Visualizing Events for Large and Heterogenous Power Consumption Data

- Prototype Design -

Christopher Gress, Sundar Sekar, and Sung Ho Kim,

Advisor - Ting Zhu and Jian Chen

Abstract—We make a visualization tool to explore time-series data from multiple real-time sources for trends acrossed the sources. The visualization provides an overview of aggregated data over different time scales, a query interface to find similar events, and a method to view data at a particular time. Together these views can be used to identify normal events in the data, which can help develop predictive models. Our contributions include a methodology to show multiple time-varying data streams to find trends.

1. Introduction

We use electrical energy in our life such as cooking, heating, lighting, and TV. As time goes by, lots of electronic devices are used, and the energy consumption increases. However, that means we are damaging our environments because the generation of electric power makes environmental pollution. Fossil fuel power plants as well as renewable energy facilities affects our environments. Major environmental issues linked to electricity are water pollution, generation of waste, and the destruction of the land. To minimize the impact, conserving energy is required. Among various ways to save the energy in our daily life, it is important to know our power consumption at home.

Our clients want to have a visualization tool to explore and analyze power consumption data collected from a few houses for a year. The tool should assist in identifying repeating patterns in the data on individual circuits and relationships between power consumption on separate circuits. This exploration can aid their work in determining normal power consumption events, and more importantly, what anomalous power consumption events look like. Additionally, the visualization may help identify ways to reduce the dataset into a more manageable size, which decreases the storage requirements for the dataset.

A secondary goal of this data, is to provide metrics of power consumption for appliances and devices that are connected on the circuits in the house. Without collecting metrics, we can not know if changes made to power consumption habits have a positive or negative effect. For power conscience consumers, the visualization could help identify what is drawing the most power in their house.

We made a visualization tool that aggregates individual power readings into events, providing the ability to explore these events. Viewing the metrics at the circuit level reduces the number of possible devices that contribute to the power use, making it more likely to determine what appliances are contributing to the power consumption. This can help our customer reduce the data set down even more by identifying what devices are associated with power usage, turning it into categorical data instead. There are three views provided to achieve the goals of the project: the piano roll, the histogram for both events by average KW and the duration of events, and the circuit levels for the loads of each circuits at certain position.

2. Literature Survey

Visualizing History of Living Spaces provides a multi view interface that includes a path based querying technique and multiple sensor sources in what they call a piano roll. The piano roll has a row for each circuit and includes a timeline horizontally. The piano roll contains a bar on the row if the sensor was active for that time period. The query interface involved a path between nodes that were spatially linked with each other, returning paths that have been traversed by people.

Cluster and Calendar Based Visualization of Time Series Data demonstrates a technique to find similarities of time-series data acrossed a certain timeframe, and generate an average curve representing each cluster. This data provides overall trends at a high level, but is unable to show data at the circuit level. The clustering techniques used for the total power consumption represent a large building, and not an individual home. When analyzing our data we determined that most power consumption on a circuit was consistent in power consumption when devices were on.

3. Design

3.1 Task Analysis

Before discussing the tasks provided by our client, we would like to briefly discuss the current tool that they use for visualizations. The tool shows a live view of the total power usage as a connected line graph. Time is on the x-axis and kilowatts are on the y-axis. Individual circuits can be plotted on this as well, but viewing all circuits at once is confusing due to the overlapping lines and occlusion.

Our client has provided us with a number of tasks that they would like to achieve with the visualization. We will discuss all tasks that were identified in this section, and in section 3.2, we will discuss the tasks that we were able to implement in the visualization with the limited time we had.

3.1.1 Task 1: Explore

The first task was to provide the capability to explore the data to get an overall sense of the distribution of data. Through exploring the data, it is possible to find potential relationships between circuits. Certain household activities where one would expect these relationships include vacuuming, where a person moves from room to room plugging and unplugging the device into different outlets.

Data exploration should also help find periodic events that occur on a particular circuit at daily, and weekly time scales.

3.1.2 Task 2: Search By Example

This task is meant to help identify sequences of events that are repeated over time. This would let the user identify a particular pattern and search by example for other instances of that pattern. Based on the frequency found in the dataset, this would help identify repeated events.

3.1.3 Task 3: Aggregate Data

This task is to identify a means to reduce the dataset to a more manageable scale. The data received from the client was only a week's worth of data, but comprised 800,000+ records, with each record containing a reading for each circuit. Extrapolating this out to a year, would be nearly 50 million records, which makes it very difficult to find relationships.

3.1.4 Task 4: Relate to Consumer Cost

The client though it would be interesting to include a price model throughout the day/week. Since power companies will vary their rate throughout the day, charging more during times of peak usage. This could potentially appeal to consumers interested in monitoring their power usage while they are attempting to reduce their power bill.

3.1.5 Task 5: Find Timeframes Circuits Are Underutilized

A complement to the task of reducing consumer cost, is determining how one may do so. If possible, devices used during peak usage, could be moved to underutilize times during less expensive times.

Another reason this could be useful, is if a household finds it is tripping circuits often. The consumer can find another circuit that the device can run on that doesn't cause the circuit breaker to trip.

3.1.6 Task 6: View Current Load Of Circuits

The visualization should show the overall trend of power usage on the circuit as well as the usage at a particular moment in time.

3.1.7 Task 7: Filter Events

The visualization should allow the user to filter out circuits from the view, allowing the user to focus on specific patterns, without being distracted by other events out of the current scope of interest.

3.1.8 Task 8: Annotate and Label Event Selections

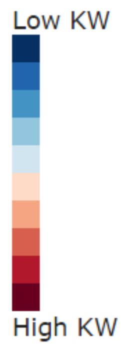
This task is focused on labeling event patterns found through the other tasks. These labels could be used to further reduce the dataset down, by showing the devices running instead of the power usage.

3.2 Prototype Design

The tasks that we have chosen to address with our visualization include exploring the data to identify repeating patterns in the data, either within a particular circuit, or between circuits, filtering events by each circuit, aggregating data, viewing current load of circuits, and partially, searching by example.

A pattern may include multiple events, with each event having characteristics such as average power consumption, average voltage, and duration, and each event separated by some time delay.

3.2.0 Legend



The colors are not linearly spaced. The blue colors represent the 0% or less of the highest observed power consumption, and the red colors range from 10% to 100%.

3.2.1 Task selection in our implementation

For the tasks above, there are three interfaces: the piano roll view, the histogram view, and the circuit levels view. Some tasks are only achievable through the interactions between these views. Each view shows some information of all circuits. If the user wants to compare or to analyze only some selected circuits, the user can choose specific circuits in the show/hide circuits view (Fig. 1). As the user selects circuits, every view changes promptly. This is for the task 7 as well.

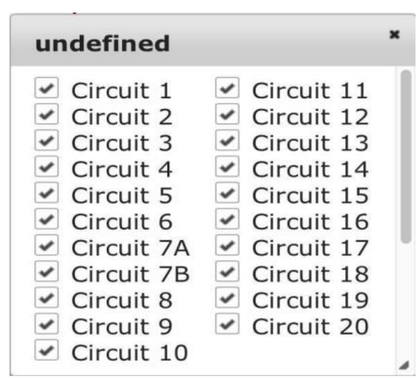


Fig.1 User interface - Show/Hide Circuits. This allows the user compare or to analyze only some selected circuits.

3.2.2 Piano Roll

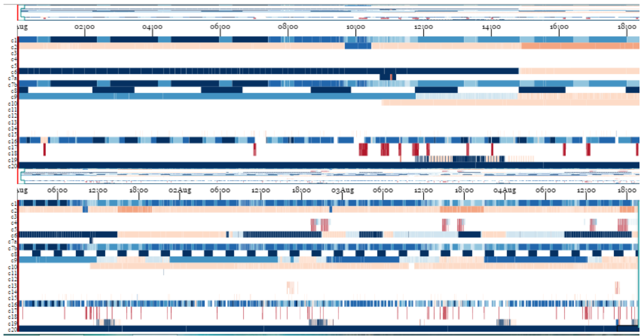


Fig 1.2 User interface- Piano Roll. This interface features an overview of a certain time range of data. Users can create two piano roll, once data for different weeks are loaded. This allows for visualizing events and patterns.

To address the first task, exploring data, we designed the piano roll interface (Fig. 1.2).

This visualization interface was inspired by the Visualizing the History of Living Spaces paper. The y-axis has each circuit and the x-axis represents time. Each line is colored by changing hue. Comparing with the maximum value of the corresponding circuit, the color of the circuit changes from green to red as the value changes from low to high. The white color represents nearly zero usage, the green for low usage, yellow for medium usage, and the red for high usage. The color representation is less accurate than position would be, but makes it easier to spot the areas of heavy power usage.

In order to handle the data for up to a month, we found that the raw data as received was not conducive to building the visualization we wanted. To address Task 3, we clustered data into power events. In order to keep the time complexity down, we decided on a streaming algorithm. The algorithm used the following criteria to determine if the events are related: same circuit, consecutive timestamps, within 10 watts of a running average. A new event is started as soon as a data point outside of the 10 watt range is found. Events have the following attributes: avg kilowatt usage, start time, end time, and circuit. We were pleasantly surprised to find that we reduced the dataset for the week into approximately 100,000 events, which considering an event only considers one circuit, it reduced the dataset by a factor near 160.

This features an overview of a certain time range of data (e.g. day, week, month, etc.), compared to another time-range of data (could be the same time range). There are two piano rolls at the bottom for this. Each piano bar has its own slider bar, allowing the user to line up the start or end of a time-shifted event. The user can zoom in or zoom out using this slider, and the scale of the data is changed. The slider bar has a bird's eye view of all the data in the range, showing where individual power events are.

There is a red vertical line through each piano roll that shows the current timestamp of the circuit level view. As the data is scrolled over, the circuit levels view changes (See 3.2.4).

3.2.3 Histogram view

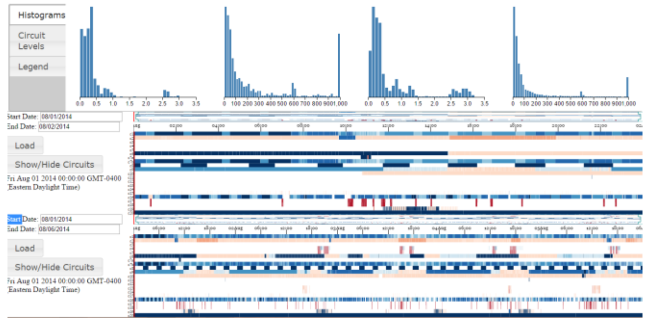


Fig.2 User interface - Main window with Histograms. The histogram view is generated once the user loads the data.

To address task 2 partially, we added the histogram view (Fig. 2).

At the top, there are two tabs: the histogram tab and the circuit levels tab. The histogram tab has four sections. The first two from the left are for the top piano roll. The second two are for the bottom piano roll. In each, the left histogram bins the events by the average KW of events, and the right histogram is the duration of events in seconds. The final bin, contains all events larger than the other bins. Clicking on a bar highlights the events with black in the piano roll. Choosing another bar will cancel the other highlighting, and use the newly selected.

We had hoped that we could provide a more ad hoc query function, where values were supplied by events selected by the circuit. In the absence of time, this was changed to choosing events that are found in that bin.

3.2.4 Circuit levels view

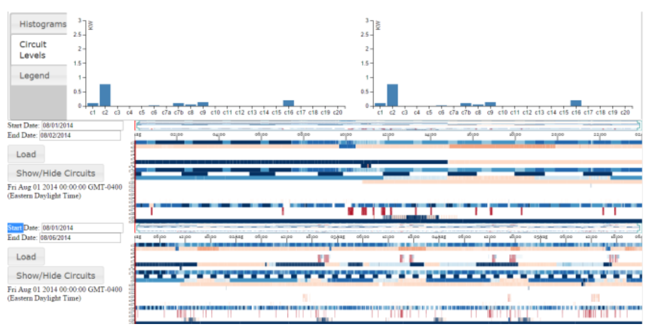


Fig.3 User interface - Main window with Circuit Levels.

To address task 6, we made the bar graph for the power load of all circuits.

The other tab at the top is a bar graph of all circuits at the location of the red vertical line in the piano bar (Fig. 3). It is for the power load in KW for each circuit. It has two sections. The left one is for the top piano roll, and the right one is for the bottom piano roll. This bar chart updates as the user changes the position of the red line in each piano bar.

4. Pilot Study

The pilot study was designed to validate the tasks that were most important to the customer, which was the ability to explore the dataset and find patterns and trends. The questions, and the results are found below.

4.1 Questions

The pilot study was divided into four categories which made it easier for us to collect results for our visualization tool. The four categories and questions related to it are as follows:

4.1.1 Data Overview

- Are the majority of events in the data set high powered, or low powered?
- Which circuit seems to draw the most power overall?
- Which circuit seems to draw the least power?
- Which time of day appears to have the least power draw?
- Which time of day appears to have the most power draw?

4.1.2 Identifying Patterns on Same Day

- Identify all patterns that repeat throughout the day, noting the circuit and timestamps of each event involved.

4.1.3 Identifying Patterns on Separate Days

- Identify at least two similar patterns on piano roll 1 and piano roll 2, noting the circuit and timestamps that they occur.

4.1.4 Tying Events to Appliances

- For the events in each day, identify as many appliances you can based on what you see. Note the timestamp and circuit, as well as the device/appliance you believe is on the circuit.

4.2 Results

4.2.1 Data Overview

To study the effectiveness of our visualization tool, we asked four user to study our visualization tool and answer questions mentioned in section 4.1.1. All the four participants were able to differentiate between high powered and low powered events in the data set. All of them answered low powered events as majority of the data sets on the piano roll were dark blue referring to low power consumption.

Three out of the four were able to correctly answer that circuit 2 drew the most power overall. Two out of the four were able to correctly answer that circuit 3 drew the least power overall. By using the color legend, users were able to distinguish between circuits that drew the most power and least power. All participants had a hard time determining the time of day that the most power was consumed.

From this we conclude that color is good for inexact comparisons. Every participant could identify the majority of events were low powered, although, this may have been guided by the histogram. Comparing the colors of each circuit for total consumption was generally easy to do as well. However, it was much harder for participants to compare power usage all circuits at any particular time. This is backed up by the numerous wrong answers about the time of day that the most and least power is used.

4.2.2 Patterns on Same Day

To study if the user can identify patterns on the same day, we asked the question and participants responded as follows.

All participants identified C17 and C18 had patterns although timestamps of corresponding events they mentioned were different. Two participants answered C8 had a pattern. One participant pointed out C1 and C16, and another mentioned C2 and C10.

It is difficult to grade these on their accuracy, as we never defined what a pattern is. C17 and C18 are active at the same time, so those could be considered a pattern. C8 maintains a constant pattern across all days. Overall, all patterns found by participants are indeed patterns, although, there were patterns that were not observed.

4.2.3 Patterns on Separate Days

For the question about the patterns on separate days, participants responded as follows.

The first participant answered that both two rolls were the same on the same period. Especially, C8 was pointed out - 17:55 to 19:22 on Piano Roll 1 and 15:56 to 17:43 on Piano Roll 2. The second one identified C9, C2, and C7A had the same pattern when Piano Roll 1 was between 8/3 and 8/4, and Piano Roll 2 was between 8/4 ~ 8/5. The third one responded that C17 and C18 had the same pattern on Piano Roll 1 and Piano Roll 2. Both the third participant and the last one answered that C2 had the same power consumption on Piano Roll 1 and Piano Roll 2.

Just like the previous section, all participants did identify actual patterns, but not all patterns were observed by participants. The piano rolls were helpful in providing the ability to find events that occur only once a day, and don’t repeat in a single day.

4.2.4 Tying Events to Appliances

One of our task was that based on the events, identify as many appliances by looking at piano roll. Only one person was able to answer this correctly, and that was because they had prior knowledge as to what was on the circuits. Just by looking and finding patterns by comparing different circuits is not enough to identify appliances.

5. Future Work

From the feedbacks and criticism we received for our visualization tool, we have proposed some future design implementations. The main problem was the color scheme. The color scale from blue to red is not equal. Sometimes it was difficult to differentiate between circuits as there were no lines or boundaries that were dividing each circuits due to color overlapping. We can add different color scheme that differentiate each circuits so that it is easier for user to understand and interpret information.

Another problem was axis labels. The histograms and circuit level diagrams need legends as it was harder for users to interpret from what they were looking at. we can add axis labels that could help user to understand and compare information much better.

Another feedback on piano rolls were that it was very useful find the overall power consumption of different circuits easily. Distribution of event length is useful to find some repeated events. For future references, correlation of power consumption in different circuits can be very useful. Better event detection algorithm can be applied to current design.

The other problem was the need of prior knowledge. It was hard for participants to interpret the visualization without the knowledge in the field. We can make a manual and a simple tutorial with several examples. A user can learn background information and how to operate the tool from the manual. In addition, the user can practice performing a task and analyzing the visualization by following short examples in the tutorial.

Based on this feedback, the next steps for the visualization includes using known patterns of pattern usage to show when certain appliances are running on a circuit. This would show running appliances on a circuit instead of the power consumption. The color scheme also needs more attention. The distribution of colors for low power events and high power events is good, with more colors on the lower end than the higher end, but the colors of blue and red used at the transition are not intuitive, as a darker color is assumed to have more importance or a higher value than a lighter color.

Additionally, some tasks that we would like to complete still include searching by example and annotating and tagging patterns that are identified. We would also like to work on other ways to aggregate the power information, to provide a more informative view.

6. Conclusions

The visualization tool provides good overview for each circuit. Color does not provide enough information to answer questions about all circuits. Color can be used to differentiate between low powered and high powered events in the data sets. It also can be used to identify repeating patterns. Diverging color scheme works well, but should not use white in the middle of the scale. It is nearly impossible for a user to identify devices with accuracy that are running on each circuit without prior knowledge. Overall, the visualization is useful for our customer, providing a view at the circuit level without occluding information.

References

[1] Jarke J. Van Wijk , Edward R. Van Selow, “Cluster and Calendar Based Visualization of Time Series Data,” *Proceedings of the 1999 IEEE Symposium on Information Visualization*, p.4, October 24-29, 1999.

[2] Sean Arietta, Alexei Efros, Ravi Ramamoorthi, Maneesh Agrawala, “City Forensics: Using Visual Elements to Predict Non-Visual City Attributes,” http://vis.berkeley.edu/papers/cityforensics/paper.pdf.

[3] Xiaofan Jiang , Minh Van Ly , Jay Taneja , Prabal Dutta , David Culler, Experiences with a high-fidelity wireless building energy auditing network, *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, November 04-06, 2009, Berkeley, California [doi>10.1145/1644038.1644050]

[4] Weaver, C., Fyfe, D., Robinson, A., Holdsworth, D., Peuquet, D. and MacEachren, A. M. (2006) Visual Analysis of Historic Hotel Visitation Patterns, *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, Baltimore, MD.

[5] Yuri A. Ivanov, Christopher R. Wren, Alexander Sorokin, and Ishwinder Kaur. Visualizing the history of living spaces. *Transactions on Visualization and Computer Graphics*, 13(6):1153–1160, 2007.