Optimized Statistical Test for Event Detection in Non-Intrusive Load Monitoring

Leen De Baets, Joeri Ruyssinck, Chris Develder, Tom Dhaene and Dirk Deschrijver
Department of Information Technology, Ghent University - imec
Technologiepark-Zwijnaarde 15, 9052 Ghent
Email: leen.debaets@intec.ugent.be

***Abstract*—Event detection plays an important role in non-intrusive load monitoring to accurately detect when appliances are switched on or off in a residential environment. Besides being accurate, it is important that these methods are robust on real-life power traces. This paper shows that some state-of-the-art event detection methods may miss events when there is a substantial base load caused by active power consuming devices. In order to address this problem, this paper extends the existing chi-squared goodness-of-fit test with a a voting scheme. Furthermore, a work flow is proposed using surrogate-based optimisation for tuning the parameters of these tests in an efficient way. Results on the BLUED dataset indicate that the novel voting chi-squared GOF method outperforms the standard chi-squared GOF test when applied to traces with a higher base load.**

I. Introduction

Non-intrusive load monitoring (NILM) identifies the per-appliance energy consumption by first measuring the aggregated energy trace at a single, centralized point in the home using a sensor and then disaggregating this power consumption for individual devices using machine learning techniques. This energy disaggregation method makes it possible to decompose the electricity bill into the consumption of individual appliances, which can lead to insight and empowerment for the consumers to reduce their energy consumption in an informed way. This per-appliance energy consumption is also useful for energy management tools, as it facilitates the detection of high energy consuming appliances. This empowers the users to devise energy conservation strategies, such as re-scheduling of high power demanding operations for the off-peak times.

This paper focuses on event detection methods that are capable of determining the step-changes in the power signals, which occur whenever a device is turned on or off. Once the events are detected, these changes can in turn be analysed for appliance recognition. The most commonly used statistical event detection method is the Generalized Likelihood Ratio Test (GLRT) [1], which tests if two neighboring windows representing consecutive time frames share a common distribution. The possible presence of an event in two neighboring windows is determined by calculating a decision statistic from the natural log of a ratio of probability density functions in those neighbouring windows. CUmulative SUM (CUSUM) filtering [2] is a method to determine changes in the quality number (e.g., the mean of the values in the windows or the difference between the predicted and real value) by testing it against a stopping criterion that describes when an event occurs. More recently, a χ^2 GOF test was introduced that detects events by assuming, like the GLRT, that two neighbouring windows share a common distribution. A χ^2 test statistic is applied on two neighbouring windows and an event is assumed if the null hypothesis is rejected [3]. A more straightforward heuristic is defined by the Real-time Energy Activity Classification Technique (REACT) [4], which detects events by comparing two consecutive windows of power signals p and q with equal length n and reports the occurrence of an event when $min(|p_i - q_i|)$ $(\forall i = 1, ..., n)$ exceeds a certain threshold.

This paper focuses on the χ^2 GOF method as event detection test as it is widely used, simple and has good performance [3], [5], [6], [7]. Section II explains the preprocessing of the power signal. Section III shows that the performance of this method degrades when high consuming devices are present. This paper extends the method with a voting scheme, such that its performance can be improved substantially. Moreover, section IV proposes a surrogate-based method to identify the optimal parameter configurations. The performance of the novel approach is benchmarked and discussed in Section V. Section VI concludes this paper.

II. Preprocessing the data

The input for the standard and voting χ^2 GOF method is a power signal. In [8], it is mentioned that noise and spikes in the power trace can lead to false detected events. As a solution to remove the noise, a median filter is applied on the power signal. Each sample in the power signal p_i is replaced by the median of its m neighbours:

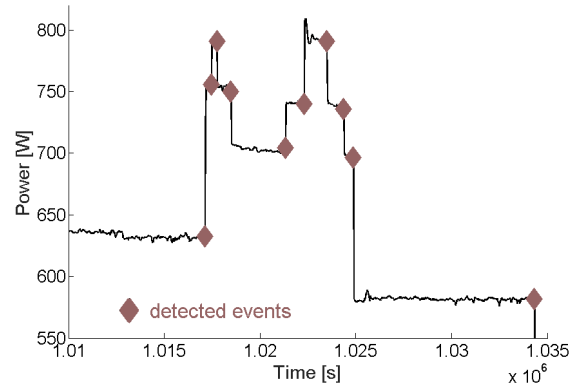
$$p_i = median(p_{i-m/2} + ... + p_{i+m/2}) \tag{1}$$

The parameter m can be trained, see Section IV.

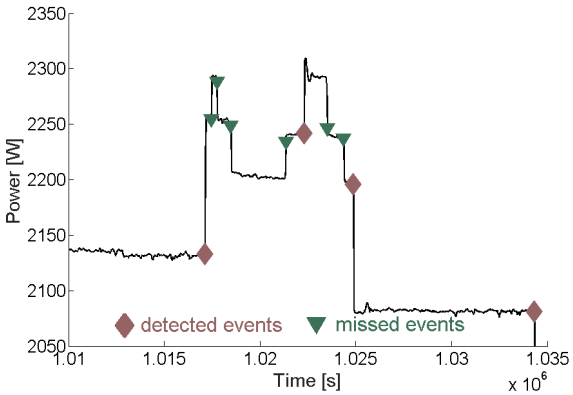
III. Voting χ^2 GOF Method

The standard χ^2 GOF method [3] detects events by relying on the fact that the distribution of power values before/after the occurrence of an event are different. To assess this difference, a probabilistic χ^2 test can be used. Assume two consecutive non-overlapping windows $q = (q_1, q_2, \cdots, q_n)$ and $p = (p_1, p_2, \cdots, p_n)$, each containing n data samples from the power signal. Then, an event occurs at the end of

978-1-5386-3917-7/17/$31.00 ©2017 IEEE



(a)



(b)

Fig. 1: Detected events (diamond) using the standard χ^2 GOF method with $n = 40$: (a) sample power trace from [17], (b) same trace with an added base load of $1500W$.

window q with a confidence level of $100(1-\alpha)\%$ and $n-1$ degrees of freedom, if

$$l_{GOF} = \sum_{i=1}^{n} \frac{(q_i - p_i)^2}{p_i} > \chi^2_{\alpha,n-1}. \quad (2)$$

The values of $\chi^2_{\alpha,n-1}$ can be looked up in a table [9].

Although this standard χ^2 GOF method has shown to be effective, its performance is sensitive to the base load of the signal. Figure 1 illustrates this problem: when the power base level is around 600W and an appliance using 50W is switched on, the event is correctly detected (Figure 1a). However, the event is missed if a base load of 1500W is added (Figure 1b). It is seen from equation (2) that events are characterized as a change in power $(q_i - p_i)$ relative to the power signal itself (p_i). Therefore, the method is prone to miss smaller events when the base load of the signal is high. This can lead to poor results. As a solution, a voting mechanism (based on the idea presented in [1]) is proposed in this paper to solve the problem.

In the voting χ^2 GOF method, the GOF is calculated for each sample in the power signal as given by equation (2). A voting window of length w slides over the resulting time-series of GOF-values and a sample gets a vote if its GOF-value is the highest among all points in the voting window. This results in a maximum of w votes. Each sample receiving at least v_{thr} votes is flagged as an event. The robustness of the voting method against changes in base load is shown in Section V.

Although some guidelines on how to choose the parameters of the statistical tests are given in [3], it is possible to further increase the detection ratios by optimizing them as shown in the next section.

TABLE I: The configuration parameters **g** and their range for the standard χ^2 GOF event detection.

standard χ^2 GOF		
name		range
window median filter	m	$1-100$
window event detection	n	$1-100$
confidence level	α	$[90, 95, 97.5, 99, 99.9]$

TABLE II: The configuration parameters **g** and their range for the voting χ^2 GOF event detection.

voting χ^2 GOF		
name		range
window median filter	m	$1-100$
window event detection	n	$1-100$
window voting system	w	$1-100$
voting threshold	v_{thr}	$w * [0.1 : 0.1 : 1]$

IV. Parameter Optimization

Both the standard and voting χ^2 GOF method have configuration parameters **g**, see Table I and II respectively. The parameter ranges that are investigated in this paper, lead to respectively 50.000 and $10.000.000$ possible configuration parameter combinations, which makes a brute-force optimization infeasible.

Surrogate-Based Optimization (SBO) [10] is proposed, which can significantly speed up the process of optimizing the configuration parameters without reducing the granularity of their ranges. The method assumes that smooth changes in the configuration parameters will lead to smooth changes in detection ratios. Under that condition, an exhaustive search of the overall model parameter space is not required to find the optimum solution. The training needs following steps, also depicted in Figure 2:

1) The performance is quantified using a utility function needing to be maximized. For the event detection methods, the F-measure is used as suggested in [11]. The utility function is thus defined as:

$$F(\mathbf{g}) = 2 \cdot \frac{precision(\mathbf{g}) \cdot recall(\mathbf{g})}{precision(\mathbf{g}) + recall(\mathbf{g})}, \quad (3)$$

The ultimate goal is to chose **g** in such a way that the F-measure is as high as possible.

2) A limited number of configurations $K(= 10)$ for the model parameters are chosen according a latin hypercube and the event detection method is applied to obtain

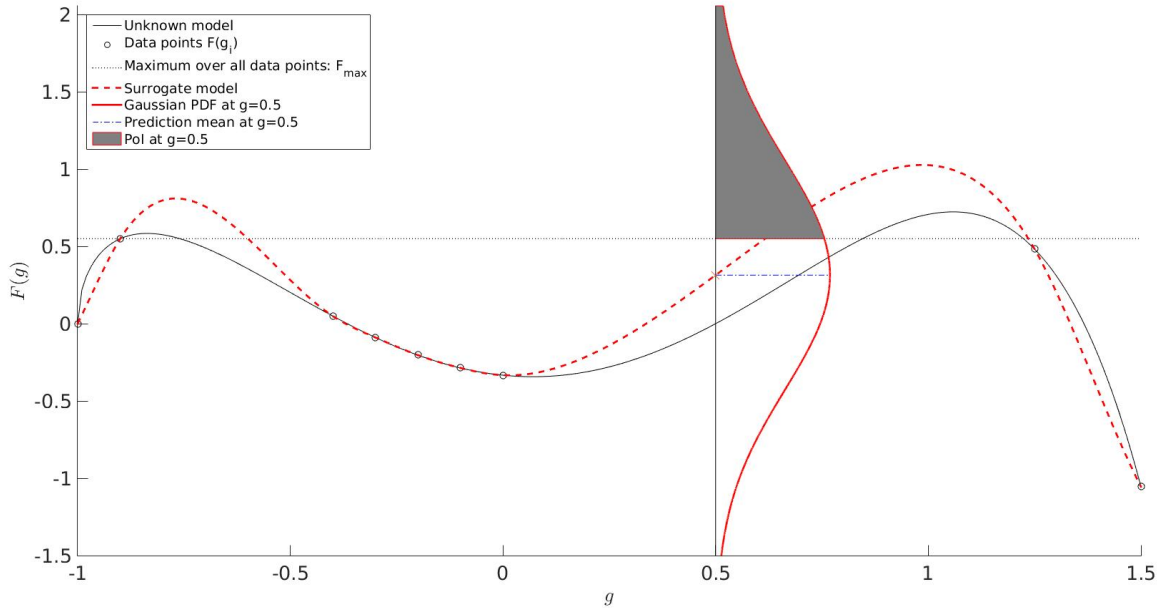


Fig. 3: A graphical illustration of expected improvement: a surrogate model (dashed line) is constructed based on some data points (circles) of an unknown function $F(\mathbf{g})$. For each point the surrogate model predicts a Gaussian probability density function (PDF). An example of such a PDF is drawn at $g = 0.5$.

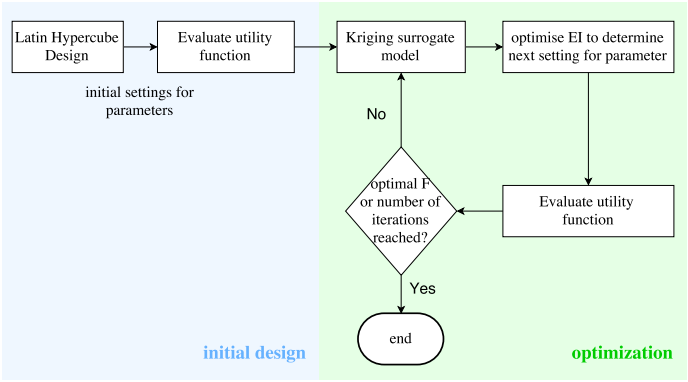


Fig. 2: A flowchart of surrogate-based optimization.

the F-measure. At the end of this step, the following set is obtained:

$$S = \{(\mathbf{g}_k, F(\mathbf{g}_k)), k = 1 \dots K\}. \quad (4)$$

3) A Kriging model of the F-measure is built as a function of the parameters listed in Table I or II. This model predicts the uncertainty of the F-measure as the realization of a normally distributed random variable $Y(\mathbf{g}) \sim N(\mu(\mathbf{g}), \sigma(\mathbf{g}))$ where $\mu(\mathbf{g})$ denotes the predicted value ($\mu(\mathbf{g}) \sim F(\mathbf{g})$) and $\sigma(\mathbf{g})$ denotes the prediction variance.. More details about Kriging can be found in literature, e.g., [12].
4) Once a Kriging model is built, the Expected Improvement (EI) measure determines the optimal configuration of the next parameter configuration $\mathbf{g}$ where higher F-measure values are most likely to occur. In other words, EI quantifies the Probability of Improvement (PoI), the amount of improvement that is expected to occur when a certain configuration is explored as compared to the optimal value found so far. This is done by considering every possible improvement over the current best value, multiplied with the associated likelihood. The corresponding F-measure of this optimal configuration point is calculated and is added to the set S.
5) The Kriging model is rebuilt and the process is repeated until a satisfactory solution is found, i.e., until the maximum is reached (in our case $F = 1$), or when the maximum number of iterations is exceeded. Here, the maximum iterations is set to 90.

Figure 3 illustrates one iteration of the SBO for the case where only one parameter g is optimized. The black line represent the unknown utility function $F(g)$, and the final goal is to construct a surrogate model of the line equivalent to the original one. The black circles are the data points for which the F-measure is calculated and they belong to set S. For each possible value of the parameter g, a Kriging model is built. The output of this model for the values in the set S is the Gaussian $N(F(g), 0))$. For points not in this set, the output $N(\mu(g), \sigma(g))$ is an estimation on the F-measure ($\mu(g)$) accompanied with a prediction variance ($\sigma(g)$). In the figure, an example of an output is given for $g = 0.5$. Also, the PoI for this value is depicted by the gray area enclosed by the Gaussian. This area equals the difference between the

maximum value in set S (the left most circle) and $\mu(0.5)$, multiplied with the likelihood of parameter 0.5. The parameter value with the highest gray area is chosen, the according F-measure is calculated and added to set S.

Thus, rather than computing the results for all possible model parameters, a surrogate model of the optimization objective function (the utility function) is generated that is locally accurate in the regions of interest (the optimum). SBO has already been applied successfully in other research areas, like e.g., wireless communication [13], electromagnetics [10], and microwave filter design [14]. An implementation of the SBO routine is available in the Surrogate Modelling (SUMO) Toolbox [15], [16] (available online[1]).

V. Results

The robustness against different baseload levels of the methods is tested on the BLUED dataset [17]. The aggregated active power signal of 60Hz from a family residence in the United States for a whole week is considered. Every state transition of each appliance is manually labeled, providing the ground truth. The considered house has a two-phase power consumption, where 904 transitions are recorded in phase A and 1578 in phase B. Since phase B is more noisy than phase A, they are both optimized and tested separately. For each method, the data is processed with a median filter.

Performance is evaluated on 20% of the data, whereas the remaining 80% is used for training. Performance is reported averaged over 10 runs (each with a random 20% test split). For training, 5-fold cross validation is used on the other 80% to set the optimal parameter values. Note that for the division, the trace of an entire day is taken as a whole unit.

To find out if each method is robust against changes in the base load of the power signal, the three methods are applied in three different cases. In all cases, a base load of respectively 0W, 1500W and 3000W is added to the power signal. In practice, such high base load conditions arise when multiple high-power devices are operating in the background, such as electrical heaters (that can easily consume 1500W). Considering the results of these use cases, a conclusion can be made about each method's robustness.

The results of the standard χ^2 GOF method when applied on the first three cases are given in Figure 4, showing the spread of the F-measure caused by running the 5-fold cross validation ten times. When no offset is added, the performance for phase A is almost perfect ($F \approx 0.98$) and for phase B the performance is $F \approx 0.80$. However, this changes when the base load is increased by adding an offset to the signal. The F-measure keeps dropping as the offset increases, indicating that the standard χ^2 GOF method is not robust against higher base loads.

The results of the voting χ^2 GOF method when applied on the first three cases are also shown in Figure 4. As can be seen, the voting χ^2 GOF method gives comparable results to the standard χ^2 GOF method for phase A and B when no offset is added. When offsets are added to the signal, the F-measure remains the same, indicating the robustness of the voting χ^2 GOF method. This in contrast to the standard χ^2 GOF method.

[1] http://sumo.intec.ugent.be

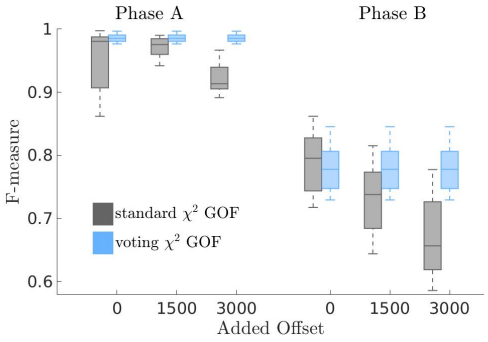


Fig. 4: The F-measure when detecting events with the standard χ^2 GOF, and voting χ^2 GOF.

When comparing the running time of SBO and the brute-force approach, it is found that for the standard χ^2 GOF method it is reduced from approximately 124 days to 5.9 hours, and for the voting χ^2 GOF method from 189.2 days to 45 hours, resulting in substantial speed up factors.

VI. Conclusion

This paper focuses on the robustness of event detection methods in NILM on real-life power traces (e.g. traces containing high power consuming devices and having a high base load). A novel event detection method (the voting χ^2 GOF method) is proposed and benchmarked to the standard χ^2 GOF method. A performance increase of $7-12\%$ in terms of F-measure is observed. The parameters of the methods are trained with the proposed surrogate-based optimisation which runs 100.000 times faster than the standard brute force approach.

References

[1] Anderson, K., et al., *Event detection for non intrusive load monitoring.*, 38th Annual Conference on IEEE Industrial Electronics Society. IECON, 2012, 3312-3317.

[2] Trung, K. N., et al. , *Event Detection and Disaggregation Algorithms for NIALM System.*, Proceedings of 2nd International Non-Intrusive Load Monitoring (NILM) Workshop. 2014.

[3] Jin, Y., et al., *A time-frequency approach for event detection in non-intrusive load monitoring*, SPIE Defense, Security, and Sensing. International Society for Optics and Photonics, 2011, 80501U-80501U).

[4] Kazmi, A. H., et al. , *Real-Time Event Detection for Energy Data Streams.*, Ambient Intelligence, 2014. 221-225.

[5] Berges, M., et al., *"Learning systems for electric consumption of buildings."*, ASCI international workshop on computing in civil engineering (38), 2009.

[6] Yang, C. C., et al., *A systematic approach to ON-OFF event detection and clustering analysis of non-intrusive appliance load monitoring.*, Frontiers in Energy 9.2, 2015, 231-237.

[7] Wong, Y. F., et al., *Recent approaches to non-intrusive load monitoring techniques in residential settings.*, Computational Intelligence Applications In Proceedings of the IEEE Symposium on the Smart Grid (CIASG), 2013, 79-79.

[8] Wu, T., and Srivastava, M., *Low-cost appliance state sensing for energy disaggregation.*, Proceedings of the 4th ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, 2012,53-55.

[9] Abramowitz, Milton, and Irene A. Stegun. Handbook of mathematical functions: with formulas, graphs, and mathematical tables. Vol. 55. Courier Corporation, 1964.
[10] Couckuyt, I, et al., *Surrogate-based infill optimization applied to electromagnetic problems.*, International Journal of RF and Microwave ComputerAided Engineering 20.5, 2010, 492-501.
[11] Makonin, S., and Popowich, F., *Nonintrusive load monitoring (NILM) performance evaluation.*, Energy Efficiency 8.4, 2015, 809-814
[12] Couckuyt, I., et al. , *Kriging, co-kriging and space mapping for microwave circuit modeling.*, Proceedings of the 41st European. IEEE Microwave Conference (EuMC), 2011, 444-447.
[13] Mehari, M. T., et al., *Efficient global optimization of multi-parameter network problems on wireless testbeds.*, Ad Hoc Networks 29, 2015, 15-31.
[14] Singh, P., et al., *Accurate hotspot localization by sampling the near-field pattern of electronic devices.*, IEEE Transactions on Electromagnetic Compatibility 55.6, 2013, 1365-1368.
[15] Gorissen, D., et al., *A surrogate modeling and adaptive sampling toolbox for computer based design.*, The Journal of Machine Learning Research 11, 2010, 2051-2055.
[16] van der Herten, J., et al. , *Adaptive classification under computational budget constraints using sequential data gathering.*, Advances in Engineering Software, 2016, 137-146
[17] Anderson, K., et al., *BLUED: A fully labeled public dataset for event-based non-intrusive load monitoring research*, Proceedings of the 2nd KDD workshop on data mining applications in sustainability (SustKDD), 2012,1-5.