Evolving Non-Intrusive Load Monitoring

Dominik Egarter[1], Anita Sobe[2], and Wilfried Elmenreich[1,3]

[1] Institute of Networked and Embedded Systems / Lakeside Labs
Alpen-Adria-Universität Klagenfurt, Austria
[2] Institut d'informatique, Université de Neuchâtel, Switzerland
[3] Complex Systems Engineering, Universität Passau, Germany
dominik.egarter@aau.at, anita.sobe@unine.ch, wilfried.elmenreich@aau.at

Abstract. Non-intrusive load monitoring (NILM) identifies used appliances in a total power load according to their individual load characteristics. In this paper we propose an evolutionary optimization algorithm to identify appliances, which are modeled as on/off appliances. We evaluate our proposed evolutionary optimization by simulation with Matlab, where we use a random total load and randomly generated power profiles to make a statement of the applicability of the evolutionary algorithm as optimization technique for NILM. Our results shows that the evolutionary approach is feasible to be used in NILM systems and can reach satisfying detection probabilities. **Keywords:** Evolutionary Algorithm, Knapsack Problem, Evolution, Non-Intrusive Load Monitoring, NILM

1 Introduction

With the upcoming of decentralized regenerative energy sources, the amount of available energy at a particular time and, due to network capacity constraints, location becomes dependent on the current weather situation (photovoltaic production depends on amount of sunshine, windmill-powered plants on wind speed). One way to mitigate this issue is to provide energy storage (e. g., by batteries, pumped-storage hydropower plants, conversion to methane, etc). The other way is shaping the energy consumption at the consumer side. A typical household contains hundreds of electric appliances, whereof a few dozen are relevant in terms of energy consumption. In order to keep the convenience level for the customer high, we need an intelligent control system that identifies devices currently turned on and proposes minimal-invasive changes to their usage. To get this information, each relevant appliance could be equipped with a smart meter or an embedded communication and control interface able to deliver power information and characteristics [5]. Upgrading all devices in a current household this way would be painstaking and costly. An alternative approach is non-intrusive load monitoring (NILM)[8], which determines and classifies individual appliances based on characteristic load profiles. For identification only a *single* smart meter measuring the total power consumption with appropriate timely resolution is sufficient. NILM extracts features like active power, frequency etc., classifies appliances and identifies appliances by matching the measured data to a reference database. Thus, the identification can be described as an optimization problem of superimposed power profiles. Possible solutions for this problem are optimization techniques like integer linear programming [16] or pattern recognition methods like artificial neural networks [4]. In recent years, the technique of NILM has been extended and improved, but up to now no universal solution has been developed [18].

We propose an evolutionary optimization approach that identifies a variable number of appliances by their given power profile. The idea is that the potential appliance profiles (out of a database) have to be matched with the given power profile with minimum error [2]. The presented problem is related to the Knapsack problem, which is NP-hard [14, 6]. Possible techniques to tackle the Knapsack problem are either exact, heuristic or meta-heuristic solutions [11]. Genetic algorithms have successfully been used for handling the Knapsack problem during the last twenty years. Implementations are ranging from solving the simple 0-1 knapsack problem [15] to more lavish techniques like hybrid optimization [17] and multidimensional Knapsack problems techniques [9].

In the context of NILM, the genetic algorithm is typically used for detecting features and patterns of appliance power profiles [3, 1] and for optimizing existing parameters which are used in fuzzy systems [13]. Furthermore, Leung, Ng and Cheng presented in [12] a possible approach to use the genetic algorithm to identify appliances. In their paper they grouped power signatures out of one load signature data set into the groups sinusoid, quasi-sinusoid and non-sinusoid load signatures by averaging 50 consecutive one-cycle steady state current waveforms. They built a composition of load signatures of the same group between each other and a composition of load signatures among the groups. Finally, they used the genetic algorithm to identify the wanted load signatures. In contrast to this approach, we use the entire power load signal of a household over two hours and not the mean current waveform of 50 consecutive one-cycles. Further, we do not split up the appliances into groups. We randomly generate common power profiles in steady state and detect these power profiles in a random superimposed composition of power profiles over a time window of two hours. Accordingly, we make a statement about which appliance have been used and also at which point in time. The remainder of this paper is organized as follows: in Section 2 we describe the optimization problem of overlapping power profiles in more detail and how it can be solved with the help of the evolutionary algorithm. In Section 3 we evaluate the presented genetic algorithm by different test scenarios like algorithm dependence on the number of wanted power profiles or the detection behavior under the influence of noise. Finally, in Section 4 we conclude this paper and present future work.

2 Evolutionary Appliance Detection

The knapsack problem is a well-known optimization problem with the aim of packing a set of n items with a certain weight w_i and profit d_i into a knapsack of capacity C in the most profitable way. If it is possible to place a item into the knapsack without exceeding the capacity C by using $x_i \in \{0, 1\}$, which is responsible for whether or not a certain item is used, a profit d_i is earned. This context can be summarized as follows:

$$\text{maximize} \sum_{i=1}^{n} d_i \cdot x_i \tag{1}$$

$$\text{subject to} \sum_{i=1}^{n} w_i \cdot x_i \leq C, \tag{2}$$

The problem of packing items into a desired shape can easily be compared to the appliance detection and classification in NILM systems. NILM has the major aim of detecting and identifying appliances according to their own power profile P_i in the measured total load $P(t)$. The power profiles P_i are characterized by their power magnitude m_i and time duration τ_i and the total load is given by:

$$P(t) = \sum_{i=1}^{n} P_i \cdot a_i(t) + e(t), \tag{3}$$

where n is the number of known and used appliances, $a_i(t) \in [0, 1]$ represents the state timing vector of the appliance being on $(a_i(ts) = 1)$ at switching time t_s or off $(a_i(t) = 0)$ and $e(t)$ describes an error term. Therefore the general optimization problem of NILM can be formulated as the minimum error $e(t)$ of the total power load and the composite appliance power profiles P_i:

$$e(t) = \arg\min \left| P(t) - \sum_{i=1}^{n} P_i \cdot a_i(t) \right|. \tag{4}$$

In contrast to the traditional knapsack problem, which allows only bounded values smaller than capacity C, we allow positive and negative error values and take the absolute error value for our fitness evaluation. Turning back, a NILM system tries to find the right switching points $a_i(t)$ and their corresponding appliances to minimize the error between the sum of superimposed appliance power profiles and the total load $P(t)$. This relates to the knapsack problem, where in the case of NILM the capacity C of the knapsack corresponds to the total load $P(t)$ and the items of the knapsack correspond to the appliance power profiles P_i. Further, we assume that the profit d_i equals 1, because we suppose that all appliances in the household are of equal importance concerning their usage. The aim of the evolutionary approach is to find a composition of power profiles P_i, which can be packed into the measured total load $P(t)$ with minimum error. Therefore, we modify the general knapsack problem by dismissing the profit maximization with an error minimization. An illustration of the basic principle can be seen in Figure 1, where a collection of possible power profiles P_i and the trend of the total power load are presented. Out of the collection a selection of power profiles P_i is met and this selection is then packed into the trend of the total power load to best approximate the trend of the total power load.

To solve this error minimization, we use an evolutionary algorithm as described in [7] with uniform mutation, single point crossover and elite selection. In detail, the used evolutionary algorithm has to evolve the set of used power profiles P_i.

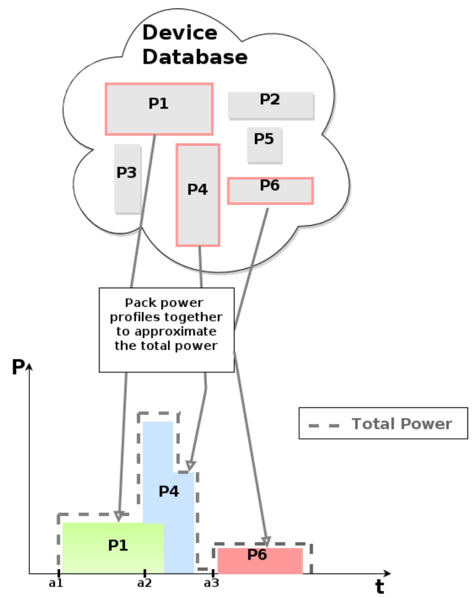


Fig. 1: Basic principle of the ON/OFF time genome appliance detection. Find a composition of saved power profiles P_i, place it at the switching times t_s and try to minimize the error between the evolved power load and the given total power load.

To be able to reduce the complexity of the optimization problem we assume that we know the starting points of the appliances' power profiles. A possible technique to detect the switching times t_s is mentioned by Hart in [8] and is called edge detection of $P(t)$. The edge detection calculates the difference of the current power signal $P(t)$ and the delayed power signal $P(t-1)$ and tries to detect the switching event by thresholding the calculated difference value. A switching time t_s is given if the difference value is larger than a predefined threshold d and accordingly $a_i(t_s)$ is set to 1.

Thus, the evolutionary algorithm examines a composition of appliance power profiles P_i, place them at the switching time t_s by multiplying P_i with its corresponding state timing vector $a_i(t)$ and approximate the total load $P(t)$. Therefore, a genome maps a set of power profiles P_{xi}, where x_i represents the index of the power profile[1] P_i stored in the database. The fitness function F_s for the optimization is given by:

$$F_s = -\left| P(t) - \sum_{i=1}^{Nb} P_{x_i} \cdot a_i(t) \right|, \tag{5}$$

[1] We assume that each power load profile P_i is only stored once in the database. The database has a size of db.

where Nb represents the number of switching times t_s and correspondingly, also the number of used appliances, because we assume that every appliance occurs only ones at different switching times t_s. According to presented fitness function F_s, the best achievable fitness value is $F_s = 0$, which corresponds to an error of 0 between the evolved power load and the total power load $P(t)$.
In the following section we evaluate the detection ability of the presented modified knapsack problem by the evolutionary algorithm.

3 Evaluation

Variable	Description	Value range
P_{elite}	Elite selection	10%
P_{mutate}	Uniform mutation	40%
$P_{crossover}$	Single point crossover	40%
P_{new}	New individuals	10%
$P_{mutateRate}$	Mutation rate	10%
G	Number of generations for the GA	500
N	Number of populations for the GA	500
P_i	Power profile P_i of appliances	
m_i	Randomly generated power magnitude[2] m_i	$m_i \in [100, 4000]$
τ_i	Randomly generated time duration[3] τ_i	$\tau_i \in [60, 3600]$
P_g	Random generated total power load over two hours in seconds resolution ($T = [1, 7200]$). Total power load equals a random set of power profiles P_i out of the database.	
db	Number of random generated and stored power profiles P_i	50
Nb	Number of used appliances	5
S	Simulation runs	10

Table 1: Parameters used for the evolutionary algorithm and the simulations

To be able to evaluate the presented evolutionary algorithm with its genome representation and fitness function, we compute simulations of the evolutionary algorithm in Matlab. The parameter properties for the evolutionary algorithm can be found in Table 1 and were determined empirically. We performed S simulation runs for each test case and generated the mean fitness $F = \sum_{s \in S} \frac{F_s}{S}$ and the mean detection probability $\bar{P}_{det}$. The mean detection probability $\bar{P}_{det}$ is given by $\bar{P}_{det} = \sum_{s \in S} P_{det}/S$, where P_{det} is given by $P_{det} = \frac{\#det}{Nb}$ and is the detection probability by simulation run. The variable $\#det$ is the counted

[2]The power magnitude m_i was chosen in a common power range of household appliances.

[3]The time duration τ_i represents a common usage of household appliances between one minute and one hour. The time window of 2 hours will produce a variety of superimposed power loads

number of correctly detected power profiles P_i and Nb is, as mentioned before, the number of switching events and also the number of used power profiles[4] P_i. Beyond that, Table 2 show how many errors $e_s = N_b - \#det$ occurred for every simulation run $s \in S$ and further, we calculate the mean error $\bar{e} = \sum_{s \in S} e_s / S$ per the simulation runs $s \in S$. For the evaluation we used the following different test scenarios:

- influence of the number of wanted and stored power profiles and
- influence of disturbances like noise and unknown power loads

to make a statement on their influence on the detection ability of the presented NILM technique, which we will describe in the following sections in more detail.

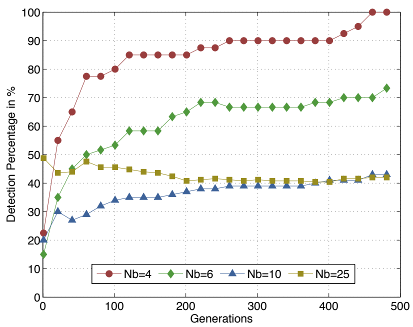
3.1 Variation of wanted and stored appliances

In Figure 2 the results for different numbers of active devices and different sizes of the power profile database is shown. First, we consider the case of varying number of power profiles. For this test scenario we chose a database size $db = 50$ and the number of wanted power profiles P_i was $Nb = [4, 6, 10, 25]$. We have taken these values of Nb to cover the cases of household common and high Nb appropriate to the size of db and the considered period of time (2 hours). Considering Figure 2, the fitness F and detection percentage Θ reach satisfying and sufficient results of a detection percentage up to 100%. We can see that the fitness value F of Figure 2(c) [4] depends on the number of devices (curves are from low Nb at top of the figure to high Nb at the bottom). The lower the number, the better the fitness, because it is harder to find a set of correctly ordered power loads of size 10 than of size 5. The evolutionary algorithm is able to find sufficient results after 100 to 200 generations, which is acceptable for the intended application scenario. The detection behavior is similar in the case of the detection percentage Θ in Figure 2(a). The lower the number of used power loads, the better is the result of detected power loads. We claim that the worst case in this example is finding 25 power loads, reasoning that our optimization problem can be seen as a problem of combining several things k out of a larger group n, where the order is not taken into consideration and accordingly, can be considered as the well known combination problem. A combination can be formulated as $\frac{n!}{k! \cdot (n-k)!}$ and the worst case for this problem is, if $k = n/2$, which is in our case with $k = Nb = 25$ and $n = db = 50$.

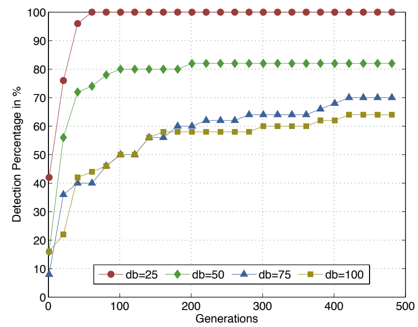
Beyond that, we varied the number of stored power profiles db in the second test scenario in Figure 2(b). For this we used a database size of $db = [25, 50, 75, 100]$ and $Nb = 5$ to make a statement regarding scalability of the

[4]Every appliance power load P_i can only be used once and accordingly, the number of wanted power profiles is the same like the number of switching times t_s

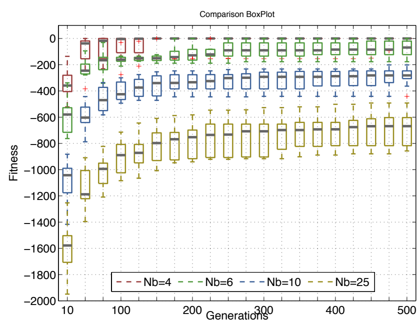
[4]The boxplot of the fitness trend for Nb variation should give a general impression how the fitness is evolving over generations and is comparable for fitness evaluation and detection probability, because the fitness has a direct relation to the detection probability.



(a) Detection percentage Θ for Nb Variation



(b) Detection percentage Θ for db Variation



(c) Boxplot of fitness F for Nb Variation

Fig. 2: This figure shows the trend for the varying $Nb = [4, 6, 10, 25]$, for varying $db = [25, 50, 75, 100]$ with $Nb = 5$ and also the boxplot of the mean fitness F for varying Nb

evolutionary algorithm. We can see that the evolutionary algorithm is able to reach a high detection percentage Θ dependent on used db. If the number of db is increased, the evolutionary algorithm evolves a lower Θ, because the search space is becoming bigger.

In addition, we present the mean error $\bar{e}$ and the mean detection probability $\bar{P}_{det}$ of the detection process in Table 2. According to this table we can claim that the lower Nb, the better the result of no errors and that the evolutionary algorithm is dependent on db. Finally, the simulations shows, that the detection depends on the characteristic of overlapping power loads. The more power loads are superimposed, the more difficult it is to make a correct decision and to minimize the error between the total power load and the evolved power load. In our test scenarios the probability of overlapping power loads is rather high, because of the chosen time duration of $\tau_i = [60, 3600]$ in a time window of 2 hours ($T = [1, 7200]$) and correspondingly, the detection algorithm still works sufficiently and satisfying.

	Detection errror e_s and mean detection probability $\bar{P}_{det}$ by								
	Nb-Variation					db-Variation[5]	with Noise[6]	with unknown[5]	
	4	5	6	10	25	100		1	2
$\bar{e}$	0	0.7	1.6	5.7	14.6	1.8	0.9	1.9	3
$\bar{P}_{det}$ in %	100	86	73	43	41	64	82	62	40

Table 2: Table for error e_s by simulation run S and the mean error $\bar{e}$ for varying $Nb = [6, 10, 25, 45]$ with $db = 50$, varying $db = 100$ with $Nb = 5$, under the influence of noise and under the influence of unknown power loads $Unkn = [1, 2]$

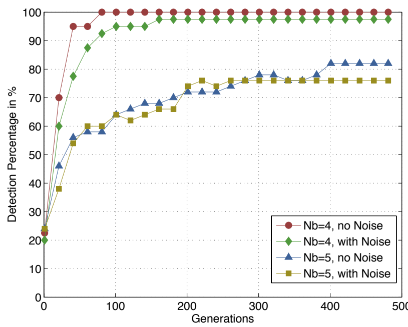
3.2 Influence of Disturbances:

To make a better statement of the ability and the quality of the presented algorithm, we examined the detection behavior of the evolutionary algorithm under the influence of noise and unknown, not stored appliances. In Figure 3(a) noise with zero mean $\mu = 0$ and the standard deviation $\sigma = \sqrt{max(P(t))}$ was added to our simulated total power load $P(t)$ and we added unknown power loads $Unkn = [1, 2, 3]$ to the total load $P(t)$ in Figure 3(b). At first, we consider the detection scenario under the influence of noise. In this scenario, the simulation results show that the detection percentage Θ in Figure 3(a) is slightly influenced by noise and therefore, we claim that the presented algorithm is robust to noise. We observe this behavior in the cases of $Nb = 4$ and $Nb = 5$. Both test scenarios show satisfying detection results of 0 errors. Further, we consider the case of adding unknown appliances to the total load $P(t)$ in Figure 3(b). This figure indicates that the detection percentage Θ depends on the number of unknown power loads[7], because the more unknown information is in the system the more complicated it is for the evolutionary algorithm to establish a correct evolved composition of power loads P_i to approximate $P(t)$. Finally, we also present the mean error $\bar{e}$ and the mean detection probability $\bar{P}_{det}$ in Table 2, which confirms our statements that our algorithm is noise robust and is dependent on the number of known and accordingly, on the number of unknown appliances.

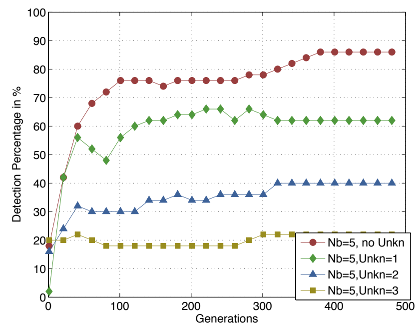
4 Conclusion and Future Work

In this paper we present an evolutionary algorithm to solve the task of detecting an appliance based on their power loads in the total load of a household over

[5] $Nb = 5$
[6] $Nb = 5$, $\mu = 0$,$\sigma = \sqrt{max(P(t))}$
[7] Nb=5 for this test scenario



(a) Trend of the detection percentage Θ with noise



(b) Trend of the detection percentage Θ with unknown power loads

Fig. 3: This figure shows the trend under the influence of noise and unknown power loads $Unkn = [1, 2, 3]$. The size of $db = 50$ and $Nb = [4, 5]$

a time window of 2 hours. Our algorithm provides promising results to detect superimposed respective power loads with up to 100% certainty. In more detail, the presented algorithm has the following detection characteristics:

- Detection percentage Θ up to 100% depending on Nb and db
- The higher Nb and the higher db, the lower the detection percentage Θ
- Sufficient results at generation $G > 100$
- Robustness against noise
- Dependent on additive unknown and the quantity of overlapping power loads

Our results show that the presented algorithm is feasible for use in NILM systems and can achieve a detection probability of 100% in case of low number of devices Nb and records used power loads even if not all power loads are detected correctly. With our algorithm applied to a real household the results can be used for tracking the used power consumption and the usage of appliances and can improve the energy-awareness concerning the energy consumption of appliances. The current version of the algorithm was tested for on/off appliances with constant time duration. In future work, we plan to extend our algorithm for arbitrary shapes of power profiles and to evaluate the approach using real appliances [10].

5 Acknowledgments

This work was supported by Lakeside Labs GmbH, Klagenfurt, Austria and funding from the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF) under grant KWF-20214 | 22935 | 24445. We would like to thank Kornelia Lienbacher for proofreading the paper.

References

[1] Baranski, M., Voss, J.: Genetic algorithm for pattern detection in nialm systems. In: IEEE International Conference on Systems, Man and Cybernetics. Volume 4. (2004) 3462 – 3468 vol.4

[2] Bijker, A., Xia, X., Zhang, J.: Active power residential non-intrusive appliance load monitoring system. In: AFRICON, 2009. AFRICON '09. (sept. 2009) 1 –6

[3] Chang, H.H., Chien, P.C., Lin, L.S., Chen, N.: Feature extraction of non-intrusive load-monitoring system using genetic algorithm in smart meters. In: IEEE 8th International Conference on e-Business Engineering (ICEBE). (2011) 299 –304

[4] Chang, H.H., Lin, C.L., Lee, J.K.: Load identification in nonintrusive load monitoring using steady-state and turn-on transient energy algorithms. In: 14th International Conference on Computer Supported Cooperative Work in Design (CSCWD). (2010) 27 –32

[5] Elmenreich, W., Egarter, D.: Design guidelines for smart appliances. In: Proc. 10th International Workshop on Intelligent Solutions in Embedded Systems (WISES12). (2012) 76–82

[6] Garey, M.R., Johnson, D.S.: Computers and Intractability; A Guide to the Theory of NP-Completeness. W. H. Freeman & Co., New York, NY, USA (1990)

[7] Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning. 1st edn. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (1989)

[8] Hart, G.: Nonintrusive appliance load monitoring. Proceedings of the IEEE **80**(12) (1992) 1870–1891

[9] Hoff, A., Løkketangen, A., Mittet, I.: Genetic Algorithms for 0/1 Multidimensional Knapsack Problems. Proceedings Norsk Informatikk Konferanse, NIK '96 (1996)

[10] Kolter, J.Z., Johnson, M.J.: REDD : A Public Data Set for Energy Disaggregation Research. In: Proceedings of the SustKDD workshop on Data Mining Applications in Sustainability. (2011)

[11] Lagoudakis, M.G.: The 0-1 Knapsack Problem An Introductory Survey. Technical report

[12] Leung, S.K.J., Ng, S.H.K., Cheng, W.M.J.: Identifying Appliances Using Load Signatures and Genetic Algorithms. In: International Conference on Electrical Engineering (ICEE). (2007)

[13] Lin, Y.H., Tsai, M.S., Chen, C.S.: Applications of fuzzy classification with fuzzy c-means clustering and optimization strategies for load identification in nilm systems. In: 2011 IEEE International Conference on Fuzzy Systems (FUZZ). (2011) 859 –866

[14] Martello, S., Toth, P.: Knapsack problems: algorithms and computer implementations. John Wiley & Sons, Inc., New York, NY, USA (1990)

[15] Singh, R.P.: Solving 0/1 Knapsack problem using Genetic Algorithms. In: IEEE 3rd International Conference on Communication Software and Networks (ICCSN), IEEE (2011) 591–595

[16] Suzuki, K., Inagaki, S., Suzuki, T., Nakamura, H., Ito, K.: Nonintrusive appliance load monitoring based on integer programming. In: SICE Annual Conference. (2008) 2742 –2747

[17] Wien, S.v., Troya, J.M., York, N., Cotta, C., Troya, J.M.: A Hybrid Genetic Algorithm for the 0-1 Multiple Knapsack Problem (1998)

[18] Zeifman, M., Roth, K.: Nonintrusive appliance load monitoring: Review and outlook. IEEE Transactions on Consumer Electronics **57**(1) (2011) 76–84