Disaggregation of digital meter data for synthetic load profile classification

Toon Bogaerts[1], Stef Jacobs[1,2], Sara Ghane[1], Freek Van Riet[2], Wim Casteels[1], Siegfried Mercelis[1], Ivan Verhaert[2], Peter Hellinckx[1]
[1] University of Antwerp – imec, IDLab - Faculty of Applied Engineering
Sint-Pietersvliet 7, 2000 Antwerp, Belgium
[2] University of Antwerp, EMIB Research Group,
Groenenborgerlaan 171, B-2020 Antwerp, Belgium

Abstract

The electrical consumption has to be taken into account in building simulations. Empirically-based profiles are required, which can be generated by central measurements and using non-intrusive load monitoring (NILM) for disaggregation. In this work, we present an overview of NILM techniques, a comparison between two frequently used deep neural networks for individual appliance identification and we investigate the influence of the sampling rate with regards to the accuracy. Our best performing neural network is a combination of convolution and long-short-term memory networks. Furthermore, the sampling rate has a significant influence on the performance of neural networks in this context. There should be a trade-off between sampling rate and efficiency when applied in real-world devices.

Key Innovations

- A comparison between frequently used deep neural networks for individual appliance identification
- An overview of Non-Intrusive Load monitoring techniques
- Investigation into the relationship between sampling rate and accuracy for individual appliance identification when using deep learning starting from the current implemented 0,1Hz up to 60Hz.

Practical Implications

The combination of convolution neural networks and long-short term memory neural networks outperforms the state-of-the-art non-intrusive load monitoring (NILM) techniques for disaggregating the central measured electricity consumption.

Introduction

The building stock in Europe is responsible for around 40% of the consumed energy and 36% of the CO_2 emissions (Burman et al., 2014; European Commission, 2008). The climate goals of the European Union by 2050 require increased use of renewable energy sources (RES) and the development of efficient energy systems. Higher insulation rates of buildings, advanced control strategies and more insight into energy components and systems are the first steps to a zero-carbon future. To develop new concepts or to gain more insight, building simulations are an appropriate method. In these simulations, every aspect of an energy system can be studied and different sensitivity analyses can be elaborated.

Typically an energy system can be subdivided into three main parts, namely the producers, the distribution, and the consumers. Examples of the energy demand (i.e. the consumers) in buildings are space heating, space cooling, domestic hot water (DHW), electricity consumption, etc. In this context, different trends can be seen regarding the energy demand.

Firstly, the relative share of heating is decreasing, while the relative share of domestic hot water (DHW) and cooling is increasing (European Commission, 2011; Rijksdienst Voor Ondernemend Nederland, 2018). This is because of the increased buildings' insulation rate and reduced ventilation losses. Consequently, the internal heat gains will play a key role in the heating and cooling demand of dwellings and cannot be neglected when performing an energetic analysis of energy concepts through simulations. These internal heat gains have two main sources, namely the occupancy and heat losses inside the building, such as from the electrical appliances. It can be stated that the electrical consumption of these appliances is linear to their heat losses (i.e. the internal heat gains) and depends on the efficiency of the appliances. Thus, by considering the electrical consumption of the electric appliances, and knowing the efficiency, these internal heat gains can be simulated in the building simulations.

Secondly, as mentioned before, more RES are being implemented in the electrical grid. However, these sources are intermittent and unstable. For instance, a solar panel only generates electricity when the sun is shining. As a consequence, the electrical grids need flexibility, e.g. through energy storage systems or by demand-side management (e.g. Gelazanskas et al., 2014). With respect to this, the electrification of heating and cooling offers an interesting potential (Thomaßen et al., 2021; Neirotti et al., 2020), which can be optimized throughout building simulations. For instance, a heat pump could produce thermal energy when an overload of electricity is available and this thermal energy can be stored in thermal storage tanks. Later, heat conversion to electricity is possible with cogeneration of heat and power (CHP) or the heat is directly usable for heating purposes (Thomaßen et al., 2021; Neirotti et al., 2020).

For these reasons, the electricity consumption of buildings is strongly present in the buildings' total energy consumption and has to be included in building simulations. In this respect, synthetic load profiles (SLPs) can be used. These are databases of the electrical consumption per chosen time interval throughout a year, which should meet two criteria.

First, the generation of SLPs should be based on in situ measurements, to be empirically-based and thus to represent actual user behaviour. Second, the SLPs contain labelled data of the different appliances. In this way, the dataset can be adopted to different case studies, where the electrical appliances‘ composition can vary. To meet both criteria, measurements of every single appliance are required. However, this is labour and cost-intensive. An alternative is to perform a central measurement and disaggregate the data into the data of the individual appliances, a concept which is called Non-Intrusive Load Monitoring (NILM). While great efforts have been made to investigate the potential of NILM-algorithms, a comparison between Deep Learning architectures for load classification is still missing. This will be further discussed in the second section of this paper.

Another hurdle for the real-life application of NILM is the lack of knowledge on the use of high temporal resolution data. In Belgium, a high-frequency read-out port, the S1-gate, is being added to the smart energy meters delivered by Fluvius. This gate has a sampling frequency of 2000 up to 4000Hz and provides very detailed information on the current and voltage sine-wave in the different phases. However, the data is not yet interpreted by this high-frequency sensor. When this data is used and interpreted in combination with the standard P1-gate’s data (at a sampling rate of 0,1Hz) an advanced control and detailed feedback is possible. The roll-out of such smart energy meters is ongoing in Europe (European Commission, 2014), thus it will be possible to develop an SLP using the S1-gate and a NILM technique. Particularly, the higher frequencies might be useful to distinguish two electrical appliances with a similar usage profile. In this respect, it is probable that the starting currents, which are present for a short time, are different. However, a higher sampling rate requires more data storage and calculation speed. To the best of the authors’ knowledge, no trade-off between the high sampling rate and the accuracy of the NILM algorithm has been investigated.

Because of the discussed shortcomings of existing NILM techniques and the lack of knowledge about the effect of temporal resolution, this paper focuses on three aspects:

1. The paper starts with an overview of NILM-techniques.
2. Based on “1.”, different techniques are compared including convolutional neural network, long-short term memory (LSTM), and a combination of both. This includes the optimization of the hyperparameters.
3. Finally, an evaluation is made on the added value of high temporal resolution readouts and the current standard P1-gate (0.1Hz).

All three steps contribute to developing a tool to measure electricity consumption to set up SLPs. The tool itself might be the subject of future work.

Topic “1.” is discussed in the second section. In the third section, the data set that is used for topics “2.” and “3.” is discussed and the results are given in the fourth section. Section five concludes this research.

Overview of NILM techniques

Non-Intrusive Load Monitoring (NILM) was first introduced by Hart et al. (1992). It is a technique for energy disaggregation, which involves the task of decomposing the total aggregated energy consumption of a digital meter into individual electrical appliances. Applying machine learning methods to detect signatures of electrical devices, could assist NILM in disaggregating digital meter data and appliance identification. This can be done by adopting supervised or unsupervised approaches.

Supervised learning algorithms usually require a huge amount of labeled data for training which is not cost-effective in NILM. Examples of such approaches that applied to the NILM domain include k-nearest neighbour (kNN) (Khan et al., 2019; Yang et al., 2018), naïve Bayes classifier (Yang et al., 2018), support vector machine (SVM) (Figueiredo et al., 2012; Gong et al., 2019), decision tree (Lin et al., 2020), artificial neural networks (ANN) (Xu et al., 2014), etc. Unsupervised learning algorithms, such as clustering (Barsim et al., 2014), expectation-maximization algorithm (EM) (Figueiredo et al., 2014), etc., do not involve manual data labeling. Another method that could be utilized as both supervised and unsupervised methods is the hidden Markov Model (HMM) and its variants (Bonfigli et al., 2017; Liu, 2020). The efficiency of the NILM system in both supervised and unsupervised cases is attributed to the employed features that could be obtained through a feature extraction process, which is the process of transforming input data into a set of useful features. On the other hand, deep learning techniques do not require manual feature extraction and are capable of extracting most discriminative and robust features automatically from data through neural networks. Therefore, in the past few years, to improve meter data disaggregation performance, the possibility of applying them to the NILM has been investigated. The most used algorithms in this area are deep neural network (DNN), convolutional neural network (CNN), recurrent neural network (RNN), long short-term memory (LSTM), and denoising auto-encoders (DAE).

Deep learning techniques have been first applied to NILM in 2015 (Kelly et al., 2015). They applied LSTM, DAE, and a custom DNN architecture to energy disaggregation. In the work by Bonfigli et al. (2018) load disaggregation problem was treated as a noise reduction problem and they presented a DNN architecture based on DAE. A framework to identify individual appliances from the aggregated data based on DAE and LSTM was presented by Wang et al. (2019). The DAE was utilized to reconstruct the single appliance signal and LSTM was applied to identify that the signal belongs to which electrical appliance. Xia et al. (2020) regarded the disaggregation problem as a signal separation process and presented a deep composite LSTM. To tackle the time-dependency problem in multi-state appliances for non-intrusive load disaggregation, they introduced an encoder-separation-decoder structure that consists of deep LSTMs.

Kaselimi et al. (2020), proposed a model for robust energy disaggregation in the presence of noise, which was based on generative adversarial networks (GAN). The discrimination was done using a gated recurrent unit (GRU) to improve the robustness and precision accuracy of the model. Kaselimi et al. (2019) presented a deep recurrent multi-input/multi-output regression model based on CNN. The recurrent aspect helps to capture the temporal interdependencies of the power signals more effectively. The multi-channel properties of this model improve the accuracy by providing more information.

Harell et al. (2019) proposed a 1-D dilated causal CNN based on WaveNet for load disaggregation. They showed the strength of a CNN in learning from various energy features. A convolutional sequence to sequence architecture was proposed by Chen et al. (2018). To extract information from aggregated power consumption, gated linear unit convolution blocks and max-pooling layers were utilized. They further refined the output using residual blocks of fully connected layers. A multiscale deep residual neural network architecture based on dilated convolution for load disaggregation was presented by Zhou et al. (2020). Using residual blocks ensures that the network avoids degradation problems due to the vanishing/exploding gradient that is caused by increasing the number of layers and improves performance. In a work by De Baets et al. (2018), weighted pixelated image of the voltage-current trajectory was used as the input for a CNN which then extracts the most representative spatial features for appliance classification. Kaselimi et al. (2020) introduced a hybrid CNN-LSTM architecture that is both spatially and temporally deep. Feature extraction was done using CNN and sequence to sequence modelling is performed by LSTM.

Basu et al. (2016) implemented data-driven event-based NILM techniques and evaluated their performance based on two different sampling rates (10 seconds and 15 minutes). They mainly focused on residential building's sector and compared the results for two sampling rates for available appliances in a house, and they obtained a better result with 10 seconds sampling rate. Huchtkoetter, J., & Reinhardt, A. (2019) assessed the impact of the temporal resolution on the accuracy of load signature event detection, by comparing event detection methods which are based on chi-square method and threshold analysis. The results revealed that higher resolutions increase the chance of being falsely detected as events. Besides, they indicated sampling rates between 925Hz and 1.2kHz as suitable for event detection. The evaluation was solely done based on the F1 score. Lynch, S., & Longo, L. (2017) conducted a study on the effect of sampling rate on disaggregation accuracy of Hidden Markov Model-based (HMM) algorithms. In other words, they focused on finding a relationship between sampling rate (in a range of 1 second to 6 minutes) for feature extraction and selection in NILM and HMM model accuracy. They used REDD dataset as a baseline for comparison and they considered precision, recall, and F1 score as evaluation metrics. Their findings on REDD dataset, verified the correlation between sampling frequency and model accuracy at the building block level of HMM, meaning that higher sampling rates generally lead to more accurate results. Comparing to the state-of-the-art, we distinguish our research using the current available P1-gate frequency and explore the possible improvement higher sampling rates can generate when using deep learning with regards to NILM applications. Higher sampling rates will be available with the deployment of the S1-gate in future applications. The importance of higher sampling rates will be investigated with respect to the current P1-gate (0,1Hz) up to 60Hz (dataset limitation) which becomes possible with the future S1-Gate (up to 4kHz) as well as a comparison between the two most frequently used deep learning architectures for individual appliance identification, namely an LSTM network, a one-dimensional convolution network, and a combination of both.

Description of the data set

An algorithm requires a dataset to train itself and learn the environment in which it will operate. For this research, we use an online-available dataset, namely BLUED (Anderson et al., 2012). It contains the electrical information of 43 electrical appliances, ranging from larger appliances, such as a refrigerator and an oven, to smaller appliances, such as the lights, in an American single-family house during one week. The house has a two-phase (phase A and B) electricity network with a neuter. Different types of information are recorded in the dataset as listed in Table 1.

Table 1: Overview of the available data types in the BLUED dataset.

Type	Frequency	Unit
Current A	12kHz	Amps [A]
Current B	12kHz	[A]
Voltage A	12kHz	Volts [V]
Active power	60Hz	Watt [W]
Reactive power	60Hz	Watt [W]
Time		hh:mm:ss
date		yy/mm/dd
Event timestamp		W at hh:mm:ss

The current of the two phases and the voltage of the whole house are monitored at a high frequency (12kHz), while the total active power is computed at 60Hz. Furthermore, every state transition is labelled and time-stamped per component and a total of 4817 events are registered. An event is defined as an appliance's power consumption of 30 W or more for at least 5 seconds. In total, 2335 events are known, while 2482 events are from an unknown source. Figure 1 gives an overview of the known-registered events for the different appliances of interest within this research.

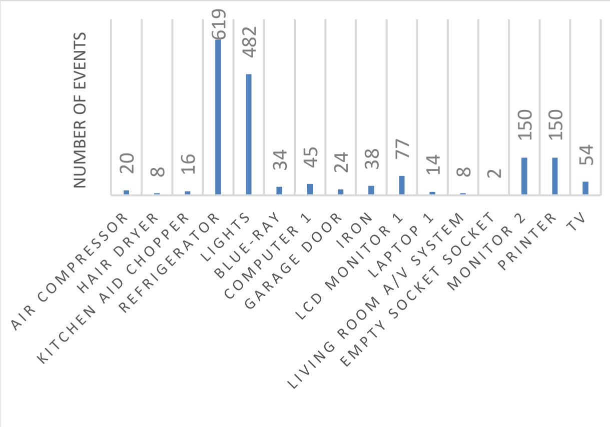

Figure 1: histogram with the total number of events registered per appliance.

The electrical appliances can be categorized into three types: 1) an on/off machine, 2) a finite state machine and 3) an infinite state machine. An on/off machine can switch between two operation modes, namely on or off. A finite state machine has multiple operation modes that are countable. For example, a hairdryer has four different ventilation rates. An infinite state machine can vary through all possible modes between off and the maximum power, e.g. dim lighting. The algorithms used in this paper neglect the category of the appliance and will solely focus on classifying a change in power with a specific appliance.

Pre-processing on the dataset

As described above, the BLUED dataset contains measurements of appliances from a central measurement. Labels of corresponding appliances are assigned to the time series when there is a power difference of 30W that lasts for at least 5 seconds. Furthermore, the current and voltage are recorded at a high sampling rate of 12kHz. Unfortunately, the precision of the timestamps describing the events and the measurements are different. In combination with the absence of the voltage of phase B and inconsistency between different files describing the dataset, we concluded to solely focus on the aggregated 60Hz measurements. These data points are more stable, but also have some issues. The first issue arose when the index of the measurements is reset to 0 after 50000 samples. This can result in a mismatch between the continuous timestamp of the events and the index of the measurements. A second issue can be seen at the end of the dataset. Here an inconsistency occurs within the measurements which results in a desynchronization between the events and measurements. Around 400 of the last events are inconsistent and thus dropped from the dataset.

After further investigating the data, it can be deduced that the classes are imbalanced. Or in other words, some appliances occur more frequently than others. Some classes are also present with less than 10 occurrences which will influence the model's capability to identify them. To solve this imbalance the model is trained using oversampling. This technique creates a container for each class and uniformly picks a random sample from each container to train the model. The evaluation of the models is done using a separate test set. For each class, 80% of the occurrences are used to train the model and 20% for evaluation.

The class of 'Unknown appliance' will be dropped as this does not contribute to individual appliance classification. Furthermore, classes describing circuits are also removed. These classes are an aggregation of multiple appliances and thus in similar time intervals, the output can be both a circuit and an appliance on this circuit. If these classes would stay present the problem would change into a multiclass classification which is not within the scope of this research. Finally, individual lights are aggregated to a single lights class, as these are mostly identical and correspond to the class of consumers. Afterward, the data is scaled down using a standard scaler, the data is scaled to increase numeric stability within the neural network. Figure 2 illustrates an example input for the neural networks originating from the refrigerator turning on.

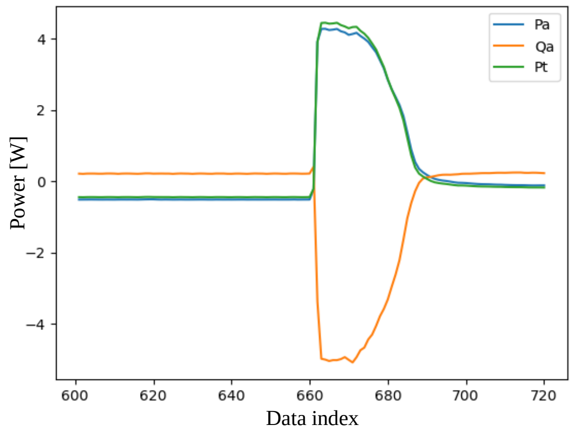

Figure 2: Example of a scaled-down event sequence originated from the refrigerator turning on. Pa: scaled-down active power, Qa: scaled-down reactive Power, Pt: scaled-down total power.

Evaluation of sampling rates

The evaluation of a high sampling rate (i.e. the future S1-gate) is established by down-sampling. The dataset already possesses the power at a high frequency (60Hz). For this reason, the power is down-sampled to a frequency of 30Hz, 15Hz, 1Hz, and 0,1Hz (P1-gate sampling rate). Afterward, the accuracy of the temporal resolutions is compared to each other. In this way, we can evaluate if the high sampling rate is beneficial or if the 0,1Hz is sufficient to provide a reliable labelling algorithm.

Results

In this section, the experiments conducted will be presented. Firstly, a comparison of deep learning architectures for appliance identification will be discussed. This is followed by an experiment looking at the influence of sampling rates.

Comparing deep learning architectures

The architectures compared in this paper are namely: an LSTM network, a one-dimensional convolution network, and a combination of both. After optimization, the best performing parameters and a description of the models are presented in Table 2. Convolutional neural networks are a class of deep neural networks that are mainly used in image recognition tasks. These networks utilize filters that traverse the input data and detect patterns. With each iteration, these patterns increase in complexity. Due to the sliding filter approach, convolutional neural networks are translational invariant. In other words, the location of a pattern in the input data does not influence the models' performance. The convolutional network used in this paper is a one-dimensional convolution neural network, where filters are applied to the sequence of measurements for multiple iterations. The number of filters defines the number of patterns that can be detected, while the window size sets the size of these patterns. The features extracted by the filters are aggregated using global average pooling. The other architecture investigated is an LSTM. This architecture is a recurrent neural network. This type of neural network utilizes feedback connections, allowing the detection of patterns in sequential data. These networks are commonly used for speech recognition, natural language processing, etc. An LSTM is a specific recurrent neural network designed to minimize the vanishing/exploding gradient phenomena (Pascanu et al., 2013). A cell is a single LSTM, a layer utilizes more cells in parallel to increase the number of patterns that can be detected. Multiple layers are used to increase the complexity of the patterns that can be detected. A final model combines both discussed architectures. This model aims to utilize the strong point of each model, starting with the convolutional neural network followed by the LSTM to generate the predicted output. Each model has a fully connected neural network as a final layer with a SoftMax activation function. This ensures the model outputs a certainty for each class using the one-hot labelling method.

Table 2: Model descriptions

Model	Layers (units/filters)	Dropout rate	Window size	# parameters
LSTM	LSTM (64) LSTM (64)	30%		50 757
Convolution	1D Conv (64) 1D Conv (64) 1D Conv (128)		4 8 16	165 509
Convolution-LSTM	1D Conv (64) 1D Conv (64) LSTM (64) LSTM (64)	 30% 30%	4 8	100 037

The models are trained and evaluated on the 60Hz data series containing the active power, reactive power, and total power. Each event is transformed to a sequence using 1 second before the labelled event and 1 second after. This results in a sequence of 120 samples for each event. The models are trained using the root-mean-square propagation optimizer in combination with a categorical cross-entropy loss function. This function aims to isolate predictions to a single class, which is ideal for our use case. The models are trained for 100 epochs with 20000 randomly chosen samples each epoch on a NVIDIA GeForce GTX 1660 (6GB VRAM), intel I7-9700K CPU and 16GB RAM. Due to the relative small number of parameters and parallelization of the GPU, training time averages around 5 minutes with a neglectable difference for each model.

Table 3: Accuracy scores

Accuracy	**Phase A**		**Phase B**	
Model	**Train**	**Test**	**Train**	**Test**
LSTM	97 %	90 %	88%	53%
Conv	61%	53%	55%	38%
Conv-LSTM	99%	93%	90%	59%

Table 3 contains the accuracy scores of the different algorithms applied to both phases present in the dataset. The accuracy score represent the percentage of samples correctly predicted from the 20000 randomly chosen samples. At first sight, it can be seen that the accuracy on phase A is generally better when compared to phase B. This can be explained by the number of appliances to detect in each phase, phase A containing only 6 classes while phase B has 12 classes. Another reason for this difference in accuracy can be explained with the similarity of the appliances which will be further discussed below when the confusion matrix of phase B is discussed. From these results, it can be concluded that the combination of both convolution and LSTM outperforms the stand-alone version of these networks. Even though the convolution model has the most trainable parameters, it is not the best performing model. This can be explained with the window size. On one hand, when using smaller windows there is a chance of not detecting larger patterns when traversing the network. On the other hand, when the window size becomes too large, small indications can be overlooked. A trade-off has been made between these restrictions.

Figure 3 illustrates the confusion matrix of phase A on the test data when predicted using the Conv-LSTM network. The air compressor is 50% wrongly identified due to the underrepresentation in the dataset, there are only 5 training samples and 2 testing samples. Due to the limited amount of samples, this class is more difficult to identify. Another interesting fact can be seen when looking at the lights and refrigerator. A total of 27% of the refrigerator samples are classified in the lights category. When looking in detail at these misclassifications, it can be concluded that when the refrigerator turns off, the drop in power is similar to the drop in power from a light. This could be solved with more data and/or more training time of the model.

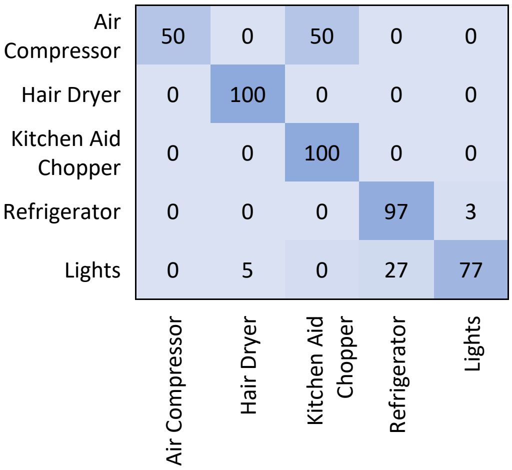

Figure 3: Confusion matrix (%) phase A, Conv-LSTM. Horizontal: predicted label, vertical: True label

Figure 4 illustrates the confusion matrix of phase B on the test data when predicted by the Conv-LSTM network. At first sight, this matrix is more disperse when compared to the confusion matrix of phase A. There are some classes which are frequently misclassified, e.g. the monitors, laptops, and computers.

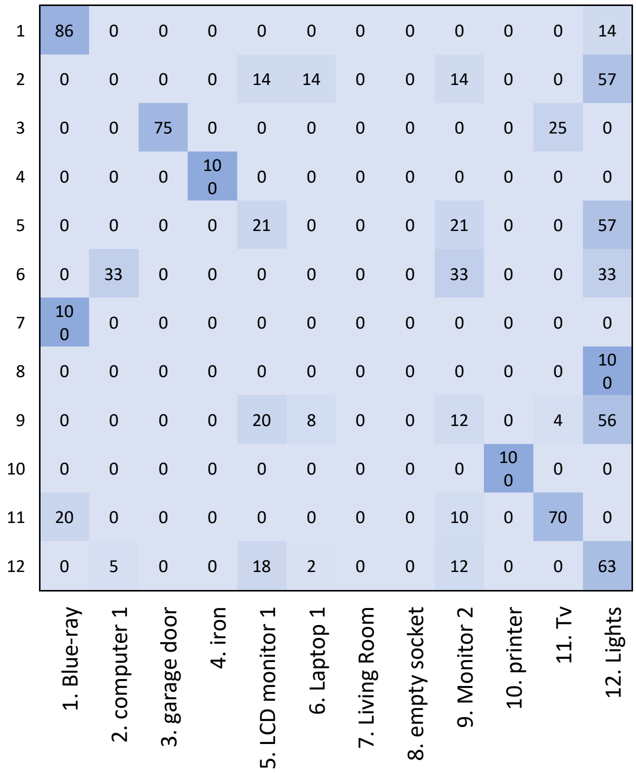

Figure 4: Confusion matrix (%) phase B, Conv-LSTM. Horizontal: predicted label, vertical: True label

These appliances have similar load profiles when switching on or off. As these appliances have a similar function they might be combined to a single category similar to the lights category. Another solution to having better accuracy on these appliances is to ensure they are not overlapping when starting up. For example, when turning on a computer the monitor will switch on shortly after, this can confuse a deep neural network as it would have difficulties separating these power profiles. Furthermore, it can be seen that the lights category is the most commonly predicted class. When the model is unsure it will predict this class causing faulty predictions. More data and longer training times might further improve the accuracy across all classes.

Evaluation of sampling rates

This experiment will focus on the influence of sampling rate in function of accuracy. The sampling rates that are going to be evaluated are 60Hz, 30Hz, 15Hz, 1Hz, and 0,1Hz (currently implemented P1-gate). For each sample rate, the Conv-LSTM architecture is retrained and evaluated using the independent test data. Sample rates 60Hz, 30Hz, and 15Hz are trained on the same time interval, namely 1s before and 1s after the labelled event. When using 1Hz and 0,1Hz, the sequence contains three data points sampled closest to the event label. Only three data points are sampled because the relevant information drops when moving further away in time from the labelled event. Down sampling is executed using the average values from the higher sample rate. The results are presented in Figure 5. Surprisingly, a small improvement can be seen on 30Hz for phase A and 15Hz for Phase B. This could be the result of the down sampling method averaging the values and thus reducing noise on the signal.

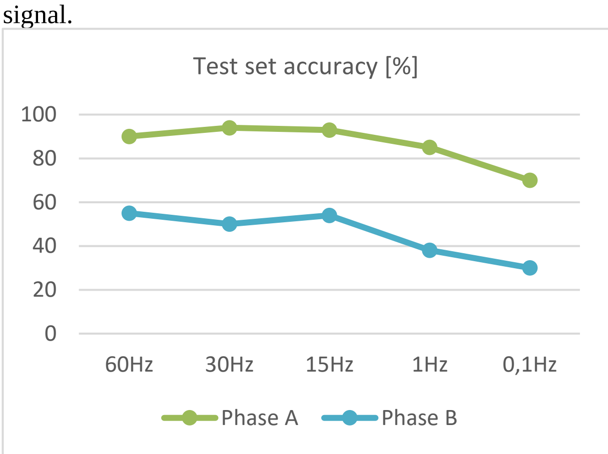

Figure 5: Accuracy in function of sample rate

Intuitively the accuracy drops on the lowest sampling rates. The information within the sequence is significantly reduced. For example, 60Hz contains 120 data points while 0,1Hz only contains 3 data points. These results indicate that a sample rate higher than 15Hz can give good results for individual classification of appliances from a central measurement. Furthermore, lower sampling rates decrease the complexity of storing the data and efficiency of the model which might be a factor for deployment on embedded devices in real-world applications.

Conclusion

In this paper, three commonly used NILM architectures for individual appliance identification are compared. From this comparison, we can conclude that the combination of Convolution and LSTM outperforms stand-alone convolutional neural networks and LSTM neural networks. The Conv-LSTM has an average accuracy of 76%. We believe this performance can be further improved using more advanced techniques. We suggest using time encodings to exploit human habits. For example each day around 8 AM, the coffee machine would be turned on. Another improvement can be the use of transformers. This new type of neural network has its origin in natural language processing where it has shown superior performance when compared to other state-of-the-art algorithms. Transformers have potential in sequence classification due to the self-attention mechanism. This mechanism can highlight relations within the event sequences to further distinguish closely related classes.

Furthermore, it can be concluded that higher sampling rates benefit the accuracy. Our approach achieves an acceptable accuracy on 15Hz up to 60Hz but fails on frequencies below 1Hz and thus the current P1-gate. These higher sampling rates facilitate to distinguish similar devices with similar behaviour and thus to generate more detailed electrical load profiles. When the S1-gate (2-4kHz) will be deployed, another comparison should be made between these new sampling rates to define a trade-off between accuracy and efficiency.

Another key aspect of NILM research is the quality of provided data. The dataset should contain sufficient events for each class to avoid miss classification due to overfitting on limited data.

Because of the high accuracy of the proposed NILM algorithm, it is found that Conv-LSTM can contribute to the development of SLPs by central electrical measurements. In this way, it will be more cost and labour efficient to generate electrical load profiles, because only one measurement is needed. Moreover, with the roll-out of smart meters, these meters can provide the data for the Conv-LSTM algorithm for disaggregation into synthetic load profiles (SLPs). These SLP are highly needed to perform realistic building simulations, because the heat demand in buildings is shifting from a heat-based demand to a heating and cooling demand. The higher insulation rates of buildings lead to less space heat demand, which means that the internal heat gains are becoming more important to take into account. The main sources of internal heat gains are the occupancy and the heat losses of electrical appliances. These losses can be deviated from the electrical load profiles and an efficiency of the different electrical appliances. For these reasons, it is important to include the SLPs in building simulations, in order to perform accurate energetic evaluations.

References

Anderson, K. & Ocneanu, A. & Benitez, D. & Carlson, D. & Rowe, A. and Berges, M. (2012). Blued: a fully labeled public dataset for event-based non-intrusive load monitoring research. *Proceedings from the 2nd KDD Workshop on Data Mining Applications in Sustainability*. Beijing (China), 12 August 2012

Barsim, K. S., Streubel, R., & Yang, B. (2014). An approach for unsupervised non-intrusive load monitoring of residential appliances. *Proceedings from the 2nd International Workshop on Non-Intrusive Load Monitoring*. Austin (USA), 3 June 2014.

Basu, K., Hably, A., Debusschere, V., Bacha, S., Driven, G. J., & Ovalle, A. (2016). A comparative study of low sampling non intrusive load dis-aggregation. In *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*. *IEEE*, 5137-5142.

Bonfigli, R., Felicetti, A., Principi, E., Fagiani, M., Squartini, S., & Piazza, F. (2018). Denoising autoencoders for non-intrusive load monitoring: improvements and comparative evaluation. *Energy and Buildings 158*, 1461-1474.

Bonfigli, R., Principi, E., Fagiani, M., Severini, M., Squartini, S., & Piazza, F. (2017). Non-intrusive load monitoring by using active and reactive power in additive Factorial Hidden Markov Models. *Applied Energy 208*, 1590-1607.

Burman, E. & Mumovic, D. & Kimpian, J. (2014). Towards measurement and verification of energy performance under the framework of the European directive for energy performance of buildings. *Energy 77*, 153-163.

Chen, K., Wang, Q., He, Z., Chen, K., Hu, J., & He, J. (2018). Convolutional sequence to sequence non-intrusive load monitoring. The Journal of Engineering, 2018 (17), 1860-1864.

De Baets, L., Ruyssinck, J., Develder, C., Dhaene, T., & Deschrijver, D. (2018). Appliance classification using VI trajectories and convolutional neural networks. *Energy and Buildings 158*, 32-36.

European Commission (2014). *Benchmarking smart metering deployment in the EU-27 with a focus on electricity*.

European Commission (2011). *Joint Research Centre, and European Technology Platform on Renewable Heating and Cooling (RHC-Platform), 2020–2030–2050, common vision for the renewable heating and cooling sector in Europe: European tech- nology platform on renewable heating and cooling.*

European commission (2008). *Energy efficiency: delvivering the 20% target.*

Figueiredo, M., De Almeida, A., & Ribeiro, B. (2012). Home electrical signal disaggregation for non-intrusive load monitoring (NILM) systems. *Neurocomputing 96*, 66-73.

Figueiredo, M., Ribeiro, B., & de Almeida, A. (2014). On the optimization of appliance loads inferred by probabilistic models. *Proceedings from the 2nd International Workshop on Non-Intrusive Load Monitoring*. Austin (USA), 3 June 2014.

Gelazanskas, L. & Gemage A.A., K. (2014). Demand side management in smart grid: A review and proposals for future direction. *Sustainable Cities and Society 11*, 22-30.

Gong, F., Han, N., Zhou, Y., Chen, S., Li, D., & Tian, S. (2019). A SVM Optimized by Particle Swarm Optimization Approach to Load Disaggregation in Non-Intrusive Load Monitoring in Smart Homes. Proceedings from the 3rd IEEE Conference on Energy Internet and Energy System Integration (EI2). Changsha (China), 8-10 November 2019.

Harell, A., Makonin, S., & Bajić, I. V. (2019). Wavenilm: A causal neural network for power disaggregation from the complex power signal. Proceedings from IEEE on 44th International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brighton (UK), 12-19 May 2019.

Hart, G. W. (1992). Nonintrusive appliance load monitoring. *IEEE 80* (12), 1870-1891.

Huchtkoetter, J., & Reinhardt, A. (2019). A study on the impact of data sampling rates on load signature event detection. Energy Informatics 2, 1-12.

Kaselimi, M., & Doulamis, N. (2020). Long-term recurrent convolutional networks for non-intrusive load monitoring. *Proceedings from the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. Corfu (Greece), June 2020.

Kaselimi, M., Protopapadakis, E., Voulodimos, A., Doulamis, N., & Doulamis, A. (2019). Multi-channel recurrent convolutional neural networks for energy disaggregation. *IEEE Access 7*, 81047-81056.

Kaselimi, M., Doulamis, N., Voulodimos, A., Doulamis, A. D., & Protopapadakis, E. (2020). EnerGAN++: A Generative Adversarial Gated Recurrent Network for Robust Energy Disaggregation. *IEEE Open Journal of Signal Processing 2*. 1-16.

Khan, M. M. R., Siddique, M. A. B., & Sakib, S. (2019). Non-Intrusive Electrical Appliances Monitoring and Classification using K-Nearest Neighbors. In *2019 2nd International Conference on Innovation in Engineering and Technology (ICIET)*. Hanoi (Vietnam), 6 May 2019.

Kelly, J., & Knottenbelt, W. (2015). Neural nilm: Deep neural networks applied to energy disaggregation. *Proceedings from the 2nd ACM international conference on embedded systems for energy-efficient built environments*. Seoul (South Korea), November 2015.

Lin, J., Ding, X., Qu, D., & Li, H. (2020). Non-intrusive load monitoring and decomposition method based on decision tree. *Journal of Mathematics in Industry 10* (1), 1-14.

Liu, H. (2020). Hidden Markov Models Based Appliance. *Non-intrusive Load Monitoring*. Springer, Singapore.

Lynch, S., & Longo, L. (2017). On the Relationship between Sampling Rate and Hidden Markov Models Accuracy in Non-Intrusive Load Monitoring. In *AICS,* 180-192.

Neirotti, F. & Noussan, M. & Simonetti, M. (2020). Towards the electrification of buildings heating – Real heat pumps electricity mixes based on high resolution operational profiles. *Energy 195,* Article 116974.

Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training recurrent neural networks. *Proceedings from the International conference on machine learning*. Atlanta (USA), 16-21 June 2013.

Rijksdienst voor Ondernemend Nederland (2018). *Study in order of the Ministery of Economic Affairs and Climate, "W/E report: Ontwikkeling van de koudevraag van woningen" (in Dutch).*

Thomaßen, G. & Kavvadias, K. & Pablo Jiménez Navarro, J. (2021). The decarbonisation of th eEU heating sector through electrification : A parametric analysis. *Energy Policy 148,* Part A article 111929.

Wang, T. S., Ji, T. Y., & Li, M. S. (2019). A new approach for supervised power disaggregation by using a denoising autoencoder and recurrent lstm network. *Proceedings from the IEEE 12th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)*. Toulouse (France), 27-30 Augustus 2019.

Xia, M., Wang, K., Song, W., Chen, C., & Li, Y. (2020). Non-intrusive load disaggregation based on composite deep long short-term memory network. *Expert Systems with Applications 160*, 113669.

Xu, Y., & Milanović, J. V. (2014). Artificial-intelligence-based methodology for load disaggregation at bulk supply point. *IEEE Transactions on Power Systems 30*(2), 795-803.

Yang, C. C., Soh, C. S., & Yap, V. V. (2018). A systematic approach in appliance disaggregation using k-nearest neighbours and naive Bayes classifiers for energy efficiency. *Energy Efficiency 11*(1), 239-259.

Zhou, G., Li, Z., Fu, M., Feng, Y., Wang, X., & Huang, C. (2020). Sequence-to-Sequence Load Disaggregation Using Multi-Scale Residual Neural Network. *IEEE Transactions on Instrumentation and Measurement 70*, 1-10.