# Summarisation & Visualisation of Large Volumes of Time-Series Sensor Data

**Zhengwei Qiu, Cathal Gurrin, Aiden R. Doherty, Alan F. Smeaton**

CLARITY: Centre for Sensor Web Technologies, Dublin City University, Ireland

{zqui,cgurrin, adoherty, asmeaton}@computing.dcu.ie

## ABSTRACT

**With the increasing ubiquity of sensor data, presenting this data in a meaningful way to users is a challenge that must be addressed before we can easily deploy real-world sensor network interfaces in the home or workplace. In this paper, we will present one solution to the visualisation of large quantities of sensor data that is easy to understand and yet provides meaningful and intuitive information to a user, even when examining many weeks or months of historical data. We will illustrate this visulalisation technique with two real-world deployments of sensing the person and sensing the home.**

**Keywords:** Heatmap, Time-Series data, Visualisation

## 1. INTRODUCTION

Sensors are ubiquitous and we are sure to rely on them increasingly in our daily lives as the costs of deployment becomes less. There are many types of sensors, from environmental sensors that monitor our cities and rivers [1], location specific sensors that monitor our immediate environment (e.g. heat of our buildings or our electricity consumption) to person sensors that monitor us as individuals (e.g. location sensing [2] or health monitoring).

We have worked with a variety of sensors in this research with a view to sensing the person and sensing the environment in which they live and work. For example we utilise Microsoft SenseCams to identify the activity of an individual by monitoring the onboard accelerometer [3]. Associated with this is a location sensor (via GPS logger) which stores the location of a person every five seconds. This presentes us with our first challenge, how to represent location data in an easy-to-read and understand summary format. In additon, we can measure home $CO_2$ emissions through a number of sensors, including an electricity usage sensor supplied by Episensor. This poses our second challenge, how to summarise an extended period of electrictiy usage data for a home user.

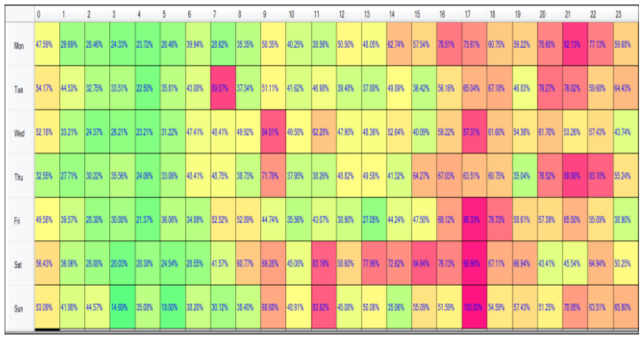

**Figure 1 – Heat Map visualisation of 3 months of overall electricity consumption data in one home.**

## 2. Background

Taking electricity consumption as an example, traditionally our usage in the home is only checked manually a few times per year and our bills come with the granularity of 1 or 2 months' worth of data. Obviously, real-time smart meter electricity logging will change this for the better [4]. Real-time smart electricity meters can infer users more about the usage of their electricity. But it is a problem to give users a clear real-time summarised report with such a vast amount of what is essentially time-series data. Such data needs to be instantly visible on a myriad of devices in the home, for example laptops, handheld devices, smart phones and even app-enabled televisions.

There are many sophisticated algorithms that can be used to manage the vast amounts of data generated by such sensors. However for the average user, it is essential that the vast quantities of logged data are summarised in an intuitive manner. In this paper we concentrate on using heatmaps as a summarisation and visualisation technology.

## 2.1. Temporal Heatmaps Visualisation

Heatmaps are a 2 dimensional display technique in which values are represented using colours of varying intensity. In this way it becomes very easy to display linear data by colouring each cell in the two dimensional display in relation to the underlying data. In our implementation, we employ a 2-dimensional heatmap to display temporally summarised linear data. In this work, quantities of raw data can naturally be constrained to certain time periods that the user is interested in. In our case, we utilise 168 data values (24 hourly values x 7 days) to represent each hour in a typical week (see Figure 1 for an example). However we may use a different scale, for example weeks and months or months and years.

In our implementation, those 168 hourly values are then normalised (against the highest hour or alternatively against a known upper bound) which means that we can associate these hourly values with colour intensity values. These intensity values on their own are relatively meaningless to the average home user, but when they are visualised on a heatmap (see Figure 1) it is quickly apparent that the user can grasp the meaning of this data.

The process through which a heatmap is generated for a large archive of time-series data, visualised over a 168 cell heatmap is as follows:

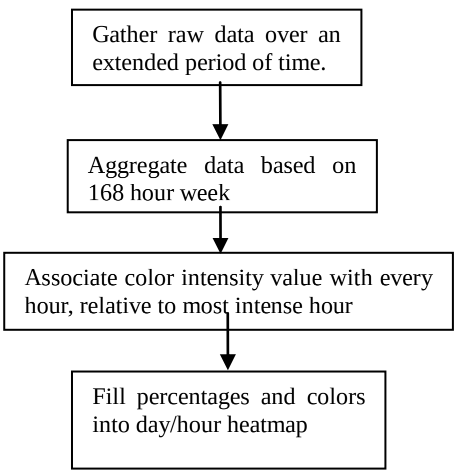


**Figure 2 – The processing stages required to generate a single heatmap which summarises vast quantities of any type of time-series sensor data.**

## 3. Use Case: Electricity Usage

Although this paper focuses on the visualisation of time-series sensor data, we must also be concerned with the gathering of the data to supply the source for visualisation. We will now briefly describe the hardware element, before focusing on the visualisation.

To gather electricity usage data from a home we employed an off-the-shelf real-time energy usage tool [5]. As shown in the Figure 1, the users can get infomation about when they use most electricity at a time granularity of one minute. From this more fine grained information, users can be better informed of their electricity usage, thus offering the potential to save money on their electricity bill.

As shown in Figure 3 and 4, the data collection system consists of an energy monitor attached to a household's main incoming power supply. This monitor records various aspects of the power consumption data. Figure 3 illustrates how in-home energy data is collected.

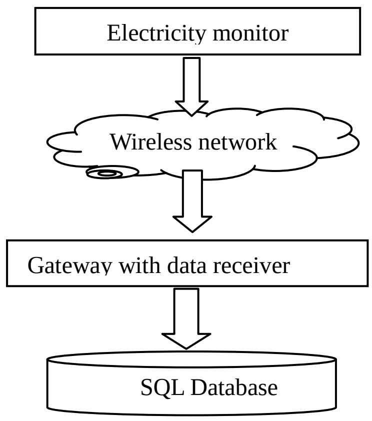


**Figure 3 – A high-level overview of our in-home data gathering setup – the software setup.**

This electricity data is simply linear sensor data with an associated timestamp. In order to be easily visualised, this time-series data needs to undergo some automated processing to prepare the raw data for visualisation.

Figures 1 displays our heat map visualisation applied on three months of home energy logged data from a single households. We can make various observations about this data, such as the most intense electricity consumption time of the week (Sunday at 17:00 in household #1) which is the maximum value that is used for normalisation of the other values. There are various other findings that we can mine from this data, simply by observing the heatmap.

We would like to stress that these visualisations can be made interactive by allowing users to select their start/end times, and investigate their electricity consumption over a given time period e.g. summer usage may show different characteristics to winter usage where children in a household are back into a normal school-going routine.

There are many applications of this data beyond simply analysing the entire power consumption of a location. Consider Figures 5 & 7 where the electric shower usage of both households is displayed. We can see clearly see the difference in the shower habits of both households.

Although the work disclosed here is related to usage of electricity data, we are not locked into this data, as other sensor's data can be easily summarised via our visualisation technique, such as the location of an individual.

**Figure 4 – A high-level overview of our in-home data gathering setup – the hardware setup.**

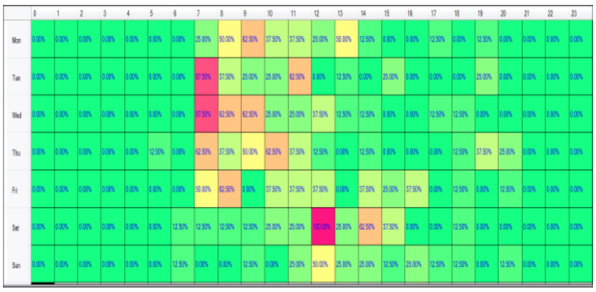

**Figure 5 – Heat Map visualisation of 3 months of electric shower usage - household #2.**

## 5. Use Case: Location Plotting

The second case study concerns an individual's location log. In this case, the subject carried a GPS logger for a number of years which logged location every five seconds. From this we can identify the location of the individual at any point in time, and also average the data over a prolonged periods of time (e.g. one year) and this allows us to answer lifestyle questions such as - what is the probability that the individual is at home on a Thursday at 3pm? Since this data can easily be averaged over time, we can employ a similarly scaled heatmap. Figures 8 and 9 show our heat map visualisation applied on 39 months of GPS data from one individual. We can once again easily mine knowledge such as the user staying at home longer during weekends, than during weekdays. Another item of note is that the user goes to work around 8-10am every day.

The different process between electricity usage data and GPS data is that the GPS data needed to be decomposed into textual locations using a gazetteer before being aggregated and plotted, thus turning it into <timestamp, single_value> time-series data. The process is as shown in the following diagram.

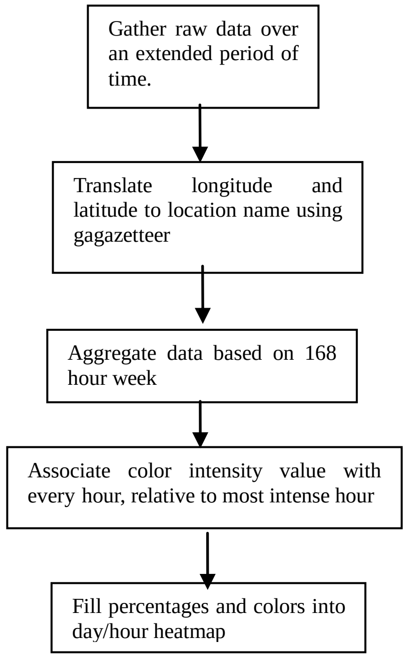


**Figure 6 – Process of converting GPS data to heat map**

## 6. Conclusions & Future Work

In this paper we have introduced a novel summarisation and visualisation technique to automatically supply users a simple and intuitive summary of time-series sensor data. Users can find their life pattern from the heatmap, and also identify areas for improvement in their lifestyle. We have shown that heatmaps are a very effective way to display large scale sensor archives of time series data. In future we'll carry out a user interaction study to evaluate this and improve it.

## ACKNOWLEDGMENTS

This material is based upon works supported by the Science Foundation Ireland under Grant No. 07/CE/I1147.

## REFERENCES

[1] Cleary, John and Maher, Damien and Slater, Conor and Diamond, Dermot (2010) In situ monitoring of environmental water quality using an autonomous microfluidic sensor. In: SAS 2010 - IEEE Sensors Applications Symposium, 23-25 February 2010 , Limerick, Ireland. ISBN 978-1-4244-4988-0

[2] Doherty, Aiden R. and Gurrin, Cathal and Jones, Gareth J.F. and Smeaton, Alan F. (2006) Retrieval of similar travel routes using GPS tracklog place names. In: Workshop on Geographic Information Retrieval 2006, 6-11 August 2006, Seattle, Washington.

[3] S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur, and K. Wood. Sensecam: A retrospective memory aid. In UbiComp: 8th International Conference on Ubiquitous Computing, volume 4602 of LNCS, pages 177–193, Berlin, Heidelberg, 2006. Springer.

[4] S. Darby. The effectiveness of feedback on energy consumption. Technical report, Environmental Change Institute, University of Oxford, 2006.

[5] Aiden R. Doherty, Zhengwei Qiu, Colum Foley, Hyowon Lee, Cathal Gurrin, and Alan F. Smeaton (2010) "Green Multimedia: Informing People of their Carbon Footprint through Two Simple Sensors", To appear in ACM Multimedia 2010, 25-29 October, Firenze, Italy.


| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mon | 0.00% | 25.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 50.00% | 0.00% | 0.00% | 0.00% | 25.00% | 50.00% | 50.00% | 0.00% | 0.00% | 0.00% |
| Tue | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 25.00% | 25.00% | 0.00% | 25.00% | 25.00% | 0.00% | 0.00% | 25.00% | 0.00% | 0.00% | 0.00% | 25.00% | 50.00% | 0.00% | 25.00% | 25.00% | 0.00% |
| Wed | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 25.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 25.00% | 0.00% | 50.00% | 0.00% | 25.00% | 25.00% | 25.00% | 0.00% | 0.00% | 25.00% | 0.00% |
| Thu | 25.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 25.00% | 0.00% | 50.00% | 25.00% | 0.00% | 0.00% | 0.00% | 0.00% | 25.00% | 0.00% | 0.00% | 50.00% | 0.00% | 75.00% | 75.00% | 25.00% |
| Fri | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 25.00% | 25.00% | 0.00% | 0.00% | 0.00% | 0.00% | 25.00% | 100.00% | 25.00% | 0.00% | 25.00% | 0.00% | 0.00% |
| Sat | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 25.00% | 0.00% | 50.00% | 25.00% | 0.00% | 0.00% | 25.00% | 0.00% | 25.00% | 75.00% | 25.00% | 25.00% | 0.00% | 0.00% |
| Sun | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 25.00% | 0.00% | 50.00% | 25.00% | 0.00% | 50.00% | 0.00% | 25.00% | 25.00% | 0.00% | 25.00% | 0.00% | 0.00% | 0.00% |

**Figure 7 – Heat Map visualisation of 3 months of electric shower usage – household #1.**

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tue | 41.06% | 37.24% | 39.67% | 40.53% | 43.44% | 46.75% | 49.24% | 50.54% | 47.39% | 44.35% | 40.41% | 36.85% | 34.07% | 27.27% | 25.91% | 26.71% | 26.32% | 30.10% | 31.69% | 37.84% | 44.66% | 48.72% | 48.30% | 49.56% |
| Wed | 49.56% | 50.79% | 51.50% | 51.50% | 52.10% | 56.15% | 55.97% | 48.83% | 37.56% | 22.41% | 11.44% | 8.80% | 7.19% | 6.97% | 6.80% | 5.25% | 5.86% | 7.23% | 15.14% | 25.25% | 36.91% | 48.54% | 52.61% | 53.77% |
| Thu | 52.86% | 52.93% | 52.98% | 52.98% | 54.76% | 56.67% | 52.64% | 44.23% | 28.27% | 14.14% | 3.90% | 2.34% | 1.19% | 1.19% | 1.54% | 2.76% | 2.19% | 2.63% | 6.77% | 13.37% | 25.98% | 37.18% | 45.90% | 49.97% |
| Fri | 51.70% | 51.19% | 50.75% | 51.65% | 53.57% | 54.72% | 51.12% | 41.29% | 26.64% | 12.83% | 3.53% | 2.40% | 2.38% | 2.34% | 1.79% | 1.59% | 1.94% | 2.38% | 4.01% | 11.11% | 27.30% | 38.19% | 47.57% | 48.50% |
| Sat | 49.93% | 49.90% | 51.24% | 51.69% | 54.17% | 57.67% | 56.24% | 47.38% | 34.23% | 17.24% | 6.87% | 3.82% | 2.98% | 3.26% | 2.98% | 3.32% | 1.47% | 1.95% | 2.73% | 4.84% | 18.96% | 30.98% | 40.36% | 45.53% |
| Sun | 47.19% | 47.73% | 48.81% | 48.81% | 51.40% | 53.89% | 54.91% | 45.62% | 34.32% | 18.33% | 5.48% | 2.36% | 1.80% | 1.64% | 1.22% | 1.80% | 0.82% | 1.37% | 4.65% | 9.40% | 17.24% | 23.28% | 30.64% | 34.17% |
| Mon | 32.74% | 37.60% | 41.77% | 41.92% | 42.51% | 43.81% | 52.10% | 51.78% | 49.68% | 43.90% | 38.01% | 31.78% | 28.77% | 25.31% | 23.20% | 23.20% | 27.08% | 29.63% | 34.83% | 37.00% | 41.97% | 39.33% | 37.38% | 41.45% |

**Figure 8 – Heat Map visualisation of 39 months of time the individual could be found at home.**

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tue | 0.14% | 0.03% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.09% | 0.92% | 2.05% | 2.99% | 2.83% | 4.35% | 6.74% | 4.61% | 1.24% | 0.00% | 0.00% | 0.00% | 1.77% | 2.17% | 4.57% |
| Wed | 5.22% | 5.99% | 5.99% | 5.99% | 4.79% | 1.80% | 0.36% | 6.97% | 16.54% | 32.38% | 47.39% | 52.89% | 51.00% | 46.72% | 38.32% | 36.28% | 34.38% | 31.93% | 22.69% | 10.88% | 5.42% | 2.68% | 2.90% | 3.13% |
| Thu | 4.43% | 5.36% | 5.36% | 5.36% | 4.76% | 2.38% | 3.41% | 7.59% | 22.14% | 42.44% | 56.45% | 61.18% | 58.33% | 57.40% | 47.90% | 41.75% | 40.90% | 36.95% | 23.74% | 14.41% | 6.29% | 2.83% | 1.28% | 2.99% |
| Fri | 3.61% | 4.76% | 4.81% | 4.76% | 2.38% | 1.46% | 2.87% | 8.38% | 23.71% | 40.23% | 52.61% | 56.55% | 57.29% | 55.30% | 47.26% | 39.62% | 37.03% | 31.95% | 22.02% | 11.09% | 5.56% | 2.21% | 0.98% | 1.82% |
| Sat | 5.49% | 7.74% | 7.14% | 7.14% | 5.36% | 2.98% | 0.80% | 4.87% | 16.92% | 31.25% | 47.07% | 52.67% | 52.38% | 54.19% | 50.00% | 41.72% | 36.95% | 33.28% | 23.51% | 13.88% | 7.77% | 2.90% | 1.55% | 3.69% |
| Sun | 4.27% | 6.44% | 5.95% | 5.36% | 2.98% | 2.98% | 2.93% | 5.98% | 15.72% | 31.21% | 46.27% | 50.95% | 48.50% | 48.02% | 38.42% | 31.97% | 25.22% | 22.73% | 17.15% | 9.95% | 4.30% | 2.18% | 0.62% | 0.14% |
| Mon | 1.36% | 1.83% | 1.80% | 1.80% | 1.80% | 1.20% | 0.60% | 0.00% | 0.00% | 1.22% | 2.14% | 2.75% | 2.48% | 2.62% | 3.09% | 5.44% | 4.15% | 1.87% | 0.61% | 0.60% | 0.60% | 0.83% | 0.36% | 0.60% |

**Figure 9 – Heat Map visualisation of 39 months of time the individual stays at the place of work**