The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

How to enable the utility data under the lower sampling rate and less attribute for making smart cities

Eiji Murakami[*1]

[*1] Azbil Kimmon Co., Ltd.

This paper discusses the utility data analysis under the different sampling rate makes variety of result. Otherwise, we use the high sampling data and much attributes, we start to think combination of the data which haven't measured at same time of utility data has measured. To come up the idea of data combination, the method of idea creation should employ for making the innovation. This paper will show the difficulties and some of the existing method to overcome them.

1. Introduction

The research for the utility data analysis has been conducted on electricity consumption of electric power meters, where is mainly in Europe. In 2019, Internet of things (IoT) of gas and water meter will be started using 4G LTE of mobile communication technology in Japan.

As a result of employing the IoT technology, the measuring consumption of utility data which has been conventionally done every once a month or every two months, now it could carry out at least every day and even more frequent to measure utility data every hour if necessary. In this paper, we show an example of utility data and discuss how difficult to analyze if the sampling rate is lower and the sampling data is reduced. We also discuss how difficult will be made in analyzing utility data with less attribute data. Finally, I will describe what is expected in the future when utility data is collected under the lower sampling rate and less attribute data.

2. Utility data analysis

2.1 Data sampling rate

The data sampling rate is important in utility data analysis. Table 1 shows the relationship between sampling rate and acquired data. When the sampling rate is 1 hr - 15 min then acquisition data is visualized, it becomes a only bar chart showing the result that is consumption of utility in each time span.

Data Sampling Frequency	1 hr – 15 min	1 min – 1 sec	10Hz – 2KHz	10KHz – 15KHz	> 1MHz
Data Appearance	1 hr PG&E	<1 min Sampling	Similar to adjacent graphs	KHz Sampling	MHz Sampling

Table1 Sampling rate and graph pattern, cited from [Armel11]

When the sampling rate is 1 min - 1 sec then acquired data is visualized, it becomes a line graph with sampling data. It can be estimate the consumption of utility, and could be inferred the purpose from gas or water service had been consumed in the consumption pattern.

2.2 Utility data of 1 min sampling rate

Fig 1 is a graph of the power consumption in the electric power meter with one minute period

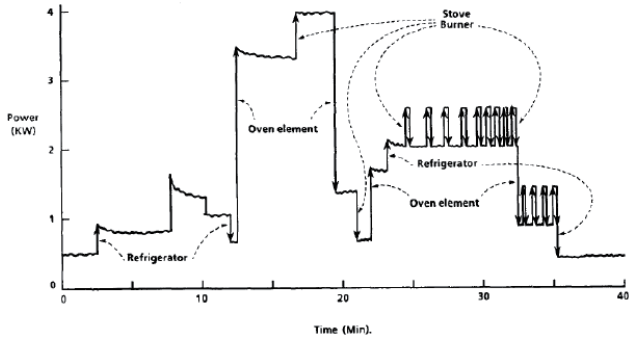

Fig 1 One min sampling rate and graph pattern, cited from [Armel11]

According to the Fig1, it can be categorized by the pattern of energy consumption for all appliances. This is an example where it is feasible if the data sampling rate is 1 minute or less with having appropriate technology.

3. Study for data analysis of REAL restaurant

Fig 2 shows an example of measuring utility data, temperature and humidity at a restaurant.

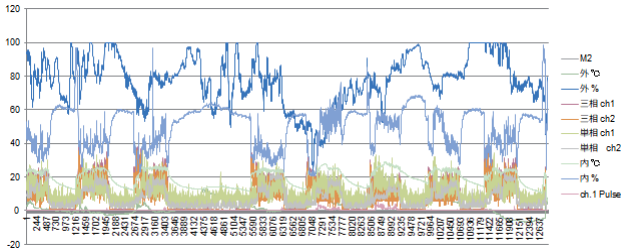

Fig 2 REAL restaurant various data

The meaning of Legend (symbol) of the graph is as follows. M2 is gas consumption, 外 is outdoor, ch1 and ch2 are electrical

Contact: Eiji Murakami, Azbil kimmon Co.,Ltd., 1-14-3 Kita-Otsuka, Toshima-ku Tokyo, 170-0004

power consumption reading of power meter, 内 is inside (room). The horizontal axis shows elapsed time from the start of measurement in seconds. In the vertical axis, the measured value was normalized between 0% and 100%. Ch1.pulse is the human sensor value installed at the entrance of the restaurant. All the data has sampled by 1 minute. So we can measure the gas consumption, outdoor temperature and humidity, room temperature and humidity, power consumption and human sensor value by each one minute.

The one analysis method is observing the graph pattern. The other analysis is we employ statistical method as follows, correlation analysis, factor analysis, and principal component analysis. From principal component analysis, the following facts are discovered.

The combination of gas consumption, indoor temperature and single-phase electrical power consumption ch 1 have same behavior (in principal component analysis, if the values of principal component are near same, these variable belong to same category). The combination of the single-phase electrical power consumption ch 2 and the three-phase electrical power consumption ch 1 and ch 2 have the same behavior according to result of principal component. Outdoor temperature and human sensor value are correlated.

On the other hand, outdoor humidity and room humidity have no correlation.

One hypothesis has been obtained, paying attention to the correlation between the outdoor temperature and the value of the human sensor, there is a tendency for customer wants to come restaurant more often when outdoor temperature gets low in winter season. As a hypothesis verification, this phenomenon can be explained from the fact that this restaurant is a ramen shop located in Sapporo.

4. Data with low sampling rate and less attribute

Section 3 shows the example where there are sufficient attribute data at high sampling rate. In section 3, let us consider a case where the sampling rate has changed from one minute to one hour. When the sampling rate is 1 minute, the graph pattern is a line graph in plotting measured values of each sample and connecting each sample with a line. When the sampling rate is 1 hour, the measured value of each sample is represented by a bar graph. The fluctuation of the sample value generated within one hour is invisible on the graph. Any changes in sample values generated within one hour are unknown. This is a specific harmful effect when the sampling rate changes to a lower rate.

Next, let us consider the case where available attribute data decreases in section 3. In section 3, we applied statistical methods to all data and discovered that there is a correlation between outdoor temperature and human sensor of restaurant door by counting how many people comes in with principal component analysis. The obtained findings are natural for ramen shop in Sapporo, but it is not easy to choose the limited sampling data for measuring with limited sensor from the beginning, rather than try and error exploratory way. Also it is not easy for hypothesis verification by one shot analysis with limited data.

5. Discussion and future work

[Toki 18] solved the vehicle routing problem of LP gas cylinder which is a special case of travelling salesman problem (TSP). The necessary data for the vehicle routing problem is the delivery address and delivery date. By the gas meter IoT, gas consumption data can be obtained every day, so we can determine the delivery date to minimize the delivery cost. Once the delivery date is determined, sort out the all address of the LP gas consumer then full charged LP gas cylinder should be delivered at the same day which is today. Then search the optimal the delivery route for all customer location.

The necessary data for vehicle routing to deliver the LP gas cylinder is only gas consumption and customer's address. Now we think about other applications that could enrich our living life with restricted information.

Fig. 3 shows an example of adding the certain color on map after the gas consumption has been calculated for each customer. There could be various meanings of the color, for example it can changes to red when the gas consumption is high. We know the whole volume of the LP gas cylinder, the remaining gas volume of the LP gas cylinder can be calculated from the gas consumption volume. If the remaining gas is very low volume, we can change the color like red on map which might have special meaning such as energy supply for customer has been critical condition at the potential risk of LP gas company. The meaning of area which is colored on map may be varied depending on the application. However, the data to use is the same as [Toki 18], and only difference is adding the geographic information. The applications that use geographic information in this way called Geographic Information System (GIS).

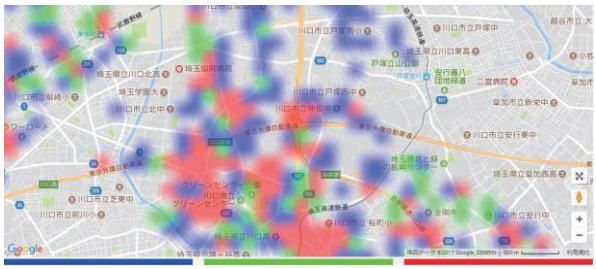

Fig 3 Example of several colors on map (The map copyright belongs to Google and ZENRIN)

The future work would be using utility data obtained by low sampling rate like [Toki 18] uses daily gas consumption and use few number of attribute data, otherwise using the geographical information have society enrich by application or service [Kumar14] under the restricted information should be studied more. Also the methodology for create the innovation like IMDJ [Ohsawa 17] which can share the same goal as I explain just before would like to apply.

References

[Armel11] Carrie Armel, Energy Disaggregation, Precourt Energy Efficiency Center, Stanford, 2011

[Kumar14] TM Vinod Kumar, Geographic Information System for Smart Cities, Copal Publishing Group, 2014

[Ohsawa17] Yukio Ohsawa,Teruaki Hayashi,Masahiro Akimoto,Noriyuki Kushiro,Jun Nakamura,Masahiko Teramoto, MoDAT Market of Data, Kindaikagaku publisher, 2017

[Toki18] Soma Toki,Naoshi Shiono,Eiji Murakami,A practical approach to the vehicle routing problem in cylinder gas distribution, INFORMS Annual meeting, 2018