different dataset to see how effective Powerstrip is with high-frequency, fractional data. Powerstrip is designed to work with device-level data recorded (at most) every second. Additionally, it only works with integer data. The UK-DALE dataset contains very large files with fractional data collected at 16 kHz. Powerstrip was able to compress UK-DALE data at a compression rate of about 1.3, compared to compression rates of 16.8 for REDD and 15.2 for Dataport. Thus, while Powerstrip can compress high-frequency data, it is far more effective in compressing lower-frequency datasets.

Faculty Mentor: Sean Barker

Funded by the Gibbons Summer Research Program