An Overview of Techniques for Non-Intrusive Load Monitoring

Milan Đurović, Marko Batić, Nikola Tomašević
University of Belgrade, Institute Mihajlo Pupin, Belgrade, Serbia
milan.djurovic@pupin.rs, marko.batic@pupin.rs, nikola.tomasevic@pupin.rs

***Abstract* — Although the financial benefits are crucial for today's market oriented industry, environment protection is recognized as an ultimate goal for modern industry development. This goal is usually performed in the frame of energy efficiency. It has been shown that non-intrusive load monitoring (NILM) is not only beneficial for planning and optimization of production but also for consumption of energy. Overview of basic problems related to NILM, key idea, theoretical frames for realization of this task and an implemented algorithm are presented in this paper.**

I. Introduction

When it comes to Non-Intrusive Load Monitoring (NILM) it is all about determining consumption of specific devices from the signal of aggregated power. This can be really challenging task due to existence of noise, similarity of devices and variety of appliances of the same type.

Surveys have shown that providing information to consumers about their consuming habits can yield up to 12% energy savings [1], [2], this is really important fact for environment conservation. Nowadays, when computational power and smart-meters are affordable, interest in non-intrusive load monitoring has significantly increased due to potential economic benefits of consumers.

In this paper we have descried concept of NILM. We have also described what are the problems usually encountered when performing NILM. An overview of state of the art of relevant algorithms from the literature is also provided. We have proposed a suitable approach based on Artificial Neural Networks (ANN). Moreover, performance evaluation of the proposed ANN based approach was conducted in this article as well.

The reminder of this paper is organized as follows. Section 2 describes state of the art literature and algorithms that we could find in the literature. In Section 3 we have proposed and described our algorithm. In Section 4 we have analyzed results scored with our approach. Finally, concluding remarks and discussion are given in Section 5.

II. State Of The Art

There are many approaches to the task of NILM. Signature of appliance is some characteristic set of patterns or features in power consumption of that appliance. Some researchers are primarily interested in transient states for example harmonics that occur for short periods of time when an appliance changes its state. In transients there is often a lot of information that is highly unique for specific appliance and then we can say that they use transient signatures. Norford and Leeb [3] use transient event detection to classify devices with similar level of consumption. Whereas others are more interested in steady state signatures. When it comes to steady state signatures the ones that are mostly used are power change, harmonics and V-I trajectories. Power change is actually a level of average power consumption when device is in specific state. Harmonics refer to patterns in frequency domain that are stable while the device is in specific state. V-I trajectories refer to voltage-current characteristics that are also stable when the device is in specific state. Some examples of V-I trajectories are shown in Fig. 1. It is important to highlight that sampling frequency one is dealing with is going to influence the choice of approach one is using. For instance, it is impossible to develop approach based on transient states if sampling period is one second or higher considering these transient states happen in split of second. It is also important to realize that decision either to use transient state signatures or steady state signatures beside sampling frequency is going to be influenced by type of device one should recognize.

In order to understand some of the difficulties one is going to deal with when performing task of NILM it is important to know that there are three basic types of appliances (See Fig. 2).

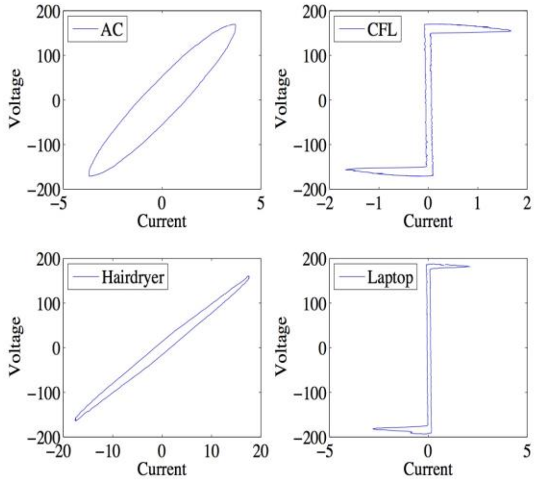

Figure 1. Examples of V-I trajectories

First type is so called single state appliances, these are devices that can be either turned on or off and in both of these states they have some constant level of power consumption that doesn't vary over time. Second type are multi state type of appliances that can be in more than two states, but every of these states are characterized by constant level of power consumption that is time-independent. Thirdly there are also continuous or infinite state type of appliances that can have two or more states but in these states appliances don't have the constant level of power consumption. Third group of appliances is the most difficult one for energy disaggregation.

There are many techniques that were presented but it seems that it doesn't exist a superior method that would give the best results in all situations.

A. Hart's approach

One of the first methods that was proposed back in eighties was Hart's approach [4], which consists of several steps. First step is to measure voltage and current, after that one should calculate normalized power in order to mitigate variations and instability of voltage of grid. Next step is to perform edge detection. That means that one should detect events when some appliance changes its state based on change in aggregated power. Next step is to perform clustering of appliances, and build a model for every appliance. After that one can use the gathered data to make a statistic for each appliance. Important disadvantages of this approach is that in this way you need labeled data for a lot of devices in order to perform training of algorithm mainly due to huge variety of appliances of the same type. This approach is also not able to learn patterns in power consumption of specific appliances and that can be really important especially when it comes to continuous state appliances. There is also no implementation of knowledge about correlation between states of different appliances. Best example how this correlation can be important is if one has Xbox and TV-set, so for example if Xbox is turned on probability that your TV is turned on is really high and this can be really beneficial for recognizing some devices. In this approach it is also not modeled a duration of states and duration of state can be really beneficial. For example if you know that your washing machine is turned on now and you know that average time of on state is 45 minutes this information can be really effective for recognizing the moment of turning off the device.

There are more than few approaches that are based on similar principle. One of modern approaches that involve feedback through mobile application from residents is [5]. In this article authors use measurements of real, reactive and distortion power in order to classify edges. In the trial they performed the accent was on small devices and in order to mitigate the noise they experimented with different filters for preprocessing the signal. They got the best results with kernel filters but mainly due to computational complexity they decided to use combination of mean and median filter instead.

Since then there were many proposed approaches that differ significantly, but they all have something in common and that is that they have advantages and disadvantages in comparison with other methods.

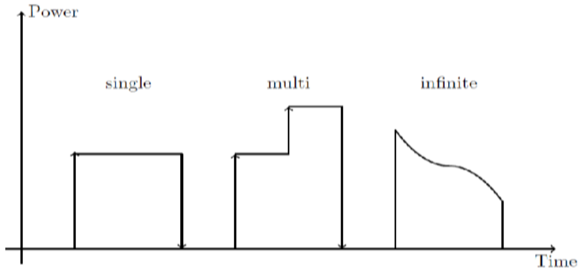

Figure 2. Graph of power consumption of three different type of appliances.

B. Hidden Markov Models

It is also interesting that most of these methods were not originally developed for task of NILM but rather for some other applications. Best example are approaches based on Hidden Markov Models (HMMs) [6], [7], [8]. Application of HMMs in speech recognition tasks was successful so many researchers led by the similarity of these two tasks tried to apply HMMs and variations of this power tool to NILM and they obtain satisfying results that are state of the art results.

Hidden Markov Models (HMMs) are probabilistic tool that has significant application at many areas. We are going to explain how it works on a simple example of a single device that can be either turned on or turned off. Lets presume that our task is to determine the states of the appliance s_t $t\in \{1,2,\ldots,T\}$ based on our series of measurements of power consumption y_t $t\in \{1,2,\ldots,T\}$. Our HMM is going to consist of the finite set of hidden states *S* (in our case it can be ON or OFF) of an appliance, the finite set of measurements *Y* per states (power consumption) observed in each state, $Y = \{y_1, y_2\ldots., y_T\}$. The observable symbol *Y* can be discrete or a continuous set. The transition matrix $A = \{a_{ij}, 1 \leq i, j \geq N\}$ represents the probability of moving from state S_i to S_j such that: $a_{ij} = P(q_{t+1} = S_j | q_t = S_i)$, with $a_{ij} \leq 0$ and where q_t denotes the state occupied by the system at time *t*. The emission matrix $B = P(y_t \mid S_j)$ representing the probability of measurement *Y* when system state is S_j. So basically there are three possible tasks when it comes to HMMs.

The first task is so called the evaluation problem and it refers to determining the probability that given sequence of measurements was created by specific model. The second task is so called decoding problem and it addresses to determining the most probable series of states that generated given measurements for given model. Third task is so called learning problem and it is all about learning the parameters of the model given series of measurements. When it comes to task of NILM at beginning one is going to deal with learning problem in order to determine parameters of the model and after that one is going to solve second task in order to determine states of appliances based on measurements.

The first idea could be to model every possible state of all appliances at specific house as a state of hidden Markov model. It is not hard to see that if you have *N* devices at your home, where every appliance has *M* states your HMM would consist of M^N states. That would be computationally too expensive. That's why Factorial Hidden Markov Models (FHMM) were proposed.

The basic idea with FHMMs is that one is going to have as many HMMs as there are appliances at home. So now if there are *N* appliances and each appliance has *M* states FHMM is going to have *MN* states which is significantly less than previously. Model of each appliance is going to generate consumption of that device and then all these consumptions are going to be aggregated and that's the reason why they are sometimes called Additive FHMMs. In spite of having computationally less expensive problem it's still challenging to perform disaggregation mainly due to local optima susceptibility. Authors in [6] presented approximate inference algorithm that overcome problem with local optima susceptibility. There is also a problem due to way of modeling duration of states of appliances. In order to illustrate this let's presume that appliance of interest is washing machine and that it is usually turned on for an hour a day. HMMs model time duration of each appliance through matrix of transitions and they model probability density functions of duration of states with geometric distributions which would be inadequate in the example of washing machine. That's why Hidden Semi Markov Models were developed. The key feature of HSMM's is that this tool is capable of modeling time duration of states in more appropriate way. Also there is a problem with approaches based on HMMs that they don't model dependencies between different appliances. If you look at the example of your monitor and your PC it is not hard to see that there is significant correlation between states of these two devices. Kim [8], presented approach based on Conditional Factorial Hidden Semi Markov Models (CFHSMM) that overcomes some of these problems by introducing dependencies between parameters of the model and other features and also he has outperformed approaches based on FHMM and FHSMM. He didn't introduce new features but only utilized states of other devices and also the time of day in determining the state of appliance of interest. He has shown that at daytime duration of ON-states of most of appliances is more appropriate to model with gamma distribution whereas at night it is better to represent it with geometric distribution. We have to outline that despite of promising results of approaches based on HMMs estimating of the parameters without having information on disaggregated power consumption of the house where task of NILM has to be performed is quite challenging and much harder in comparison with some other approaches that are going to be presented later.

C. *Graph Signal Processing*

Another interesting approach is Graph Signal Processing (GSP) [9]. In the case when there are not much data GSP approach seems to be suitable. It is a novel signal processing concept because it captures correlation among data samples in time and space by embedding the structure of signals onto a graph. It shows powerful, scalable and flexible properties that makes it suitable for many data mining and signal processing problems.

D. *Artifical Neural Networks (ANN)*

There is also application of deep learning for task of NILM [10], [11]. Artificial neural networks or just neural networks have scored amazing results in visual recognition challenges and they totally outperformed all methods whose features were human based. From 2012 when Alex Krizhevsky and his team have presented Alex-net [12], deep convolutional neural network, that has won ImageNet Large Scale Visual Recognition Competition (ILSVRC) researchers from all over the world started to implement different architectures of deep neural networks in so many domains and that's the case of NILM too. As mentioned above, one of the biggest advantages when using neural networks is that one doesn't have to think of extracting meaningful features because that's going to be done by network. There are three network architectures that are proposed by Kelly [10]: convolutional neural networks, Long Short Term Memory (LSTM) networks and autoencoder networks. When it comes to convolutional neural networks the basic idea is to have a lot of small filters that are going to slide through input and perform convolution. So every of these filters are going to have the task to find something meaningful in input and then at the end we are going to have a few dense layers that are going to make decisions based on all the information provided by these filters. Mainly due to huge number of parameters approaches based on deep neural network require a lot of labeled data, and this can be obstacle in implementation of this approach. It is also important to highlight that training of deep neural networks is computationally and time consuming. But on the other hand it is really important to outline that once trained network can perform classification or regression in a split of a second. Crucial advantage of application of neural networks for task of NILM is that they score satisfying results on unseen houses where they don't require labeled data but only signal of aggregated power consumption.

Kelly [10] scored best results with LSTM networks which are quite often used for speech recognition, grammar checking and translation. Pedro [11] also used this network architecture and it was also really interesting that when he wanted to perform regression he reduced this problem on classification by dividing interval of possible levels of consumption of devices on smaller intervals and performed quantization. He used softmax loss, which outperforms L2 loss especially when there are outliers that can significantly deteriorate results when one uses L2 norm.

Kelly [10] also implemented autoencoder network. He was led by extremely good results that were scored in image processing domain and the main advantage of deep neural networks which is that they are capable of extracting meaningful features that are in most cases more useful than human based extracted features. The main idea of autoencoder networks is to have architecture shown in Fig. 3. Every autoencoder network consists of encoder and decoder part. Encoder network is supposed to extract useful features whose dimension is going to be smaller than dimension of input, and decoder network is supposed to reconstruct input provided only extracted features.

Very interesting approach was proposed in [13], the basic idea was to train deep neural network to extract meaningful features and then to use these features as observations for hidden Markov Models.

Some authors performed frequency analysis of aggregated signal but this approach demands high sampling frequency.

E. ElectriSense

Another approach that requests really high sampling frequency is ElectriSense [14]. ElectriSense relies on the fact that most modern consumer and fluorescent lighting employ switch mode power supplies (SMPS) to achieve high efficiency. These power supplies continuously generate high frequency electromagnetic interference (EMI) during operation that propagates throughout a home's power wiring. The authors of this approach has shown that EMI signals are stable and predictable based on device's switching frequency characteristics.

III. ANN BASED APPROACH IMPLEMENTATION

We implemented a similar solution to one presented in [15]. The architecture we used is shown in Fig. 4. But unlike the authors from [14] we used this architecture for purpose of classification and our architecture also has significantly less units in dense layer so there are also less chances for overfitting which happens when there are too many parameters of the network in comparison with available data for training. In these situations it is a common thing to have really small error on training set but high one on test set. Risk of overfitting is pretty high when one uses deep neural networks with huge number of parameters. In order to fight against it we used dropout [16]. The main idea of dropout is to zero out outputs of half a neurons in a layer in every forward pass, which means essentially to exclude half a neurons of a layer. In this way it forces neural network to perform diversification and not to make decision based only on few nodes. We experimented with number of nodes per layer and chose our architecture empirically. We also used batch normalization [17] in order to normalize inputs of layers, in this way problems with initialization are significantly mitigated and it has beneficial influence on generalization.

We also used early stopping where we divided our date in three parts training, validation and test part. We also trained multiple model ensembles and average their results in test time in order to fight against overfitting.

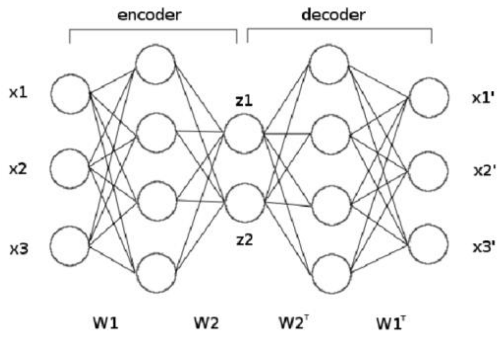

Figure 3. Architecture of autoencoder network.

Term classification refers to fact that our task was to determine if the specific appliance is turned on or turned off at some time. We used REDD data set [18], for both training and testing. In REDD data set there is information about both aggregated and disaggregated power consumption of 6 houses and appliances in these houses. Sampling period for aggregated power consumption was one second and sampling period for power consumption of appliances was three seconds. In order to get labeled data from consumption of appliances we preprocessed data with mean filter and used threshold of 10W. The level of threshold was determined by analyzing the consumption of devices we were interested in.

It is also important to highlight that in REDD dataset there are often periods when either sensors for aggregated power or sensors for power consumption of single appliances were not working. While preprocessing we decided not to use such periods.

We trained two neural networks for two appliances, first one for refrigerator and the second one for dishwasher. The input of our network is sequence of 599 measurements of aggregated power consumption and you can see example of such an input in Fig. 5

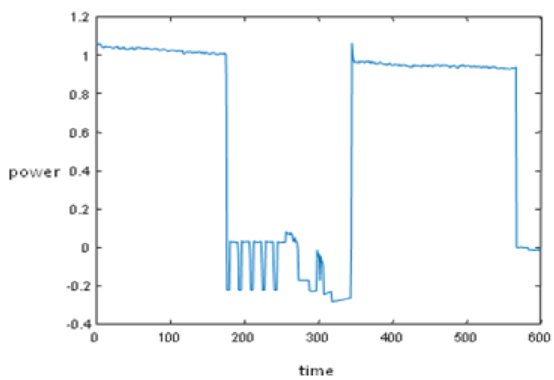

Figure 5. Example of input in neural network

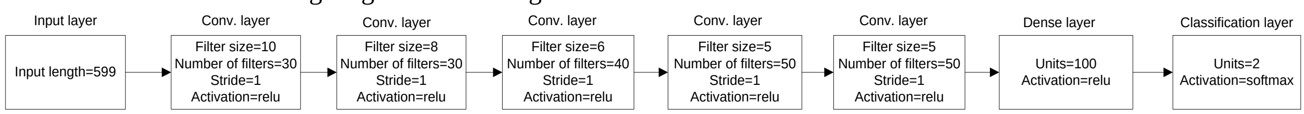

Figure 4. Architecture of neural network utilized for NILM.

Results for refrigerator are presented in Table 1. Results for unbiased data for dishwasher are presented in Table 2. Results for biased data for dishwasher are presented in Table 3. Term biased refers to situation when there are much more examples from one class than from another one.

In Tables 1, 2 and 3 there are four metrics. Accuracy is defined as a percent of correctly classified samples among all samples. Precision is defined as correctly classified samples among samples that are classified as that device of interest was turned on. Recall is defined as percent of correctly classified samples among samples that come from the class when device of interest was turned on. F1-score is harmonic mean of precision and recall.

It is important to highlight that training was performed on the data from four houses and testing was executed on the data from the house different from the ones the model was trained on. In this way we simulated realistic situation that one has to perform NILM at the house where there is no information about specific appliance one has to recognize and also there is no information how many appliances there are.

IV. Performance Evaluation

When it comes to refrigerator the results are satisfying and you can see that both accuracy and F1-score are extremely high. And it is obvious that network succeed to perform generalization based on high accuracy and F1-score in situation where it didn't have any information about the specific model of fridge it was supposed to recognize.

In Fig. 6 you can see how power consumption of refrigerator looks like and you can see that there exists obvious pattern. This is great asset of deep neural networks to learn really complex patterns and dependencies from data. In many areas it has been shown that deep neural networks totally outperform approaches and models that are handmade especially when it comes to complex problems that are hard to describe mathematically. In situations like these if one has a lot of data for training deep neural networks seem like a reasonable choice. We believe that this is the case with NILM too, and experiments we performed confirmed that.

On the other hand when it comes to dishwasher we can see that results are good but not as good as for refrigerator. Main reason for these results is that we had significantly less data about the dishwasher primarily due to the fact that refrigerator was much more often turned on. In our training data we had roughly the same number of examples where device is turned on and where it's turned off. The reason why we created our training data in this way is that if had much more examples from one class it would be very easy for model to converge to local minima where it would all time classify examples as class that's predominant.

This is why the size of training data for dishwasher is much smaller. Data for fridge consisted of measurements that effectively was gathered for 30 days whereas data for dishwasher was effectively gathered for 4 days. You can find results for dishwasher in Table 2 and Table 3. In Table 2 are presented results for unbiased data, and unbiased means that again we have roughly the same number of examples from both classes, when device is turned on and when it is turned off. In Table 3 you can find the results for whole data that was available from the house where the model was tested.

You can see now why F1-score is much better than accuracy. Accuracy was much more affected by data it was tested on, and F1-score is more representative and more credible metric.

V. Conclusions

We presented in this article basic concept of Non-Intrusive Load Monitoring (NILM) and described the problem. Global tendencies to decrease energy consumption and pollution are going to become even more important in the future. NILM is most definitely going to be the part of it because it is a relatively easy way to stimulate people to change their habits and decrease the consumption. In this article the overview of the state of the art of the relevant approaches that could be find in the literature is presented. It was obvious that this problem is really interesting and attractive for many researchers as there is variety of proposed algorithms. We developed our own architecture of neural network for task of classification which was inspired with [15] but had

Table 1.
Scores for our network architecture for refrigerator.

Accuracy	F1-score	Precision	Recall
0.92	0.92	0.98	0.89

Table 2.
Scores for our network architecture for unbiased data for dishwasher.

Accuracy	F1-score	Precision	Recall
0.92	0.92	0.98	0.89

Table 3.
Scores for our network architecture for biased data for dishwasher.

Accuracy	F1-score	Precision	Recall
0.9935	0.74	0.77	0.70

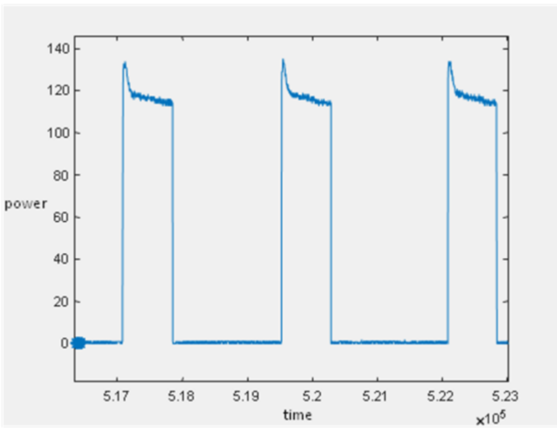

Figure 6. Power consumption of refrigerator.

significantly smaller number of parameters. We tested our model for two appliances and scored state of the art results. In the future we plan to train and test our models for bigger number of appliances and for different sampling periods. We are also planning to evaluate impact of variety among appliances of the same type to behavior of our model and to other proposed algorithms.

REFERENCES

[1] K. C. Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? The case of electricity," Energy Policy, vol. 52, no. 0, pp. 213 – 234, 2013.

[2] A. Faustine, N. H. Mvungi, S. Kaijage, K. Michael, A Survey on Non-Instrusive Load Monitoring Methoides and Techniques for Energy Disaggregation Problem, Cornell University Library, arXiv.org: 1703.00785.

[3] Norford, L.K. and Leeb, S.B. Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load detection algorithms. Energy and Buildings, 24(1):51-64, 1996

[4] G. W. Hart, "Nonintrusive appliance load monitoring," Proceedings of the IEEE, vol. 80, no. 12, pp. 1870–1891, 1992.

[5] M. Weiss, A. Helfenstein, F. Mattern, and T. Staake, "Leveraging smart meter data to recognize home appliances," in Proceedings of the 10th IEEE International Conference on Pervasive Computing and Communications (PerCom), 2012, pp. 190-197

[6] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial hmms with application to energy disaggregation," in NILM: A Review and Outlook Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics (AISTATS-12), N. D. Lawrence and M. A. Girolami, Eds., vol. 22, 2012, pp. 1472–1482. [Online]. Available: http://jmlr.csail.mit.edu/proceedings/papers/v22/zico12/zico12.pdf

[7] Johnson, M.J.; Willsky, A.S. Bayesian Nonparametric Hidden Semi-Markov Models. Available online: http://arxiv.org/abs/1203.1365 (accessed on 4 December 2012).

[8] H. Kim, M. Marwah, M. F. Arlitt, G. Lyon, and J. Han. Unsupervised Disaggregation of Low Frequency Power Measurements. Proceedings of the 11th SIAM International Conference on Data Mining, pages 747-758, 2011.

[9] Bochao Zhao, Lina Stankovic, and Senior Member. On a Training-Less Solution for Non-Intrusive Appliance Load Monitoring Using Graph Signal Processing. IEEE Transactions on Smart Grid, 4, 2016.

[10] KELLY, J., KNOTTENBELT, W. J. "Neural NILM: Deep Neural Networks Applied to Energy Disaggregation", CoRR, v. abs/1507.06594, 2015. Disponívelem: <http://arxiv.org/abs/1507.06594>.

[11] Pedro Paulo. Applications of Deep Learning techniques on NILM. PhD thesis, Universidade Federal do Rio de Janeiro, 2016.

[12] Krizhevsky, A., Sutskever, I., and Hinton, G. E. ImageNet classification with deep convolutional neural networks.
In NIPS, pp. 1106–1114, 2012.8

[13] Anders Huss. Hybrid Model Approach to Appliance Load Disaggregation. PhD thesis, KTH Royal Institute of Technology, 2015.

[14] GUPTA, S., REYNOLDS, M. S., PATEL, S. N. "ElectriSense: Single-Point Sensing Using EMI for Electrical Event Detection and Classification in the Home". In: In Proceedings of the 12th ACM International Conference on Ubiquitous Computing, pp. 139–148, 2010.

[15] Chaoyun Zhang, MingjunZhong, Zongzuo Wang, Nigel Goddard, Charles Sutton. Sequence-to-point learning with neural networks for nonintrusive load monitoring. arXiv:1612.09106

[16] N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1):1929–1958, 2014.

[17] Sergey Ioffe , Christian Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, Proceedings of the 32nd International Conference on International Conference on Machine Learning, July 06-11, 2015, Lille, France.

[18] J. Zico Kolter and Matthew J. Johnson. REDD: A public data set for energy disaggregation research. In proceedings of the SustKDD workshop on Data Mining Applications in Sustainability, 2011.