

University of Wollongong - Research Online

Thesis Collection

Title: Discovery and pattern classification of large scale harmonic measurements using data mining

Author: Ali Asheibi

Year: 2009

Repository DOI:

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following: This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of this work may be reproduced by any process, nor may any other exclusive right be exercised, without the permission of the author. Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material.

Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Unless otherwise indicated, the views expressed in this thesis are those of the author and do not necessarily represent the views of the University of Wollongong.

Research Online is the open access repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

2009

Discovery and pattern classification of large scale harmonic measurements using data mining

A. Asheibi
University of Wollongong, ali_asheibi@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/theses>

University of Wollongong

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following: This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of this work may be reproduced by any process, nor may any other exclusive right be exercised, without the permission of the author. Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material.

Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Unless otherwise indicated, the views expressed in this thesis are those of the author and do not necessarily represent the views of the University of Wollongong.

Recommended Citation

Asheibi, Ali Taher M, Discovery and pattern classification of large scale harmonic measurements using data mining, PhD thesis, School of Electrical, Computer & Telecommunications Engineering, University of Wollongong, 2009. <http://ro.uow.edu.au/theses/558>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

NOTE

This online version of the thesis may have different page formatting and pagination from the paper copy held in the University of Wollongong Library.

UNIVERSITY OF WOLLONGONG

COPYRIGHT WARNING

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Discovery and Pattern Classification of Large Scale Harmonic Measurements using Data Mining

A thesis submitted in fulfilment of the
requirements for the award of the degree

Doctor of Philosophy

from

University of Wollongong

by

Ali Taher M. Asheibi, BSc(Eng), MSc(Eng)

**School of Electrical, Computer and Telecommunications
Engineering**

March 2009

Dedicated to my parents...

Acknowledgements

It is my pleasure to thank the many people to whom I am indebted for the development of this thesis. First and foremost, thanks go to my supervisors, Dr David Stirling, Professor Danny Sutanto and Dr Duane Robinson. Their dedication, knowledge and experience could not have been surpassed.

Thanks to Sean Elphic, Neil Brown and Dr Vic Smith of the Integral Energy Power Quality and Reliability Center who have responded to many technical, administrative and software related requests for assistance.

Tim Brown, Ahsan Lateef, Matthew Field and Praboda Parनावithana, presently and previously with the Integral Energy Power Quality and Reliability Center , have been the sources of many interesting discussions which have contributed to the PhD experience.

Thanks to family who exercised considerable patience even in the face of typical thesis consequences. Particular thanks go to my wife Faesa Netfa who has suffered from many such consequences and been very kind to admit it.

Certification

I, Ali Taher M. Asheibi, declare that this thesis, submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Electrical, Computer and Telecommunications Engineering at the University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualification at any other academic institution.

.....

Ali Taher M. Asheibi

12 March 2009

Abstract

Harmonic monitoring is an important issue for electricity utilities and their customers. Continuous monitoring of voltage and current are required to identify any substantial harmonic events before they occur. This monitoring results in large volumes of multivariate data. Although researchers have realised that such large amounts of power quality (PQ) data hold much more information than that reported using classical statistical techniques for PQ monitoring, few have taken the opportunity to exploit this additional information. This hidden information might be of assistance in the identification of critical issues for diagnoses of harmonic problems such as, predicting failures in advance and giving alarms prior to the onset of dangerous situations.

Utility engineers are now seeking new tools in order to extract information that may otherwise remain hidden, especially within large volumes of data. Data mining tools are an obvious candidate for assisting in such analysis of large scale data. Data mining can be understood as a process that uses a variety of analytical tools to identify hidden patterns and relationships within data. Classification based on clustering is an important utilisation of unsupervised learning within data mining, in particular for finding and describing a variety of patterns and anomalies in multivariate data through various machine learning techniques and statistical methods. Clustering is often used to gain an initial insight into complex data and particularly in this case, to identify underlying classes within harmonic data.

The main data mining methodology used in this work is that of mixture modelling based on the Minimum Message Length (MML) algorithm which essentially searches for a model which best describes the data using a metric of an encoded message. This method of unsupervised learning, or clustering, has been shown to be able to detect anomalies and identify useful patterns within the monitored harmonic data set. Anomaly detection and pattern recognition in harmonic data can provide engineers with a rapid, visually oriented method for evaluating the underlying operational information contained within the data set.

A case study from power quality data upon which the MML method has been ap-

plied, was taken from a harmonic monitoring program installed in a typical 33/11kV MV zone substation in Australia that supplies ten 11kV radial feeders. Several patterns have been identified from using the MML technique on the harmonic data, such as significant high harmonic disturbances, footprints of the monitored sites, unusual harmonic events (capacitor switching, turn on televisions, air conditioners and the off peak hot water system) and detection of different abstractions (super-groups), each of which comprise similar clusters. The C5.0 supervised learning algorithm has been used to generate expressible and understandable rules which identify the essential features of each member cluster, and to further utilize these in predicting which ideal clusters any new observed data may best be described by.

One difficulty with the MML algorithm when used to derive various mixture models is the difficulty in establishing a suitable stopping criterion to secure the optimum number of (mixture) clusters during the clustering process. A novel technique has been developed to overcome this difficulty using the trend of the exponential of message length difference between consecutive mixture models. First, the proposed method has been tested using data from known number of clusters with randomly generated data points and also with data from a simulation of a power system. The results from these tests confirm the effectiveness of the proposed method in finding the optimum number of clusters. Second, the developed method has been applied to various two-weekly data sets from the harmonic monitoring program used on this thesis. The optimum number of clusters has been verified by the formation of super-groups using Multidimensional Scaling (MDS) and link analysis. Third, the method was benchmarked against a commonly used fitness function technique, which has underestimated the optimal number of cluster in the measured harmonic data. This resulted from the theoretical maximum entropy equation used in calculating the fitness function that assumes the attributes are independent which is not the case in the correlated nature of the harmonic attributes. Finally, generated rules from the C5.0 algorithm were used for classification and prediction of future events to determine which cluster any new data should belong to.

List of Symbols and Abbreviations

ANN	Artificial Neural Network.
Aom	Accuracy of measurement.
DFT	Discrete fourier Transform.
D	Data set.
DWT	Direct Wavelet transform.
CT Fund.	Fundamental current.
CT Harm 3	Third harmonic current.
CT Harm 5	Fifth harmonic current.
CT Harm 7	Seventh harmonic current.
CT Harm 19	Nineteenth harmonic current.
CT Harm 49	Forty-ninth harmonic current.
CT THD	Total harmonic current distortion.
f	Frequency.
FT	Fourier transform.
FFT	Fast Fourier transform.
IEC	International Electrotechnical Commission.
K	Mixture of clusters model.
KDD	Knowledge Discovery in Databases.
KL	Kulback Lieber distance.
LV	Low voltage.
MV	Medium voltage.
MDS	Multidimensional scaling.
MML	Minimum Message Length.
MVAr	Reactive power Q.
PCA	Principle Component Analysis.
PQ	Power Quality.
Ph Fund	Fundamental voltage.
Ph Harm 3	Third harmonic voltage.
Ph Harm 5	Fifth harmonic voltage.
Ph Harm 7	Seventh harmonic voltage.
Ph Harm 19	Nineteenth harmonic voltage.
Ph Harm 49	Forty ninth harmonic voltage.
Ph Total H Dist	Total harmonic voltage distortion.
rms	Root mean square.
SOM	Self Organising Map.
ST	S-transform.
SVM	Support Vector Machine.
WT	Wavelet transform.

Publications arising from this Thesis

1. A. Asheibi, D. Stirling, and D. Sutanto and D. Robinson, "*Clustering, classification and explanatory rules from harmonic monitoring data*", Book Chapter in "*Theory and Novel Applications of Machine Learning*", Men Joo Er and Yi Zhou, Eds., I-Tech Education and Publishing, Vienna, Austria, February 2009.
2. A. Asheibi, D. Stirling, and D. Sutanto, *Analyzing Harmonic Monitoring Data using Supervised and Unsupervised Learning.*, IEEE Transactions on Power Delivery, Vol. 24, No.1, pp. 293-301, January 2009.
3. A. Asheibi, D. Stirling, and D. Sutanto, *Classification and Explanatory Rules of Harmonic Data*, Proc. Australasian Universities Power Engineering Conference (AUPEC 2008), 14-17 December 2008 Sydney, Australia, Paper ID: 259.
4. A. Asheibi, D. Stirling, and D. Sutanto, *Determination of the Optimal Number of Clusters in Harmonic Data Classification*, Proc. of the 13th International Conference on Harmonics and Quality of Power (ICHQP 2008), 28 September-1 October 2008, Wollongong, NSW, Australia, Paper 1045.
5. A. Asheibi, D. Stirling, and D. Sutanto, *Analyzing Harmonic Monitoring Data using Data Mining*, Proc. Fifth Australasian Data Mining Conference(AusDM06), 29-30 November, 2006, Sydney, NSW, Australia, pp: 63-68.
6. A. Asheibi, D. Stirling, and D. Sutanto, *Analyzing Harmonic Monitoring Data using Data Mining*, Conferences in Research and Practice in Information Technology (CRPIT), 61. Peter, C., Kennedy, P.J., Li, J., Simoff, S.J. and Williams, G.J., Eds., Australian Computer Society Inc. (ACS), 2006, pp: 63-68.
7. A. Asheibi, D. Stirling, and D. Robinson, *Identification of Load Power Quality Characteristics using Data Mining. Proc. of the Canadian Conference on Electrical and Computer Engineering, 2006. CCECE '06.*, 7-10 May 2006, Ottawa, Canada, pp: 157-162.

8. A. Asheibi, D. Stirling, S. Perera and D. Robinson, *Power quality data analysis using unsupervised data mining*, Proc. Australasian Universities Power Engineering Conference (AUPEC 2004), 26-29 September 2004, Brisbane, Australia, Paper ID: 187.

Table of Contents

1	Introduction	1
1.1	Problem statements and background	1
1.2	Thesis objectives and methodology	2
1.3	Thesis outline and summary of original contributions	3
2	Literature Review	7
2.1	Introduction	7
2.2	Power Quality Monitoring	7
2.2.1	Power quality monitoring campaigns	8
2.3	Power quality reporting and data analysis	9
2.3.1	Power quality indices	9
2.4	Signal processing in power quality data analysis	10
2.4.1	Fourier transform (FT)	10
2.4.2	Wavelet transform (WT)	12
2.4.3	S-transform (ST)	14
2.5	Data mining	14
2.5.1	Data mining versus statistics	15
2.5.2	Data mining versus machine learning	16
2.5.3	Data mining applications	16
2.5.4	Data mining in power quality data analysis	19
2.6	Unsupervised learning and supervised learning	19
2.7	Classification of power quality events using supervised learning	21
2.7.1	Supervised learning classification of power quality events using artificial neural network (ANN)	22
2.7.2	Classification of power quality events using Bayesian classifiers	23
2.7.3	Classification of power quality events using Support Vector Ma- chines (SVM)	24
2.7.4	Classification of power quality events using expert systems . .	25
2.8	Clustering as unsupervised learning	25
2.8.1	Why clustering?	26
2.8.2	Clustering objectives	27
2.8.3	Clustering algorithms and types	27
2.9	Summary	34
3	Mixture Modelling Method using Minimum Message Length (MML) Technique	36
3.1	Introduction	36
3.2	Mixture modelling	37
3.2.1	Parameter estimation and model selection	38
3.2.2	The Expectation Maximisation (EM) algorithm [49]	40
3.2.3	Fitting a model to a mixture of statistical distributions	41

3.3	Minimum Message Length (MML) Technique in Mixture Modelling Method	43
3.3.1	Minimum Message Length	46
3.4	Comparison between Mixture Modelling Method using MML technique and other clustering and feature extraction algorithms	55
3.4.1	Comparison between the Mixture Modelling using MML technique and traditional feature extraction methods based on signal processing techniques	55
3.4.2	Comparison between Mixture Modelling using MML with other distance base clustering methods.	56
3.5	Summary	59
4	Optimal Number of Clusters	60
4.1	Introduction	60
4.2	Determination of the optimal number of clusters	61
4.2.1	Effect of the number of clusters	61
4.2.2	Fitness function determination of the optimal number of clusters	63
4.2.3	Using Mixture Modelling based on MML to determine the optimum number of clusters	65
4.3	Summary	71
5	Harmonic data collection and preparation for data mining techniques	73
5.1	Introduction	73
5.2	Harmonic monitoring program and System study	73
5.2.1	Identification of load types from selected monitored sites	74
5.2.2	Harmonic monitoring equipment	76
5.2.3	Australian power quality standards	76
5.2.4	Harmonic data sampling	78
5.2.5	Harmonic data measurement	79
5.2.6	Harmonic monitoring data set	79
5.2.7	Harmonic data preparation	80
5.2.8	Harmonic voltage and current trends	82
5.2.9	Harmonic data selection	89
5.2.10	Rescaling of harmonic data	91
5.2.11	Normalisation of harmonic data	91
5.2.12	Other measured data (temperature and reactive power)	93
5.3	Summary	94
6	Anomaly detection and pattern recognition	95
6.1	Introduction	95
6.2	Data preparation	96
6.3	Anomaly detection and pattern recognition from harmonic clusters	97
6.4	Abstraction of super groups from harmonic data	102
6.4.1	Kullback Leibler Distance (KL)	102

6.4.2	Multidimensional scaling (MDS)	103
6.4.3	Segmentation of harmonic data into Super-groups using KL and MDS	104
6.5	Decision tree of supervised learning	108
6.6	Rules discovered from the super-groups using decision tree	109
6.6.1	Visualisation of the the super-groups generated rules	112
6.7	Summary	115
7	Harmonic event detection using supervised and unsupervised learning	118
7.1	Introduction	118
7.2	Results from unsupervised learning using MML	119
7.2.1	Interpretations of the generated clusters	124
7.3	Results from supervised learning using C5.0	130
7.3.1	Prediction of capacitor switching with C5.0 and lagging window	131
7.4	Summary	135
8	Determination of the Optimal Number of Clusters in Harmonic Data Classifi- cation	137
8.1	Introduction	137
8.2	Optimal number of clusters in harmonic data	138
8.3	Results from the study system harmonic monitoring data	139
8.4	Using Fitness Function to determine the optimal number of clusters .	140
8.5	Verification of the optimum model using Super-groups	141
8.6	Interpretation of the Optimal Number of Clusters in Harmonic Data using supervised learning	147
8.6.1	Rules discovered from the optimum clusters using decision tree	151
8.6.2	Rules for prediction of harmonic future data	154
8.7	Summary	155
9	Conclusions	157
9.1	Conclusions and recommendations	157
9.2	Future work	160

List of Figures

1.1	A comprehensive understanding of major building blocks of this thesis.	6
2.1	Transforming the signal from time domain to frequency domain using Fourier transform for (a) pure sine wave and (b) distorted sine wave.	11
2.2	(a) Unsupervised learning and (b) Supervised learning.	20
2.3	Bayesian network where x1 and x2 are independent and x3 is dependent variables. (adopted from [12]).	23
2.4	The XOR problem where the class A is defined if and only if x or y equal 1 but not both.	24
2.5	Flow chart of K-means algorithm.	28
2.6	Kohonen model of feature-mapping (adopted from [45]).	31
2.7	Input and output layers of SOM (adopted from [45]).	31
2.8	Principle components PC1 and PC2 of two dimensional data.	32
2.9	Typical mixture modelling of five Gamma distributions using MML (adopted from [54]).	34
3.1	Two normal distributions of similar means, standard deviations and proportions a) $\mu_1, \mu_2 = 2$ and b) $\mu_1, \mu_2 = 1$ (adapted here from various Matlab plots).	39
3.2	Most important areas in normal distribution (a) 68% and (b) 95% of values in population.	42
3.3	Fitting normal distribution to a data with 68% of population is highlighted.	43
3.4	Two variables x and y with the area of 68% population in red colour represents the intersection of the two single distributions.	44
3.5	The area of one standard deviation (68% of population) generate square shape from bivariate distribution.	44
3.6	The hyper-cube shape, unlike the ellipsoide can cover the one standard deviation area.	45
3.7	Conceptual flow chart of clustering algorithm of Mixture Modelling Method using MML technique.	50
3.8	Three cluster (30 data point) generated randomly from X1 (cluster1), X2 (cluster2) and X3 (cluster3).	51
3.9	Three randomly generated Clusters.	56
3.10	Correctly clustering of the clusters shown in Figure 3.9 using MML. .	57
3.11	False Clustering of the clusters shown in Figure 3.9 using K-means. .	57
3.12	Centre displacements of clusters 1 and 2 of the clusters shown in Figure 3.9 using Fuzzy C-means.	58
4.1	Five randomly generated clusters each with its own mean and standard deviation.	62
4.2	The clusters obtained superimposed on the randomly generated data.	63
4.3	Fitness function showing five clusters in random data.	65

4.4	Exponential of message length difference identifying five clusters as the optimum number.	67
4.5	A single line diagram of a simplified power system model used in a PSCAD®/EMTDC™ Simulation.	68
4.6	The rms values of voltage and current in phase 'a'.	69
4.7	Exponential of the message length difference of consecutive clusters.	69
4.8	The ten generated clusters superimposed on simulation data.	70
4.9	The clusters statistical parameters mean (μ), standard deviation (σ) and abundance (π).	71
4.10	Fitness function [71] also identifying that 10 is the optimum number of clusters.	72
5.1	Single line diagram illustrating the zone distribution system.	75
5.2	EMDI 2000-04XX Energy Meter.	78
5.3	Zone substation (site 1) weekly harmonic current data from the monitoring equipment.	80
5.4	Residential feeder (site 2) weekly harmonic Current data from the monitoring equipment.	81
5.5	Substation (Site 1) weekly low 3rd harmonic current.	83
5.6	Residential site (Site 5) with a relatively high weekly 3rd harmonic current.	84
5.7	Zone Substation (Site 1) weekly high 3rd harmonic voltage.	85
5.8	Commercial feeder site (site 3) high 5th harmonic currents.	85
5.9	Commercial feeder site (site 3) high 5th harmonic voltages.	86
5.10	Commercial feeder site (site 3) 7th harmonic voltage and current.	87
5.11	Substation site (site 1) low 19th harmonic voltage.	88
5.12	Substation site (site 1) total harmonic distortion (THD) and 5th harmonic current and voltage.	90
6.1	Abundance of clusters of 5th harmonic current and voltages over each phase of monitoring results.	98
6.2	Cluster of 5th harmonic current and ITHD over all three phases from Site 7.	100
6.3	Five randomly generated clusters each with its own mean and standard deviation.	101
6.4	Clusters of harmonic emissions from the different customer loads and system overall for a one week period.	102
6.5	Abundance, mean and standard deviation for each cluster of the 5 th harmonic current.	105
6.6	Super-group abstraction by MDS.	107
6.7	Super-groups in all sites over one week.	107
6.8	High 7th harmonic current at industrial site causing high 7th harmonic voltage at substation.	111
6.9	Rules A1 and D1 are synchronised on Thursday, Friday and Saturday at the industrial and the substation sites in one week time frame.	112

6.10	Evidence of Fifth harmonic producing loads at phase C due to commercial site.	113
6.11	Visualization of Rule A1 at the industrial site for one week data. . . .	114
6.12	Visualisation of Rule B1 at commercial site for a one week period. . .	115
6.13	Visualization of Rule E1 at commercial site for one week period. . . .	116
7.1	Message length vs. increasing mixture model size (number of clusters). .	120
7.2	Abundance, mean and standard deviation for each cluster of 5th harmonic current per phase.	120
7.3	Graphical profile view of model clusters indicating the statistical parameters mean (μ), standard deviation (σ) and abundance (π).	122
7.4	(a) Model of six Gaussian distribution clusters obtained at sites(1-4) and (b) The data fitted to the model.	123
7.5	Clusters at substation site in two working days (a) Clusters superimposed on the fundamental current waveform, (b) 7th harmonic current and voltage data. (c) MVar load at the 33kV.	125
7.6	Three normal temperature days at the residential site (Site 2), (a) Fundamental current and generated clusters, (b) 5th harmonic voltage and generated clusters, (c) temperature near Site 2.	127
7.7	Three hot days at the residential site (site 2), (a) fundamental current and generated clusters, (b) 5th harmonic voltage and generated clusters, (c) Temperature near site 2. (c) MVar load at the 33kV. . .	127
7.8	Normal and hot days at Residential site (site 2).	128
7.9	5th harmonic current clusters at industrial site (site 4) for different week days.	129
7.10	5th harmonic current clusters at commercial site (site 3) for two different week days.	129
7.11	Rule-1 of predicting Cluster (s2) of capacitor switching explained in Table 7.3.	133
7.12	Three rules predicting Cluster (s2) associated with capacitor switching events: (a) Rule-1 (b) Rule-2 (c) Rule 3.	134
7.13	Prediction of s2; more than one Rule can occur at the same time instant. .	135
8.1	(a) Detection of sixteen clusters of harmonic data, (b) Enlargement of (a).	140
8.2	The statistical parameters mean(μ), standard deviation (σ) and abundance (π) of the 16 clusters.	141
8.3	Sixteen clusters superimposed on four sites (a) Substation, (b) Residential, (c) Commercial and (d) Industrial.	142
8.4	Fitness function showing only five clusters as optimum number. . . .	143
8.5	Exponential curve for the maximum number of the generated clusters. .	145
8.6	The statistical parameters mean(μ), standard deviation (σ) and abundance (π) of large model with 30 clusters.	146
8.7	The KL distances between the 30 clusters sorted in ascending order. .	147

8.8	Multidimensional scaling: KL-distances are mapped as cumulative link lengths in the graph between any pair of clusters; Super group abstractions are formed through removal of links whose KL-distances exceed a pre-determined dissimilarity threshold.	148
8.9	The statistical parameters mean(μ), standard deviation (σ) and abundance (π) of the super-Groups (A, B, C, ..., P).	149
8.10	The 16 clusters(s0, s1, ..., s16) of the optimum model superimposed on the super-Groups (A, B, C, ..., P) on four sites (a) Substation, (b) Residential, (c) Commercial and (d) Industrial for two days.	150
8.11	The five regions of Gaussian distribution used to convert the numeric values.	152
8.12	Prediction Model accuracy levels for the clusters s7-s10 on training and future data.	154
8.13	Exponential of message length difference for data with and without hot days.	155

List of Tables

3.1	Percentage of values in the population within a given interval.	41
3.2	Data points shown in Figure 3.8.	52
3.3	Segmentation process of data points in Table 3.2.	54
4.1	The parameters (μ and σ) of the five generated clusters.	61
4.2	The load switching operation and timing.	68
4.3	Ten generated clusters with different means and standard deviations.	70
5.1	Proportions of each MV/LV sites based on load types	76
5.2	EMDI MK3 energy meter specifications.	77
5.3	Five clusters generated from ACPro for the 49th harmonic voltage.	88
6.1	Kullback-Lieber distances between components of the 11 cluster mixture model.	105
6.2	Generated rules from super groups (A to E) using the C5.0 algorithm.	110
7.1	Generated model detailing the abundance value (π) of the six cluster a long with the mean (μ) and standard deviation (σ).	121
7.2	Labelling the data with the clusters produced by the MML.	130
7.3	Rules describing cluster s2 generated by C5.0.	132
8.1	The 16 clusters by the method of exponential difference in message length.	143
8.2	Alignment between optimum 16 clusters and super-groups.	148
8.3	KL distances (below the threshold value) of the similar clusters.	149
8.4	The continuous data is grouped into five ranges.	151
8.5	The generated Rules by C 5.0 for clusters s12 and s13.	153
8.6	The accuracy the obtained rules using three months (Jan-Apr 2002) of training and testing data for clusters s7-s10.	153