

Deep Learning for Analysis of Time-Series in Smart Home Environments



THE UNIVERSITY OF
NEWCASTLE
AUSTRALIA

Mohammad Mobasher Kashani

BS(KHU); M.IT(UKM)

Interdisciplinary Machine Learning Research Group
School of Information and Physical Sciences
College of Engineering, Science and Environment
University of Newcastle

**A thesis submitted in fulfilment of requirements
for the degree of Doctor of Philosophy in Computer Science**

September 2022

I would like to dedicate this thesis to my loving parents, my beautiful sisters, and the memory of my brother who passed away.

Acknowledgements

Mohammad Mobasher-Kashani was supported by UNIPRS and UNRSC scholarships awarded by The University of Newcastle, Australia.

Declaration

I hereby certify that the work embodied in the thesis is my own work, conducted under normal supervision. The thesis contains no material which has been accepted, or is being examined, for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made. I give consent to the final version of my thesis being made available worldwide when deposited in the University's Digital Repository, subject to the provisions of the Copyright Act 1968 and any approved embargo.

Mohammad Mobasher Kashani
BS(KHU); M.IT(UKM)
September 2022

Acknowledgement of Authorship

I hereby certify that the work embodied in this thesis contains material of the following two published papers and manuscript under review of which I am a joint author:

- Submitted Paper : M. Mobasher-Kashani, N. Noman and S. Chalup, "Relating Dataset Complexity and Deep Learning Models in the Context of the Non-Intrusive Load Monitoring"
- M. Mobasher-Kashani, J. Li and S. Luo, "Light-Weight Recurrent Deep Learning Algorithm for Non-Intrusive Load Monitoring," 2019 IEEE 2nd International Conference on Electronic Information and Communication Technology (ICEICT), 2019, pp. 572-575, doi: 10.1109/ICEICT.2019.8846263.
- M. Mobasher-Kashani, N. Noman and S. Chalup, "Parallel LSTM Architectures for Non-Intrusive Load Monitoring in Smart Homes," 2020 IEEE Symposium Series on Computational Intelligence (SSCI), 2020, pp. 1272-1279, doi: 10.1109/SSCI47803.2020.9308592.

I have contributed to these papers by contributing to each study's conception and design, developing analysis plans, developing research material and collecting data, performing both quantitative and qualitative analyses, writing code and executing programs, interpreting data, and leading the writing of the manuscripts.

Mohammad Mobasher Kashani
BS(KHU); M.IT(UKM)
September 2022

Principal supervisor endorsement:
Professor Stephan Chalup

Abstract

Electrical energy wastage is a major problem in residential areas in modern societies. Energy production costs our environment a great deal and wasting such valuable assets will bring enormous consequences for future generations. Hence, it is a vital priority for researchers in all areas to tackle this problem. Computer science, alongside other disciplines, provides approaches to abate the overconsumption of electrical energy using a modern computation algorithm called deep learning. Deep learning models are powerful tools for the computation of complex problems, namely energy disaggregation problems. Non-Intrusive Load Monitoring (NILM) is becoming popular as an approach for extracting detailed power consumption information related to different appliances used in a household. NILM utilizes time-series analysis methods to disaggregate the signals of operating appliances from a single point in houses and, based on that, can provide advice to households about how to reduce their energy consumption. NILM is known to be a challenging task from a computational perspective, and previous solutions have not demonstrated sufficient accuracy. This thesis aims to achieve solutions with higher accuracy by utilizing deep neural network models with a Parallel Long Short Term Memory Topology (PLT) and Deep Transformer Networks. To implement machine learning techniques on low-cost devices with limited computation power, the Deep Transformer Networks are enhanced to adapt to these criteria, and then, in order to tackle the imbalance in appliances in the current datasets, a synthetic dataset is proposed and the models are applied to the synthetic dataset with the transfer learning method. The proposed method to create a synthetic dataset improves the accuracy and reduces the training time for the deep learning models in this study. To create the proposed synthetic data generator, the common time-series features of different appliances have been modelled with mathematical formulas. Since maintaining the privacy of users is vital in our design, the solution is optimized for implementation on edge devices, and all computation related to residents' data will be performed locally. The thesis describes the main structure of the deep learning models that are introduced and the proposed synthetic data generation method, then evaluates them on publicly available datasets.

Table of contents

List of figures	x
List of tables	xiv
1 Introduction	1
1.1 ILM and NILM technologies	2
1.2 Deep Learning Methods on NILM	4
1.3 Synthetic Data Generation	4
1.4 Proposed Framework	5
1.5 Benchmark Datasets	5
1.6 Motivation	6
1.7 Objectives	7
1.8 Research Questions	7
1.9 Conclusion	8
2 Literature Review	9
2.1 Smart Energy Management Technologies	9
2.2 ILM and NILM Technologies	10
2.3 NILM Technology	10
2.3.1 Candidate Appliances	12
2.3.2 Data Sampling Interval	12
2.3.3 Synthetic Datasets	13
2.4 Disaggregation	14
2.4.1 Feature Extraction Approaches	15
2.4.2 Load Identification	16
2.4.3 Deep Learning Approaches	25
2.4.4 Transfer Learning Methods	29
2.5 Computing on Edge	29

2.6	Data Privacy in NILM	31
2.7	Overview of Software Packages	32
2.8	Summary	32
3	Parallel LSTM Architectures for Non-Intrusive Load Monitoring	33
3.1	Background	34
3.2	Methodology	35
3.2.1	Problem Formulation	36
3.2.2	Parallel Deep Learning Topology	36
3.2.3	Light-Weight Recurrent Deep Learning Algorithm	38
3.2.4	Parallel Long Short-Term Memory Topology (PLT)	40
3.3	Experiments and Results	43
3.3.1	Evaluation Metrics	43
3.3.2	Light-Weight Recurrent Deep Learning Training	45
3.3.3	Parallel Long Short-Term Memory Topology Training	47
3.4	Summary	50
4	Induction and Transduction Methods on Signal Disaggregation	52
4.1	Background	52
4.2	Methodology	55
4.2.1	Deep Neural Networks	56
4.2.2	Deep Transducer Methods	56
4.2.3	Attention Mechanism	57
4.2.4	Deep Transformer Networks	60
4.2.5	Proposed Transformer Models	61
4.2.6	Optimizing Transformers for the Edge Computing	65
4.3	Experiments and Results	68
4.3.1	Transducer Model Training	69
4.3.2	Transformer Models' Training	72
4.3.3	Time Feature Extraction for Transformer Networks' Training	72
4.3.4	Transformer Models for Edge Computing Training	74
4.4	Summary	77
5	Relating Dataset Complexity and Deep Learning Models in the Context of Non-Intrusive Load Monitoring	78
5.1	Background	79
5.1.1	Data Set Complexity	80

5.1.2	Objectives	80
5.2	Methodology	81
5.2.1	Meta-Dataset Design	82
5.2.2	Variable Importance Analysis	88
5.2.3	Modelling the Importance Relationship	89
5.3	Experiments and Results	90
5.3.1	Meta-Dataset Parameters	90
5.3.2	Importance Analysis Results	93
5.3.3	Gaussian Process Regression Results	100
5.4	Summary	106
6	Mathematical Modelling for Synthetic Data in Signal Disaggregation	108
6.1	Background	109
6.1.1	Synthetic Data Generation	109
6.1.2	Transfer Learning for Signal Disaggregation	110
6.2	Methodology	111
6.2.1	Mathematical Modelling of Appliances	111
6.2.2	Transfer Learning with MatSync Data	115
6.3	Experiments and Results	118
6.3.1	MatSync Data Generation	119
6.3.2	Training on Synthetic Data	120
6.3.3	Transfer Learning on Synthetic Data	122
6.4	Summary	142
7	Discussion and Conclusion	144
7.1	Main Contributions of the Thesis	145
7.2	Future Work	149
	References	151

List of figures

1.1	The Non-Intrusive Appliance Load Monitoring application is a technique to detect the devices inside the total aggregated power consumption of a house.	3
1.2	Formal procedure for NILM [48].	3
1.3	The main approach to transfer learning on the edge devices for NILM. . . .	5
2.1	The main structure of the literature review on NILM in this thesis.	10
2.2	The major categories of smart city technologies suggested by Lu et al. [119].	11
2.3	NILM Approach	14
2.4	The structure of Autoencoder Models	27
3.1	Topology of the proposed deep interpreter network framework for NILM. . .	37
3.2	Topology of the proposed parallel deep feature extraction framework for NILM.	37
3.3	Bidirectional LSTM architecture for NILM [96].	38
3.4	Light-weight LSTM model for non-intrusive load monitoring.	39
3.5	Parallel Long Short-Term Memory Topology (PLT) mode 1.	41
3.6	Parallel Long Short-Term Memory Topology (PLT) mode 2.	42
3.7	Microwave energy signal disaggregated from the REDD dataset using PLT models based on 1 second time interval.	47
3.8	Dishwasher energy signal disaggregated from the REDD dataset using PLT models based on 1 second time interval.	48
3.9	Fridge energy signal disaggregated from the REDD dataset using PLT models based on 1 second time interval.	48
3.10	Dryer energy signal disaggregated from the REDD dataset using PLT models based on 1 second time interval.	49
4.1	Transduction and Induction concepts [183].	54
4.2	The induction sequence for signal disaggregation.	54
4.3	The proposed transduction sequence for signal disaggregation.	55
4.4	NILM Framework [150].	56

4.5	The proposed deep transducer network for signal disaggregation.	58
4.6	The Proposed LSTM Transducer Model for NILM.	59
4.7	Scaled Dot-Product Attention Architecture.	60
4.8	Basic Multi-Head Attention Architecture.	61
4.9	Combination of Classification and Transformer on NILM.	62
4.10	The Proposed Transformer Model T1.	62
4.11	The Proposed Transformer Model T2 on the NILM.	63
4.12	The Proposed Transformer Model T3.	64
4.13	Reshaping the time-series input data to reduce the training time for the proposed deep learning models.	66
4.14	Comparison between the Sinus and Bessel functions to use as a time feature extraction layer.	67
4.15	Time Feature Extraction layer structure for transformer networks.	67
4.16	The architecture of the proposed transformer network with the time feature extraction layer.	68
4.17	Bidirectional Transducer on Fridge with the time interval of 1 second. . . .	70
4.18	Bidirectional Transducer on Microwave with the time interval of 1 second. .	70
5.1	Deep Learning model for Metadataset: Parallel LSTM Mode 2 [146]	85
5.2	Deep Learning model for the Meta-dataset: Parallel LSTM Mode 3	86
5.3	Comparison of four deep learning models on the REDD dataset.	92
5.4	The importance of dataset attributes in the outcome of deep learning models for the fridge in the REDD dataset.	94
5.5	The importance of dataset attributes in the outcome of deep learning models for the lights in the REDD dataset.	95
5.6	The importance of dataset attributes in the outcome of deep learning models for the microwave in the REDD dataset.	95
5.7	The importance of dataset attributes in the outcome of deep learning models for the computer in the REDD dataset.	96
5.8	The importance of dataset attributes in the outcome of deep learning models for the television in the REDD dataset.	97
5.9	The importance of dataset attributes in the outcome of deep learning models for the kettle in the REDD dataset.	97
5.10	The importance of dataset attributes in the outcome of deep learning models for the dishwasher in the REDD dataset.	98
5.11	The importance of dataset attributes in the outcome of deep learning models for the washer dryer in the REDD dataset.	99

5.12	The importance of dataset attributes in the outcome of deep learning models for the washing machine in the REDD dataset.	99
5.13	Validation of the important attributes of the fridge using GP.	101
5.14	Validation of the important attributes of the light using GP.	102
5.15	Validation of the important attributes of the microwave using GP.	102
5.16	Validation of the important attributes of the computer using GP.	103
5.17	Validation of the important attributes of the television using GP.	103
5.18	Validation of the important attributes of the kettle using GP.	104
5.19	Validation of the important attributes of the dishwasher using GP.	104
5.20	Validation of the important attributes of the washer dryer using GP.	105
5.21	Validation of the important attributes of the washing machine using GP. . .	105
6.1	The MatSync square shape.	112
6.2	The MatSync curve shape.	113
6.3	The MatSync point shape.	114
6.4	Fridge power consumption pattern extracted from house 1 of the REDD dataset.	114
6.5	The MatSync fridge shape.	116
6.6	Simulated dataset based on four basic shapes.	116
6.7	TL1 and TL2 Transfer Learning Scenarios on the Transformer Network. . .	117
6.8	The comparison of training time for the fridge in the deep learning models with transfer learning.	124
6.9	The comparison of training time for the microwave in the deep learning models with transfer learning.	125
6.10	The comparison of training time for the dishwasher in the deep learning models with transfer learning.	126
6.11	The comparison of training time for the washer dryer in the deep learning models with transfer learning.	127
6.12	The comparison of accuracy metrics for the fridge in the deep learning models with transfer learning.	128
6.13	The comparison of accuracy metrics for the microwave in the deep learning models with transfer learning.	129
6.14	The comparison of accuracy metrics for the dishwasher in the deep learning models with transfer learning.	130
6.15	The comparison of accuracy metrics for the washer dryer in the deep learning models with transfer learning.	131
6.16	The comparison of precision metrics for the fridge in the deep learning models with transfer learning.	132

6.17	The comparison of precision metrics for the microwave in the deep learning models with transfer learning.	133
6.18	The comparison of precision metrics for the dishwasher in the deep learning models with transfer learning.	133
6.19	The comparison of precision metrics for the washer dryer in the deep learning models with transfer learning.	134
6.20	The comparison of F1-Score on the two transfer learning schemes for the fridge.	135
6.21	The comparison of F1-Score for the two transfer learning schemes for the microwave.	136
6.22	The comparison of F1-Score on the two transfer learning schemes for the dishwasher.	136
6.23	The comparison of F1-Score on the two transfer learning schemes for the washer dryer.	137
6.24	Comparison of MAE for the Transfer Learning with Synthetic Data for fridge.	138
6.25	Comparison of MAE for Transfer Learning with Synthetic Data for microwave.	139
6.26	Comparison of MAE for Transfer Learning with Synthetic Data for dishwasher.	139
6.27	Comparison of MAE for Transfer Learning with Synthetic Data for washer dryer.	141

List of tables

1.1	Comparison of public datasets on the electric signal disaggregation problem.	6
2.1	Review of literature on machine learning algorithms in NILM	19
2.2	Review of Software Packages for Deep Learning Models	31
3.1	Comparison of light-weight deep learning model on the IAWWE dataset. . . .	46
3.2	Comparison of three introduced parallel models on REDD dataset.	49
4.1	Comparison of the results of LSTM and bidirectional transducer networks and dAE models on the REDD dataset.	71
4.2	Comparison of results for four introduced transformer models on the REDD dataset with classification. The main purpose of the comparison is to observe the model performance with CNN or PLT2 classifiers, hence, the related comparison between these two classifiers is highlighted with the grey colour.	73
4.3	Comparison of transducer and transformer networks on NILM.	75
4.4	Comparison of training time between transformer and transducer models regarding their computation hardware on the REDD dataset.	76
5.1	Comparison of dataset features in signal disaggregation.	81
5.2	Denoising Autoencoder architecture that has been used for signal disaggre- gation with 2,329 parameters [96].	83
5.3	LSTM Model Mode 2 architecture that has been used for signal disaggrega- tion with 538,385 parameters.	84
5.4	LSTM Model Mode 3 architecture that has been used for signal disaggrega- tion with 1,727,641 parameters.	87
5.5	Bidirectional LSTM architecture that has been used for signal disaggregation with 1,264,977 parameters [96].	87
5.6	Feature symbols	93
5.7	Summary of Feature Importance of Nine Appliances.	100

6.1	Results from the Transformer network on the Synthetic Data for four basic mathematical shapes.	121
6.2	Results from the PLT2 network on the Synthetic Data for four basic mathematical shapes.	121
6.3	Results from the PLT3 network on the Synthetic Data for four basic mathematical shapes.	122
6.4	Results from the dAE network on the Synthetic Data for four basic mathematical shapes.	122
6.5	Results from the Bidirectional LSTM network on the Synthetic Data for four mathematical shapes.	123
6.6	Comparison of results of transfer learning models on the REDD dataset. . .	140