

University of Technology Sydney

Faculty of Engineering and Information Technology

**Analysis of Residential Load Data and Its Applications for
Smart Grids**

A thesis submitted for the degree of

Doctor of Philosophy

Amin Rajabi

(2020)

Title of the thesis:

Analysis of Residential Load Data and Its Applications for Smart Grids

Ph.D. student:

Amin Rajabi

Email: @student.uts.edu.au

Supervisor:

A/Prof. Li Li

E-mail: Li.Li@uts.edu.au

Co-supervisor:

Dr. Jiangfeng Zhang

E-mail: Jiangfeng.Zhang@uts.edu.au

Address:

School of Electrical and Data Engineering,

University of Technology Sydney, 15 Broadway, Ultimo, NSW 2007, Australia

Certificate of Original Authorship

I, Amin Rajabi, declare that this thesis is submitted in fulfilment of the requirements for the award of PhD degree, in the school of Electrical and Data Engineering, faculty of Engineering and Information Technology at the University of Technology Sydney. This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program Scholarship.

Signature:

Production Note:
Signature removed prior to publication.

Amin Rajabi

Date: 30/11/2019

Acknowledgments

I would like to take this opportunity to thank the following people and organisations for their assistance and support during my candidature.

First, I would like to express my deep gratitude to my principal supervisor Assoc. Prof. Li Li and my co-supervisor Dr Jiangfeng Zhang for their guidance, encouragement, and belief in my abilities. Their constructive comments contributed significantly to the quality of this research. I also would like to thank Prof. Jianguo Zhu for his help and advice.

I wish to thank the Irish Social Science Data Archive (ISSDA) for providing me with the datasets which are used in this thesis. I thank the UTS Advanced Research Computing Laboratory and its manager Dr Matthew Gaston for giving me substantial technical supports. I also appreciate the help and constructive advice of Prof. David McGloin.

I thank all my colleagues and friends, especially Sahand Ghavidel, Mojtaba Jabbari Ghadi, Ali Azizivahed, Mohsen Eskandari, and Mohammad Abuhilaleh with whom I have shared significant knowledge and experience during the course of my student life.

I should thank my friend Dr Ravindra Palavalli Nettimi for his special support, encouragement, and kind help during my studies.

Finally, none of this would be made possible without the endless love and unconditional supports of my family.

Publications

The following publications are part of the thesis.

Journal publications

- [1] A. Rajabi, M. Eskandari, M. Jabbari Ghadi, S. Ghavidel, L. Li, J. Zhang, P. Siano, "A pattern recognition methodology for analyzing residential customers load data and targeting demand response applications," *Energy and Buildings*, vol. 203, p. 109455, 2019.

➤ **This paper is incorporated in Chapter 4.**

- [2] A. Rajabi, M. Eskandari, M. Jabbari Ghadi, L. Li, J. Zhang, P. Siano, "A Comparative Study of Clustering Techniques for Electrical Load Pattern Segmentation," *Renewable and Sustainable Energy Reviews*, vol. 120, p.109628, 2020.

➤ **This paper is incorporated mainly in Chapter 3 and partially in Chapter 4.**

- [3] A. Rajabi, M. Eskandari, L. Li, J. Zhang, "Trends in Applications of Load Data Clustering in Smart Grid Environment," (Under review, *IEEE Systems Journal*)

➤ **This paper is incorporated in Chapter 2.**

[4] A. Rajabi, L. Li, J. Zhang, K. Muttaqi, "A Feature-Based Data Mining Approach for Characterizing Residential Consumption Behavior in Smart Electricity Grids," (Under review, *Energy Research and Social Sciences*)

➤ **This paper is incorporated in Chapter 5.**

[5] A. Rajabi, M. Jabbari Ghadi, S. Ghavidel, L. Li, J. Zhang, "A Clustering-Based Framework for Development of Time of Use Tariffs for Residential Electricity Customers," (To be submitted)

➤ **This paper is incorporated in Chapter 6.**

Conference publications

[6] A. Rajabi, L. Li, J. Zhang, and J. Zhu, "Aggregation of small loads for demand response programs—Implementation and challenges: A review," in *Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), 2017 IEEE International Conference on*, 2017, pp. 1-6

➤ **This paper is incorporated in Chapter 2.**

[7] A. Rajabi, L. Li, J. Zhang, J. Zhu, S. Ghavidel, and M. J. Ghadi, "A review on clustering of residential electricity customers and its applications," in *2017 20th International Conference on Electrical Machines and Systems (ICEMS)*, 2017, pp. 1-6.

➤ **This paper is incorporated in Chapter 2.**

Abstract

Smart grids are equipped with advanced smart metering infrastructures that enable the two-way communication between end-users and utilities and record the consumption data of electricity customers. The gathered smart meter data have opened up possibilities for analyzing the consumption behavior of customers and understanding the underlying factors affecting it. However, the dimensionality of recorded data necessitates the use of data analysis techniques to extract valuable information from the load profiles. In this regard, this thesis utilizes various methods to analyze the consumption and survey data of residential electricity customers.

In the first two chapters, the concepts of smart metering, residential customers' load patterns, and clustering of load data are extensively discussed, and a comprehensive discussion of the extant literature is presented.

Chapter 3 presents a comparative study of five main clustering approaches including K-means, fuzzy c-means, hierarchical, self-organizing map, and Gaussian mixture models for load pattern segmentation. Various parameters of each of these methods are explained in detail and their performances are compared using six cluster validity indexes. The obtained results are analyzed to find out the characteristic load shapes among the load curves of customers and to identify the main consumption patterns.

The problems of data deluge and residential DR establishment are addressed in Chapter 4 using a combination of symbolic aggregate approximation (SAX), as a suitable dimensionality reduction technique, a clustering algorithm, and the entropy concept. The use of SAX can assist in the clustering of residential load patterns, which usually display

high variability. Moreover, the results are utilized for ranking the customers based on their stability in usage patterns over time which is beneficial for different DR programs.

In Chapter 5, both the consumption data and survey data of residential electricity customers are used to find out the effects of the households' socio-demographic attributes and building characteristics on load patterns.

In Chapter 6, a combination of clustering algorithms and optimization models are used to design TOU tariffs for electricity customers. The problem is modeled as a mixed-integer linear programming problem with the objective of maximizing the profits of an electricity retailer that participates in different market settlements. The stochastic programming technique is used to address the uncertainties in future load and price.

Finally, in the last chapter, future directions in the analysis of smart meter data and the clustering of load patterns are briefly reported. Furthermore, the proposals for future work are elaborated in this chapter.

Keywords: Residential Load Data; Data Mining; Clustering; Demand Response; Data Size Reduction; Time of Use Tariffs

Contents

Certificate of Original Authorship	i
Acknowledgments	ii
Publications.....	iii
Abstract.....	v
List of Tables	xii
List of Figures.....	xiv
List of Abbreviations	xvi
1 Introduction.....	1
1.1 Background and Research Question	1
1.2 Research Objectives and Scope	4
1.3 Dataset.....	5
1.3.1 Consumption dataset.....	5
1.3.2 Survey dataset	6
1.4 Software	8
1.5 Contributions and Organization of the Thesis	10
2 Background and literature review	13
2.1 Smart Metering	13
2.2 Residential Load Shape Characteristics.....	16
2.3 Effects of New Technologies on Load Patterns.....	20
2.3.1 Renewable energy status.....	20
2.3.2 Storage systems.....	22
2.3.3 Electric vehicles.....	23
2.3.4 Impact of newer technologies on load profiles	24
2.4 Load Pattern Clustering	29
2.4.1 Clustering concepts.....	29
2.4.2 History of electricity customer clustering.....	30
2.4.3 Stages of load pattern clustering	32
2.5 Literature on Clustering of Load Data	35

2.6 Clustering Applications in Smart Grid Environment.....	39
2.6.1 Load forecasting.....	41
2.6.2 Demand response.....	48
2.6.3 Tariff design.....	53
2.6.4 Classification of new electricity customers.....	57
2.6.5 Non-technical loss detection.....	60
2.6.6 Other applications.....	63
i) Finding the relationship between household characteristics and consumption patterns.....	63
ii) Defining new class load curves.....	65
2.7 Summary.....	67
3 A Comparative Study of Clustering Techniques for Electrical Load Pattern Segmentation.....	68
3.1 Motivation and Objectives.....	68
3.2 Clustering Algorithms.....	69
3.2.1 Distance-based methods.....	69
3.2.2 Self-Organizing Map.....	73
3.2.3 Probabilistic and generative models.....	74
3.3 Discussion on the Algorithms.....	76
3.3.1 K-center family.....	76
3.3.2 Hierarchical.....	77
3.3.3 SOM.....	79
3.3.4 GMM.....	80
3.4 Application of Clustering Algorithms to the Load Curves of Customers.....	81
3.4.1 Cluster validity indexes.....	82
3.4.2 Fuzzy c-means.....	88
3.4.3 Hierarchical clustering.....	89
3.4.4 SOM.....	90
3.4.5 GMM.....	92
3.4.6 Comparing clustering methods.....	93
3.4.7 Clustering of a large number of electricity customers.....	97
3.4.8 Method comparisons based on the computation time.....	99
3.5 Summary.....	101

4	A Pattern Recognition Methodology for Analyzing Load Data and Targeting Demand Response Applications.....	103
4.1	Background and Motivation.....	103
4.2	Preliminary Stages Before the Clustering.....	105
4.2.1	Feature definition.....	106
4.2.2	Feature extraction.....	107
4.2.3	Dimensionality reduction methods.....	108
4.3	Problem Statement.....	110
4.3.1	Stages of the method.....	112
4.4	Methodology.....	113
4.4.1	SAX method.....	113
4.4.2	Clustering stage.....	115
4.4.3	DR application.....	118
4.5	Preliminary Analysis of the Dataset.....	120
4.6	Case Study.....	123
4.6.1	Application of SAX and clustering algorithms.....	123
4.6.2	Analysis of weekday and weekend clusters.....	126
4.6.3	Entropy analysis.....	128
4.6.4	Effect of amplitude partitioning and number of clusters.....	129
4.6.5	Entropy analysis for a large number of customers.....	131
4.7	Comparison with current methods and applications.....	134
4.7.1	Current practices for DR aggregation.....	134
4.7.2	Challenges of DR aggregation.....	136
4.7.3	Application of the proposed method.....	139
4.8	Summary.....	140
5	Investigating the Effect of Household Characteristics on Consumption Patterns	
5.1	Background and Motivation.....	142
5.2	Stages of the Method.....	144
5.3	Clustering.....	146
5.4	Variable Selection.....	148
5.4.1	Considerations about the variables.....	148
5.4.2	Determining the association among variables.....	153

5.5 Final Variable Selection and Classification	154
5.5.1 Final Variable Selection.....	154
5.5.2 Classification method.....	156
5.6 Applications	158
5.6.1 Clustering results.....	158
5.6.2 Chi-squared test	162
5.6.3 Classification.....	163
5.7 Implications of the study findings.....	167
5.8 Summary	169
6 Tariff Design.....	170
6.1 Background and Motivation.....	170
6.2 Procedure	172
6.2.1 Problem statement and stages	172
6.2.2 Clustering.....	172
6.2.3 Stochastic programming	173
6.2.4 Risk measure	174
6.3 Formulation.....	175
6.3.1 Forward contracts costs.....	175
6.3.2 Expected day-ahead market costs/revenues	176
6.3.3 Revenues.....	177
6.3.4 TOU structure	179
6.3.5 Energy storage units.....	180
6.3.6 CVaR formulation.....	181
6.4 Objective Function.....	181
6.4.1 Linearization	184
6.4.2 Other considerations	188
6.5 Numerical Results.....	190
6.5.1 Data.....	190
6.5.2 Characteristics of clusters	192
6.5.3 Results.....	194
6.6 Summary	207
7 Conclusions and Suggestions for Future Work	208

7.1 Summary and Conclusions.....208
7.2 Future Trends in Data Analytics of Smart Grids211
7.3 Future Directions of Load Data Clustering.....213
7.4 Future Work.....216
References 219

List of Tables

Table 2-1 Smart meters' projected rollouts globally.....	15
Table 2-2 Renewable energy capacity	20
Table 2-3 The most important variables of in-home surveys	31
Table 2-4 Clustering methods for load data clustering	35
Table 2-5 Studies that used clustering techniques for the improvement of load forecasts	44
Table 2-6 Studies that used clustering techniques for DR analysis	51
Table 2-7 Studies that used clustering techniques for designing tariff structures.....	55
Table 2-8 Studies that used clustering techniques as a preliminary stage prior to the classification..	58
Table 2-9 Studies on NTL which used clustering techniques.....	62
Table 2-10 Use of clustering techniques for studying the characteristics of households	64
Table 2-11 Studies that used clustering techniques for defining new class load curves.....	67
Table 3-1 Characteristics of main methods of K-centers family	77
Table 3-2 Linkage criteria for hierarchical clustering.....	78
Table 3-3 List of CVIs	84
Table 3-4 Comparison of processing time for different clustering algorithms	99
Table 3-5 Comparison of processing time for different CVIs	100
Table 4-1 A sample assignment of daily curves of two customers to different clusters.....	119
Table 4-2 Characteristic time intervals of the day (used for the SAX method).....	122
Table 4-3 Managerial and technical challenges of DR aggregation	136
Table 5-1 The main time periods of household activity during the day	146
Table 5-2 Defined features for each customer	147
Table 5-3 Selected survey variables.....	149
Table 5-4 Definition of social classes	150
Table 5-5 Selected variables after the initial evaluation	152
Table 5-6 GKT values for the variables with a relatively high strength of association.....	154
Table 5-7 Characteristics of formed clusters	160
Table 5-8 Contingency table for the CFL and cluster membership	162
Table 5-9 Results of MLR classification	165
Table 6-1 Attitudes toward energy saving and knowledge about energy reduction: the responses are on a scale of 1 to 5 where 1 is “strongly agree” and 5 is “strongly disagree”.	189
Table 6-2 The values of different parameters for numerical studies.....	191
Table 6-3 Characteristics of forward contracts	192
Table 6-4 Summer retail prices for benchmark scheme (without clustering).....	194
Table 6-5 Winter retail prices for benchmark scheme (without clustering)	195
Table 6-6 Obtained high, medium, and low retail prices for each cluster for summer.....	196
Table 6-7 Obtained high, medium, and low retail prices for each cluster for winter	197
Table 6-8 Purchased energy from forward contracts (summer).....	200
Table 6-9 Purchased energy from forward contracts (winter)	200
Table 6-10 Forward contracts when the battery is not considered (summer)	204
Table 6-11 Forward contracts when selling in day-ahead market is allowed (summer).....	205

Table 6-12 Retail prices when selling in day-ahead market is allowed (summer).....205
Table 6-13 Comparison of retailer’s profit and risk for different values of β (summer).....206
Table 6-14 Comparison of retailer’s profit and risk for different values of β (winter)207

List of Figures

Fig. 2.1 Structure of smart metering system	14
Fig. 2.2 Variability of the consumption of a household over a week	17
Fig. 2.3. Effect of seasonality on the consumption of a household	17
Fig. 2.4 Construction of representative load patterns for electricity customers.....	19
Fig. 2.5 Annual additions of solar PV and wind power	21
Fig. 2.6 Cost of battery in different years	23
Fig. 2.7 Stages of load pattern clustering.....	33
Fig. 2.8 Architecture of BI systems of companies	40
Fig. 2.9 Different methods for load forecasting: (a) Completely aggregated. (b) Completely disaggregated. (c) Clustering-based forecasting	43
Fig. 2.10 Fundamental functions of DR programs.....	49
Fig. 2.11. Customized tariff structures for different clusters	54
Fig. 2.12. Classification of electricity customers.....	58
Fig. 3.1 Dendrogram formed by a hierarchical clustering method	72
Fig. 3.2 Effect of fuzziness degree on the clustering results for FCM method.....	89
Fig. 3.3. Comparison of hierarchical algorithms	90
Fig. 3.4. Dendrograms of (a) ward method and (b) single method.....	90
Fig. 3.5. A 16×16 SOM grid and the corresponding clusters after applying the hierarchical method	91
Fig. 3.6. Effect of grid size and topology on the two-level clustering of load curves using SOM and hierarchical method (R: Rectangular, H: Hexagonal).....	92
Fig. 3.7. Effects of parameters of covariance matrix on the GMM clustering.....	93
Fig. 3.8. The CVI values for four different clustering algorithms	94
Fig. 3.9 Final clusters of 4 different clustering algorithms	96
Fig. 3.10. GMM clustering results	97
Fig. 3.11 Clusters of the weekday RLPs of 4141 customers	98
Fig. 3.12. Clusters of the weekend RLPs of 4141 customers	99
Fig. 4.1. Performance of a combined clustering of PCA and K-means for different number of clusters and PCs	109
Fig. 4.2. Consumption behavior of two customers	111
Fig. 4.3. Stages of the methodology.....	112
Fig. 4.4. PAA and SAX representation of a load curve for 5 days	115
Fig. 4.5. Boxplots of weekday consumption in different seasons.....	121
Fig. 4.6. Boxplots of weekend consumption in different seasons.....	121
Fig. 4.7. Daily total consumption and temperature variation.....	124
Fig. 4.8. CDF of the whole data.....	125
Fig. 4.9. Centers of the clusters with the highest number of daily load curves for the weekday data set.....	126
Fig. 4.10. Centers of the clusters with the highest number of daily load curves for the weekend data set.....	128

Fig. 4.11. Load curves of two customers with stable consumption behavior (top) and two customers with variable consumption behavior (bottom)	129
Fig. 4.12. DBI values for different size of alphabets and different number of clusters	130
Fig. 4.13. MIA values for different size of alphabets and different number of clusters	130
Fig. 4.14. Cluster assignment distribution	132
Fig. 4.15. Load curves of sample customers with stable consumption behavior for weekday dataset	133
Fig. 4.16. Load curves of sample customers with stable consumption behavior for weekend dataset	133
Fig. 4.17. Demand response aggregator model.....	135
Fig. 5.1 Stages of the method including clustering, variable selection, and classification modules.....	145
Fig. 5.2 Selecting the best number of clusters	160
Fig. 5.3 Visualizing the contribution of each cell of contingency table to the Chi-squared value ...	163
Fig. 6.1 Stages of designing customized TOU structures	172
Fig. 6.2 Forward contracts curve	176
Fig. 6.3 Revenue function and approximated piecewise revenue function.....	186
Fig. 6.4 Six formed clusters for summer.....	193
Fig. 6.5 Six formed clusters for winter	193
Fig. 6.6 The summer retail structure for the benchmark scheme (without clustering)	195
Fig. 6.7 The winter retail structure for the benchmark scheme (without clustering).....	195
Fig. 6.8 TOU price structure for each cluster for summer	197
Fig. 6.9 TOU price structure for each cluster for winter.....	198
Fig. 6.10 Purchased energy from day-ahead market (summer)	199
Fig. 6.11 Purchased energy from day-ahead market (winter).....	200
Fig. 6.12 Charge/discharge states of the storage unit (summer).....	201
Fig. 6.13 Charge/discharge states of the storage unit (winter).....	201
Fig. 6.14 Summer TOUs without forward contract	202
Fig. 6.15 Winter TOUs without forward contract.....	203
Fig. 6.16 Purchased/sold energy in day-ahead market (summer)	204

List of Abbreviations

AIC	Akaike's information criterion
AMI	Advanced Metering Infrastructure
AMR	Automatic Meter Reading
ARMA	Auto Regressive with Moving Average
BI	Business Intelligence
BIC	Bayesian Information Criterion
CER	Commission for Energy Regulation
CFL	Compact Fluorescent Lamp
CFSFDP	Fast Search and Find of Density Peaks
CVI	Clustering Validity Index
DBI	Davies-Bouldin Indicator
D ² R	Dynamic Demand Response
DFT	Discrete Fourier Transform
DLC	Direct Load Control
DR	Demand Response
DMS	Data Management System
DTW	Dynamic Time Warping
DWT	Discrete Wavelet Transform
EM	Expectation-Maximization
FCM	Fuzzy C-Means
GMM	Gaussian mixture model

HMM	Hidden Markov Model
HDF	Household and Dwelling Features
I/C	Interruptible/Curtailable load
ISODATA	Iterative Self-Organizing Data Analysis Technique Algorithm
KEPCO	Korea Electric Power Corporation
KNN	K-Nearest Neighbor
LR	Logistic Regression
MAPE	Mean Absolute Percentage Error
MIA	Mean Index Adequacy
MSE	Mean Square Error
NTL	Non-Technical Losses
OPF	Optimum-Path Forest
PAA	Piecewise Aggregate Approximation
PAM	Partitioning Around Medoids
PG&E	Pacific Gas and Energy Company
PCA	Principal Component Analysis
RLP	Representative Load Pattern
SAX	Symbolic Aggregate approxXimation
SIL	Silhouette index
SOM	Self-Organizing Maps
SVM	Support Vector Machines
SVR	Support Vector Regression
TLP	Typical Load Profile
TOU	Time of Use
WCBCR	The ratio of within-cluster sum of squares to between-cluster variation