None Intrusive load monitoring

RUNE A. HEICK 11061

Aarhus University, Department of Engineering

1-12-2015

Abstract

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Contents

Introduction 1

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Data Quality 2

Various projects today is focused on gathering data and analysing it. The gathered data is used for obtaining behaviours, habits and properties of the observed objects. This is done by using powerful statistical leaning algorithms, that are able to deduce these properties from the data. This approach is called data driven development, since the success is mainly determined by the data and not the algorithm.

When data is the central role of the system, the quality of the data are very important. Poor data can let to wrong assumptions, and have a negative effect on the application. Choosing the correct dataset is therefore a key factor [1]. Looking at the quality of the data can help you chose what dataset to use. Data quality can be described many ways, one of the more formal is from the ISO 8402 standard that describes quality as:

> *"The totality of characteristics of an entity that bear upon its ability of satisfy stated and implied needs"* [2].

This indicate that data quality is something that is very depended on the intended use and is therefore hard to generalize.

Quality of data is a subject that is gaining more and more attention due to the fact that we are moving data gathering from controlled labs, to the public. Many of these projects is the citizen science project, where it is the citizen who collect the data, and not interested parties [3]. This enables researchers to gather enormous amount of data, but they are no longer in control of the conditions the data is collected in, which introduces errors and other quality decreasing factors.

2.1 Quality Parameters

It is not uncommon that deficient areas of of research has its own quality criteria. This is due to the fact that quality is a very domain specific subject.

One of the areas that have been dealing with citizen data for many years is the Geographic information area, that are used for maps, weather prediction and climate research. They have come up whit several ways of describing quality in spatial data [4]. Method for defining quality in time series data have also been developed [5].

One of the things all the methods have in common is trying to look at the completeness of the data. Some of the most low level criteria is the sample availability. Here we asses if there is gaps in the data, that is caused of unknown interference, and look at how the samples is distributed in the measurement period.

2.2   Appliance Citizen Data

As a part of the SmartHG project 25 households have been equipped with meters on selected appliances and the main meter. The data collected from this experiment are are prone with errors due to malfunctioning test equipment or unexpected interference from the resident there have been unintentionally turning off the measurement equipment for a period of time.

The sample availability quality of the data in the SmartHG dataset have been measured on a hour basis. The quality for each meter is assessed by looking at the available samples over the maximum expected samples in the period.

$$N_{max}^{(m,T)} = \lfloor (\phi_{start}^{(m,T)} - T_P) \times T_s^{(m)} \rfloor \tag{2.1}$$

$$q^{(m,T)} = \frac{N_{observed}^{(m,T)}}{N_{max}^{(m,T)}} \tag{2.2}$$

As shown on equation 2.1 is the maximum number of samples for a meter m in the period T calculated by taking the period time T_P, corrected with the sample phase (ϕ_{start} for the given period, and dividing it with the sample time T_s. The quality of the meter is calculated as the ratio of observed samples in the timeslot T to the maximum samples, shown in equation 2.2.

To find the quality of a house in a given period T, that have a set of meters M, we take the mean value of all the meter quality's, as shown in equitation 2.3.

$$\mu_{q(M,T)} = \frac{1}{\mathbf{card}(M)} \sum_{m \in M} q^{(m,T)} \tag{2.3}$$

Each house has a quality vector Q, with the house quality found whit a period T_P on one hour. This have been done from April t_{start} to October t_{end}.

$$Q^{(M,t_{start},t_{end})} = \{\mu_{q(M,T)} | T \in \{t_{start}, t_{start} + T_p, t_{start} + 2 \times T_p, ..., t_{end}\}\} \tag{2.4}$$

This is shown in equation 2.4 where M is a set of the meters in a given house. This can be graphically shown on the figure show the figure of quality where the color is a gradient running from light green for the best quality to red for bad quality.

2.3   Related Work

Reconstruction 3

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Recondition 4

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Bibliography

[1] N. Regnauld, "Generalisation and data quality," 2015-08-01T00:00:00Z. Type: article; CC BY.

[2] I. 8402:1994, *Quality management and quality assurance – Vocabulary.* ISO, 2 ed., 1994-03-24.

[3] S. Kelling, D. Fink, F. A. L. Sorte, A. Johnston, N. E. Bruns, and W. M. Hochachka, "Taking a ?big data? approach to data quality in a citizen science project," 2015-10-27; 2015-11. Type: Text; © The Author(s) 2015; Open AccessThis article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

[4] I. 19157:2013, *Geographic information – Data quality.* ISO, 1 ed., 2013-12-15.

[5] G. Pastorello, D. Agarwal, T. Samak, C. Poindexter, B. Faybishenko, D. Gunter, R. Hollowgrass, D. Papale, C. Trotta, A. Ribeca, and E. Canfora, "Observational data patterns for time series data quality assessment," in *e-Science (e-Science), 2014 IEEE 10th International Conference on*, vol. 1, pp. 271–278, 2014. ID: 1.