LOAD PROFILING AND DATA ANALYTICS FOR DISTRIBUTION SYSTEMS

MUDASSIR AZIZAHMED MANIAR

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI
OCTOBER 2019

LOAD PROFILING AND DATA ANALYTICS FOR DISTRIBUTION SYSTEMS

by

MUDASSIR AZIZAHMED MANIAR

Department of Electrical Engineering

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2019

I dedicate this thesis to my parents for all their love and affection and to my wife and children for their patience and support. I appreciate their sacrifices and unconditional love without which this journey wouldn't have been possible.

Certificate

This is to certify that the dissertation entitled '**Load Profiling and Data Analytics for Distribution Systems**', being submitted by **Mr. Mudassir Azizahmed Maniar** for the award of the degree of **Doctor of Philosophy** is a record of bonafide research work carried out by him in the Department of Electrical Engineering at Indian Institute of Technology Delhi, New Delhi.

Mr. Mudassir Azizahmed Maniar has worked under my supervision and has fulfilled the requirements for the submission of this dissertation, which to our knowledge has reached the requisite standard. The results obtained here have not been submitted to any other University or Institute for the award of any degree.

Dr. A. R. Abhyankar

Professor

Department of Electrical Engineering

Indian Institute of Technology Delhi

Hauz Khas, New Delhi 110016, India

Date: October 25, 2019

Acknowledgements

To acknowledge the help and affection of all the people during this journey would perhaps be impossible, nor would it be in any way a measure of my gratitude for them. However, as a duty, it is sacred.

First, I express a deep sense of gratitude for my supervisor Dr A. R. Abhyankar for finding me worthy of being his disciple, suggesting a problem, providing guidance and support throughout the research. No words are adequate to quantify his patience, tolerance and understanding. His unconditional support and constant motivation at the study and personal level have been exemplary and worth emulating. It is an honour and a privilege to be associated with him as a student; without him, this thesis would not have been materialized.

I am grateful to the research committee members, Dr B. K. Panigrahi, Dr N. Sen Roy and Dr Ashu Verma, for their thorough reviews and valuable suggestions during the progress of my research. Their comments and suggestions have been instrumental in making this research worthwhile. A sincere thanks to Dr P. R. Bijwe, Emeritus Professor, IIT Delhi, for moral support.

I thank the authorities of Commissioner of Technical Education, Gandhinagar, for deputing me to pursue PhD degree under Quality Improvement Programme. I want to extend my sincere thanks to Dr U.K. Khare, then Principal, Dr P.K. Brahmbhatt, Principal, GEC, Dahod, Prof. J. I. Patel, Head Department of Electrical Engineering and staff members - Prof. A D Joshi, Ashesh Shah, Megha Mahida, Prafulla K. Bamaniya, Mumtaz Makrani, Hardik Patel, Rakshit Patel, Anil N Navle, Mohammaed

Ishak Shaikh, Arvind S Mishal, Hardik Surti, Megha B Patel, and Mittal M Parmar for all the cooperation and support.

It gives me immense pleasure to acknowledge the splendid company, help and moral support given my fellow research scholars at power computation lab - Ganesh Prajapat, Rishi Sharma, Manas Kumar Jena, Shivraman Mudaliyar, Srikanth Konda, Rajeev Jha, Sayri Das, Ayesha Firdaus, Saleem Mir, Prathyasa Bhui, Sirin Dutta Chaudhary, Kri- tika Saxena, Deep Kiran, Shruti Ranjan, Shazia Rasheed, Megha Gupta, Tabia Ahmed, Snigdha Behra, Deepak Reddy, Dushyant Sharma, Vibhuit Nougain, Rimita Saha, Ashwini Arora and Priyank Shah. I want to thank all of them for their wonderful companionship and making my stay at IIT, Delhi, a memorable one. A sincere thanks to Dr Shri Ram Vaishya, Post Doctoral Fellow, Department of Electrical Engineering, IIT Delhi, for reviewing and providing valuable feedback during the preparation of the thesis. I would like to thanks supporting staff of Power Computation Lab Gauravji and Anandji for all their cooperation.

Among many close friends, Kunal A Bhatt, Yogesh Makwana research scholars IIT, Roorkee and D. A. Patel, former research scholar, Civil Engg Dept. IIT, Delhi and Parth Raval, LDCE, Ahmedabad require special mention for being a pillar of support during the thick and thin phase of this journey.

With the remembrance of my late father Azizahmed Maniar, this PhD is possible because of his love and blessings. I pray to Almighty for our reunion and deliverance on the day of the judgement. I thank my mother, Nasima Khatoon a shade of my life for her unconditional love, blessings and moral support. I thank my Siblings Mussadik, Muzammil, Abdullah and Kulsum for their affection and motivation. And, of course, the list is incomplete without the words of appreciations for my caring wife, Asema for her patience, cooperation, and shouldering the entire responsibility of the family, my children Sadiyah and Shahid for their forbearance throughout this journey.

Place: New Delhi Mudassir A Maniar

Abstract

The Load Profiling and Research started way back in 1934 by AEIC (Association of Edison Illuminating Companies) for tariff design, planning, expansion, etc. With competition and conflicting interests of various entities in the competitive market, these exercises are now being considered essential tools by all players for survival and growth. These activities involve the acquisition, processing and analysis of the massive amount of data acquired continuously by load-serving entities. In this thesis, the load profiling procedure is revisited, an improved partitional clustering algorithm is developed and subspace clustering based approach 'Time Aligned Subspace Clustering' has been proposed for better consumer behavioural analysis. The approach will be more suitable for various demand-side management applications.

A Literature survey indicates that consumer end data is widely preferred for load profiling. There are several developing countries like India, where automated smart metering infrastructure is not well in place. Also, load profiling based on individual household metered data have some ethical and security issues associated with it. Post deregulation, there are many actors in the industry having the opposite positions on market outcome. Further, each player is required to adjust its position as the real-time load-generation balance has to be followed for system stability and non-compliance may result in penalties. Thus, accurate knowledge of future demand is essential, and load profiling and research are tools which provide pre-requisite information. These actors can make use of this publicly available data to devise their strategies in the market.

The study has proposed and used distribution feeder data for load profiling, which is made available in the public domain by many distribution utilities. Also, since

the feeder data represents a combined consumption pattern of a particular area, the outcome of profiling will help the utility to identify areas where Distribution System Management (DSM) activities like load shifting, planning, expansion, etc. would be more beneficial.

Load profiling is a process of grouping load curves of equal time interval based on the similarity of shape. Thus, any algorithm suitable for time series clustering can be adopted. Various load profiling methods suggested in the literature can be adopted, but these require initial guess about a number of clusters. Except hierarchical clustering, most of the algorithms are inherently iterative optimization programs. The algorithm tries to decompose data-set into a set of disjoint clusters by optimizing specific criterion function that emphasizes the local structure of the data. Among all the proposed methods in the literature, the *K-means* and its variants are emerging as a leading choice for load profiling.

The outcome of *K-means* is heavily dependent on the initial conditions, and it leads to the convergence to sub-optimal solutions. The above limitation is addressed to some extent by running the algorithm multiple times with different random initialization and selecting the best outcome from the available solution. The mean pattern of all clusters might look similar to some extent, but the members of clusters are not always unique and change with each run. The effect is very significant if the data-set is big with high dimensionality. All partitional clustering require initial guess about a number of clusters. The initial guess regarding a number of clusters for load profiling is seldom available.

Further, the conventional clustering algorithms do not scale well to cluster high-dimensional data-set regarding effectiveness and efficiency because of their inherent sparsity. This limitation of conventional clustering algorithms in dealing with high dimensional data-sets has motivated the concept of subspace clustering or projected clustering, whose goal is to find clusters embedded in subspaces of the original data space with their associated dimensions. The existing subspace clustering methods have too many parameters, and and it is difficult to find optimal values of parameters combination as per requirement.

The thesis has customized and developed a plug and play tool for load profiling application by reducing stochasticity of *K-means* algorithm and using a *g-Means* algorithm (a variant of *K-means* algorithm which does not require pre-requisite information regarding the number of clusters and evaluate them from input data based on statistical test). The proposed methodology requires one parameter, i.e., significance level α to be specified as an input. Thus, eliminating the requirement of multiple runs of clustering algorithms to finalize the number of clusters and their respective profiles.

The traditional profiling assumes that the behaviour of a particular class of consumer represented by load profiles will be uniform throughout the day. However, the clusters and their respective memberships may be different during different intervals of the day. The thesis proposes the subspace philosophy of clustering for load profiling application. The available subspace clustering algorithms require large numbers of parameters to be specified as an input. It is difficult to find the optimal parameter setting for load profiling application. A novel plug and play type "Time Aligned Subspace Clustering" (TASC) clustering algorithm suitable for load profiling application (time series clustering) has been developed and proposed. .

सार

लोड प्रोफाइलिंग एवम रिसर्च को AEIC (एसोसिएशन ऑफ़ एडिसन इल्युमिनेटिंग कंपनीज़) द्वारा 1934 में टैरिफ डिज़ाइन, योजना, विस्तार आदि के लिए शुरू किया गया था। विध्युत उर्जा के प्रतिस्पर्धात्मक बाजार में विभिन्न संस्थाओं की प्रतिस्पर्धा और परस्पर विरोधी हितों के कारण अस्तित्व और विकास के लिए सभी विध्युत इकाईओं द्वारा इन अभ्यासों को अब आवश्यक उपकरण माना जा रहा है। । इन गतिविधियों में लोड सर्विंग संस्थाओं द्वारा लगातार प्राप्त किए गए डेटा की विशाल मात्रा का अधिग्रहण, प्रसंस्करण और विश्लेषण शामिल है। इस थीसिस में, लोड प्रोफाइलिंग प्रक्रिया को फिर से संशोधित किया गया है, एक बेहतर आंशिक क्लस्टरिंग एल्गोरिथ्म विकसित किया गया है और बेहतर उपभोक्ता व्यवहार विश्लेषण के लिए सबस्पेस क्लस्टरिंग के सिद्धांत पर आधारित **`टाइम अलाटेड सबस्पेस क्लस्टरिंग'** प्रस्तावित किया गया है। विभिन्न मांग पक्ष प्रबंधन के प्रयोगों के लिए दृष्टिकोण अधिक उपयुक्त होगा।

एक साहित्य सर्वेक्षण इंगित करता है कि लोड प्रोफाइलिंग के लिए उपभोक्ता अंत डेटा व्यापक रूप से पसंद किया जाता है। भारत जैसे कई विकासशील देश हैं जहां स्वचालित स्मार्ट मीटरिंग बुनियादी ढांचा ठीक नहीं है। साथ ही, व्यक्तिगत घरेलू मीटर आधारित डेटा पर आधारित क्लस्टरिंग और प्रोफाइलिंग के साथ कुछ नैतिक और सुरक्षा आधारित मुद्दे जुड़े हैं। डीरेग्यूलेशन के बाद, इलेक्ट्रिक पावर सिस्टम कई विध्युत उद्योग इकाईओं में विभाजित हो गया, जो बाजार के परिणाम पर विपरीत स्थिति रखते हैं। इसके अलावा, प्रत्येक विध्युत इकाई को अपनी स्थिति को इस तथ्य के कारण समायोजित करने की आवश्यकता होती है कि वास्तविक समय लोड-जनरेशन संतुलन को सिस्टम स्थिरता

के लिए पालन किया जाना चाहिए और गैर-अनुपालन के परिणामस्वरूप दंड हो सकता है। इस प्रकार आगामी मांग का सटीक ज्ञान अत्यंत आवश्यक है। लोड रूपरेखा और अनुसंधान उपकरण है जो पूर्व-आवश्यक जानकारी प्रदान करता है। ये विध्युत इकाई बाजार में अपनी रणनीतियों को तैयार करने के लिए सार्वजनिक रूप से उपलब्ध डेटा का उपयोग कर सकते हैं।

अध्ययन ने वितरण फीडर डेटा का प्रस्ताव और उपयोग किया है जो लोड प्रोफाइलिंग के लिए दुनिया में कई वितरण उपयोगिताओं द्वारा सार्वजनिक डोमेन में उपलब्ध कराए जाते है। इसके अलावा, चूंकि फीडर डेटा किसी विशेष क्षेत्र के संयुक्त खपत पैटर्न का प्रतिनिधित्व करता है, इसलिए प्रोफाइलिंग के परिणाम से उपयोगिता को उन क्षेत्रों की पहचान करने में भी मदद मिलेगी, तथा वितरण प्रणाली प्रबंधन (डीएसएम) लोड शिफ्टिंग, योजना, विस्तार आदि जैसी गतिविधियां अधिक फायदेमंद बनेंगी।

लोड प्रोफाइलिंग आकार की समानता के आधार पर समान समय अंतराल के लोड पेटर्न को समूहीकृत करने की एक प्रक्रिया है। इस प्रकार, लोड प्रोफाइलिंग के लिए समय श्रृंखला क्लस्टरिंग के लिए उपयुक्त किसी भी एल्गोरिथ्म को अपनाया जा सकता है। साहित्य में सुझाए गए विभिन्न लोड प्रोफाइलिंग तरीकों को अपनाया जा सकता है, लेकिन इनमें कई समूहों के बारे में प्रारंभिक अनुमान की आवश्यकता होती है। श्रेणीबद्ध क्लस्टरिंग को छोड़कर, अधिकांश एल्गोरिदम स्वाभाविक रूप से पुनरावृत्ति अनुकूलन कार्यक्रम हैं। एल्गोरिथ्म डेटा के स्थानीय संरचना पर जोर देने वाले विशिष्ट मानदंड फ़ंक्शन का अनुकूलन करके डेटा-सेट को असंतुष्ट समूहों के सेट में विघटित करने की कोशिश करता है। साहित्य में सभी प्रस्तावित विधियों में, **K-मींस** और

इसके वेरिएंट लोड प्रोफाइलिंग के लिए एक प्रमुख विकल्प के रूप में उभर रहे हैं।

K-मींस का परिणाम प्रारंभिक स्थितियों पर बहुत अधिक निर्भर है; यह उप-इष्टतम समाधानों के लिए अभिसरण की ओर जाता है। उपरोक्त सीमा को अलग-अलग यादृच्छिक आरंभ के साथ एल्गोरिथ्म को कई बार चलाकर और उपलब्ध समाधान से सर्वश्रेष्ठ परिणाम का चयन करके कुछ हद तक संबोधित किया जाता है। सभी समूहों के माध्य पैटर्न कुछ हद तक समान दिख सकते हैं, लेकिन क्लस्टर के सदस्य हमेशा अद्‌वितीय नहीं होते हैं और प्रत्येक रन के साथ बदलते हैं। यदि डेटा-सेट उच्च आयामीता के साथ बड़ा है, तो प्रभाव बहुत महत्वपूर्ण है। सभी आंशिक क्लस्टरिंग में कई समूहों के बारे में प्रारंभिक अनुमान की आवश्यकता होती है। लोड प्रोफाइलिंग के लिए कई समूहों के बारे में प्रारंभिक अनुमान शायद ही कभी उपलब्ध होता है।

इसके अलावा, पारंपरिक क्लस्टरिंग एल्गोरिदम प्रभावशीलता और दक्षता के संबंध में उच्च आयामी डेटा सेटों को अच्छी तरह से स्केल नहीं करते हैं क्योंकि यह उनके अंतर्निहित विरलता होती है। उच्च आयामी डेटा सेट से निपटने में पारंपरिक क्लस्टरिंग एल्गोरिदम की इस सीमा ने उप-समूह क्लस्टरिंग या अनुमानित क्लस्टरिंग की अवधारणा को प्रेरित किया है, जिसका लक्ष्य उनके संबंधित आयामों के साथ मूल डेटा स्थान के उप-स्थानों में समाविष्ठ क्लस्टरो को खोजना है। मौजूदा उप-समूह क्लस्टरिंग विधियों में बहुत अधिक पैरामीटर हैं और यह सर्वोत्तम पैरामीटर मान संयोजनों को खोजने और आवश्यकता के अनुसार क्लस्टर संख्या को परिभाषित करने के लिए है।

थीसिस ने अनुकूलित किया है और लोड प्रोफाइलिंग अनुप्रयोग के लिए प्लग एंड प्ले टूल विकसित किया है ***K-मींस*** एल्गोरिथ्म को कम करके और ***g-मींस*** एल्गोरिथ्म का उपयोग करके (यह ***K-मींस***:

एल्गोरिथ्म का एक प्रकार है जिसे समूहों की संख्या के बारे में पूर्व-अपेक्षित जानकारी की आवश्यकता नहीं है और सांख्यिकीय परीक्षण के आधार पर इनपुट डेटा से उनका मूल्यांकन कर लेता है)। प्रस्तावित कार्यप्रणाली इनपुट के रूप में निर्दिष्ट करने के लिए एक पैरामीटर, अर्थात महत्व स्तर (सिग्निफिकंस लेवेल) 'α' की आवश्यकता होती है। इस प्रकार, संख्या समूहों और उनके संबंधित प्रोफाइल को अंतिम रूप देने के लिए क्लस्टरिंग एल्गोरिदम के कई बार चलाने की आवश्यकता को समाप्त करता है।

पारंपरिक प्रोफाइलिंग मानती है कि लोड प्रोफाइल द्वारा दर्शाए गए उपभोक्ताओं के एक विशेष वर्ग का व्यवहार पूरे दिन एक समान रहेगा। हालांकि, एक दिन के अलग-अलग अंतराल के दौरान क्लस्टर और उनके संबंधित सदस्यता अलग-अलग हो सकती है। थीसिस लोड प्रोफाइलिंग आवेदन के लिए क्लस्टरिंग के उप-दर्शन का प्रस्ताव करता है। उपलब्ध सबस्पेस क्लस्टरिंग एल्गोरिदम को इनपुट के रूप में निर्दिष्ट करने के लिए बड़ी संख्या में मापदंडों की आवश्यकता होती है। लोड प्रोफाइलिंग एप्लिकेशन के लिए इष्टतम पैरामीटर सेटिंग खोजना मुश्किल है। लोड प्रोफाइलिंग (टाइम सीरीज क्लस्टरिंग) के लिए एक प्लग और प्ले सबस्पेस क्लस्टरिंग के सिद्धांत पर आधारित एप्लिकेशन **"टाइम एलाइनड सबस्पेस क्लस्टरिंग" (टीएएससी)** प्राणाली को विकसित और प्रस्तावित किया गया है।

Contents

List of Figures

List of Tables

Nomenclature

Greek Symbols

Γ	Hubert Statistics
Γ'	Modified Hubert Statistic
μ	Mean Pattern of Data-set P or Overall column mean of Profile matrix
μ_i	Overall Column Mean of in i^{th} subspace
ϕ	A vector of $1 \times d$ generated from vector $\mathbf{x}$

Other Symbols

A	$[A_1, A_2, ..A_K]$: Pre-defined or true set of K clusters of a given data-set
AC	Average Centre Change factor
B_S	Index of Best Solution vector
B_V	Best Solution Vector
C	$[C_1, C_2, ..C_K]$: set of K clusters generated by clustering algorithm
CH	Calinski-Harabasz Index
C_{i_j}	Set of j^{th} Cluster of d_i dimension of i^{th} subspace
Co	Contingency Matrix of two distinct solution of a clustering algorithm.

C_R	Cross over rate
∂	distance measure
D	Dunn's Index
d	Number of dimensions or time period of pattern
d_i	Number of dimensions in i^{th} subspace
d_{max}	maximum limit of d_i.
d_{min}	minimum limit of d_i.
DB	Davies-Bouldin Index
D_{dim}	Matrix showing combination of feasible size of different subspaces
d_{dim}	Vector showing feasible size of subspaces
E	Entropy
F	F measure
F_i	Number of feasible solution in i^{th} subspace
f_o	Value of Objective function
FM	Fowlkes & Mallows Index
$\mathbb{F}$	Number of feasible solution of all subspaces
R	Goodman-Kruskal Coefficient
G_{max}	Number of Generations or Iterations of DE
I_P	Initial Population vector indicating index of feasible solution vector $\mathbb{X}$ selecting target vectors
J	Jaccard Coefficient

K	Number of clusters or partition.
K_i	Number of clusters in i^{th} subspace
M	Mirkin Metric
m	Set of Centroid vectors of the clustering outcome
$m_A(i)$	Centroid of an i^{th} cluster of set A.
$m_C(i)$	Centroid of an i^{th} cluster of set C
MC	Membership Change factor
$m(i)$	Centroid of an i^{th} cluster.
m_{i_j}	Centroid j^{th} Cluster of i^{th} subspace
MS	Minkowski Score
M_V	Mutant Vectors
n	Number of patterns
n_i	Number of patterns in i^{th} cluster
n_{i_j}	Number of patterns in j^{th} cluster of i^{th} subspace
N_P	Population Size for DE
P	$[z_{ij}]_{n\times d}$ data-set or Input matrix of n patterns of d dimension or attributes
Pu	Purity
R	Rand Index
$RMSSTD$	Root-Mean-Square Standard Deviation
RS	R-squared Index

s	index and variable to evaluate number of subspaces
S_{max}	Maximum number of subspaces
S_F	Scaling Factor for mutation in DE
SHI	Silhouette Index
SI	Stability Index Index
SSB	Between the Cluster distance
SSW	Within the Cluster distance
T_S	Trial Vectors
T_V	Target Vectors
VD	Van Dongen Criterion
$\mathbb{X}$	Set of all such X_i
$\mathbf{x}$	$[x_i]_{1\times d}$ Row Vector of $\mathbb{X}$ feasible solution vector selected during execution of the algorithm
X_i	$[x_i]_{F_i\times d}$ Dimension Selection Vector
PI	Priority Index
XB	Xie-Beni Index
z_{ij}	j^{th} attribute or value of the i^{th} pattern

Acronyms / Abbreviations

AEIC	Association of Edison Illuminating Companies
BIRCH	Balanced Iterative Reducing and Clustering using Hierarchies
CACTUS	Clustering Categorical data Using Summaries

CLIQUE	Clustering in Quest
CURE	Clustering Using REpresentatives
CVIs	Cluster Validity Indices
CVMM	C-Vine Copula Mixture Model
DEA	Differential Evolutionary Algorithm
DENCLUE	Density Based Clustering
DISCO	Distribution Company
DOC	Density-based Optimal Projective Clustering
DSM	Demand Side Management
EA	Evolutionary Algorithm
EM	Expectation Maximization
FCM	Fuzzy C Means
FERC	Federal Energy Regulatory Energy Commission
FINDIT	Fast and Intelligent Subspace Clustering Algorithm
FIRES	Filter Refinement Subspace Clustering
FWT	Fast Wavelet Transform
GMM	Gaussian Mixture Model
HV	High Voltage
IEPAA	Information Entropy based Piecewise Aggregate Approximation
IPP	Independent Power Producer

ISO	Independent System Operator
LP	Load Profiling
LR	Load Research
LV	Low Voltage
MAFIA	Merging of Adaptive Finite Intervals Algorithm
MO	Market Operator
MPMKVVCL	Madhya Pradesh Madhya Kshetra Vidhyut Vitaran Company Limited
ORCLUS	Arbitrarily Oriented Projected Clustering
PCA	Principal Component Analysis
PROCLUS	Projected Clustering
PURPA	Public Utility Regulatory Policies Act
SAX	Symbolic Aggregate Approximation
SCADA	Supervisory Control And Data Acquisition
SOM	Self Organizing Maps
SVD	Single Value Decomposition
TASC	Time Aligned Subspace Clustering
TMLP	Typical Monthly Load Profiles
TRANSCO	Transmission Company