A Study of Multivariate Behavior and Anomaly Patterns:

Tensor Decomposition for Multiway Big Data

A Dissertation submitted for

The partial fulfillment of

Master of Engineering Research

By

Alina Rakhi Ajayan

University of Technology Sydney

New South Wales, Australia

2017

CERTIFICATE OF ORIGINAL AUTHORSHIP

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as part of the collaborative doctoral degree and fully acknowledged within the text.

I also certify that I have written the thesis. Any help that I have received in my research work and the preparation of the dissertation itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student:

Date:

ACKNOWLEDGEMENT

My most profound gratitude is expressed to the University of Technology Sydney, and the School of Electrical and Data Engineering, FEIT, for this valuable opportunity to pursue impactful research. My attempts and motivation were primarily fuelled by my parents Mrs. Aleykutty Robert, Mr. G. R. Ajayan and my dear brother Akhil Minu Ajayan. I wish to express my gratitude to Dr. Zenon Chaczko, my advisor and supervisor through this two-year academic period.

The real drive and support came from my colleagues and friends at UTS – Manisha, Augustin, Hayat, Daniel, Firas, Ammar, Maria, Alaa, Mahmoud, Mohammad, Asma, Deepak, and Saha had been the constant voices in my head, egging me onward towards my goals and pursue my dreams. The entire Centre for Health Technologies Social group, starting with Dr. Adrian Bishop and Prof. Gyorgy Hutvagner had always been good mentors and advisors.

My candidature assessment panel members – the Hon'ble Chair and my Examiners, made my progress take a giant leap from the state of an in-depth Literature Review to a narrowed-down, streamlined research project. I hence extend my heartfelt gratitude to Dr. Daniel Franklin, Prof. Robin Braun, Dr. Negin Shariati Moghadam and Dr. Yi Cao, who brought me forward to the completion of this work.

Finally, but most importantly, the last few months of perseverance and dedication was hard to fight for and succeed. I could only push myself through the harsh write-up and implementation phase of research by the guidance of A/Prof. Raghav Menon and Michal Skowron.

Thanking you all from the bottom of my heart, for encouraging me to do my best, and to be at my best.

ABSTRACT

A vast majority of the today's information haul is through Cyber-Physical Systems (CPS). They represent the confluence of extensive data sets, tight time-constraints, latency issues and heterogeneous components. CPS architectures demand newer Big Data processing approaches. Typical applications span from the Internet-of-Things, across the World Wide Web to Smart Cities and Intelligent machines. A standard heterogeneous CPS installation, the Smart Energy Grid, is observed and the logistics are analyzed. The Smart Grid domain is weighed down by lack of unifying framework and systemic intelligence for autonomic management. Preliminary studies of the field under investigation shows how processing of Real-Time data, communication and control signaling is vital. Purely autonomic system governance is proven to be different from the contemporary definition. It takes the form of Interoperability (achieved through automation) instead of elemental Integration. That means autonomic (smart) management requires all elements to have fully controllable behavior.

This dissertation teste the hypothesis of applying Tensor decompositions and Factorizations - a momentum-gaining arithmetic tool - to this problem. The aim is to validate the prospects of higher order Anomaly Pattern Processing to capture intelligence along multiple modes of data flow. Tensorial Data representation captures information flows in Big Data, while Multivariate Anomaly Detection performs tracking of the time-series behavioral changes. Together, they implement Autonomic management in CPS super-systems. Uniqueness of this approach is highlighted by the novel multi-modal data flow imaging and models. Requirements of traditional anomalous event definition and cataloging in Data streams are removed. Tensor algebra is then studied for the scope of implementation concerning features, significance, and interpretation in terms of multi-modal data. Standard Decomposition rules and their derivatives, literature analysis on contemporary applications of Tensor algebra, and its scope on prominent real-world data processing problems are studied. Finally, the decomposition tool for Multi-way analysis is inferred, and proposed methodology is derived.

The Smart Grid Smart City Project commissioned by the Australian government is chosen as the data source investigated. The need for exhaustive examination of such repositories in the CPS Anomaly Detection context is also highlighted. Experimentation is done by applying Tensor Decomposition on the data set after normalization and pre-processing. Details of those phases, as well as the choice of coding platforms, the design of experimental frameworks, timelines estimated, and testing operations, are included in this work. The outcomes are the defined patterns extracted and their analysis-interpretation defended by proofs from actual events of the Project Trial phase.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 : Introduction

1.1 OUTLINE

Chapter 1 is an overview of the research work undertaken within the scope of this Master of Engineering Research Dissertation. The choice of a significant narrow problem domain is defended by extensive literature to support the vision. Contemporary works that motivate the devised solution are presented methodically. Research hypothesis and core postulates are then introduced, relevant to Masters level investigation and study, followed by future directions towards a Doctorate in Philosophy. The structure of this chapter is the framework for later chapters in this document as well.

1.2 DOMAIN OF RESEARCH

1.2.1 Background

Most technological implements today take the form of a System-of-Systems – a collection of discrete, heterogeneous, interdependent systems unified by decentralized homogeneous control. They represent hierarchical, heterarchical and composite functional relationships between the different sources of information channels, i.e., the various components and their sub-elements. So, qualitative or quantitative modeling of the computational architecture is always of a higher order. Typical examples of such intricate structures are Cyber-Physical Systems or CPS.

1.2.2 Context of Study

Each varying index of information (component, sub-system or signals), may be termed modes, and relationships between 2 or more such modes correspond to a system order. Current CPS technologies generate Big Data in higher orders and dimensions (size) with trends and associations across all modes. This issue deems state-of-the-art information processing systems invalid. Such behavioral patterns also contribute to autonomic intelligence in the system. When represented the right way, autonomicity makes any system fully self-managing.

It is thus crucial to model the autonomicity of CPSs regarding Information flows – not just the systems overall, but also individual elements and components. Complex interactions within a Cyber-Physical Ecosystem is captured this way. Patterns characterized maybe interpreted as snapshots along multiple modes or of higher indices of references. They no longer hold single numerical values of observations – instead they are symbolized by large sets of numerical variables, aggregated to visualize variation patterns. So, if successfully identified, the emergent patterns and trends will quickly pinpoint unexpected events, out-of-order behaviors. Those two incidents may be defined as

Anomalous events. When they pertain to more than one varying data streams, they take the form of Anomalous patterns.

1.2.3 Problem Domain

Cyber-Physical Systems requires dealing with of various issues and concerns. Designing a unifying architecture is the most important, as it addresses all critical matters relevant to green communications, computing, and systems. In this perspective, extracting and isolating multiway patterns to represent the subscale relationship of different data streams, i.e., the sources, is necessary. They, in turn, form the rule-base for defining autonomic intelligence of a unifying architecture for the super-system (system-of-systems).

Many existing analytics tools support ongoing periodic analysis of Time Series, to deliver Autonomic Management. They are inspected for shortcomings – which may be summarized as being limited by a single variable, dual variable or double variable-combination used as metrics to estimate behavioral patterns.

Multivariate Dynamic Pattern processing is suggested as the tool to overcome the limitations. Regarding Pattern recognition based on template matching or feature detection types, an improvement to the state-of-the-art concepts is put forward.

This work proposes enhancing Pattern Processing not limited to pre-defined or well-known feature set dictionary. The suggested Proof-of-concept model is based on dynamically varying Pattern 'Images' in multivariate Big Data, based on space-time-index-defined information streams, and is designed to capture mappings of every possible system and component behavior states.

1.2.4 Scope of Research

Another perspective of relevance to this work is the importance of Data characterization. As complex the recognized research problem domain is, the real-world multiway high-dimensional data substrate is equally composite. Hence a comprehensive analysis is performed on the chosen research setup that has generated the data sets used for analysis. Inferences are made about their behavior, trends over time and space and highlight operational points.

The project thus explores *"Multivariate Data Processing for Anomaly Detection through Behavioral and Anomaly Pattern Processing"* in Cyber-Physical systems. This dissertation was written to review and debate conceptual, material and theoretical frameworks towards Higher Order Big Data emerging from such implements and sift out the deepest level of intelligence, to manipulate them into executing autonomic management.

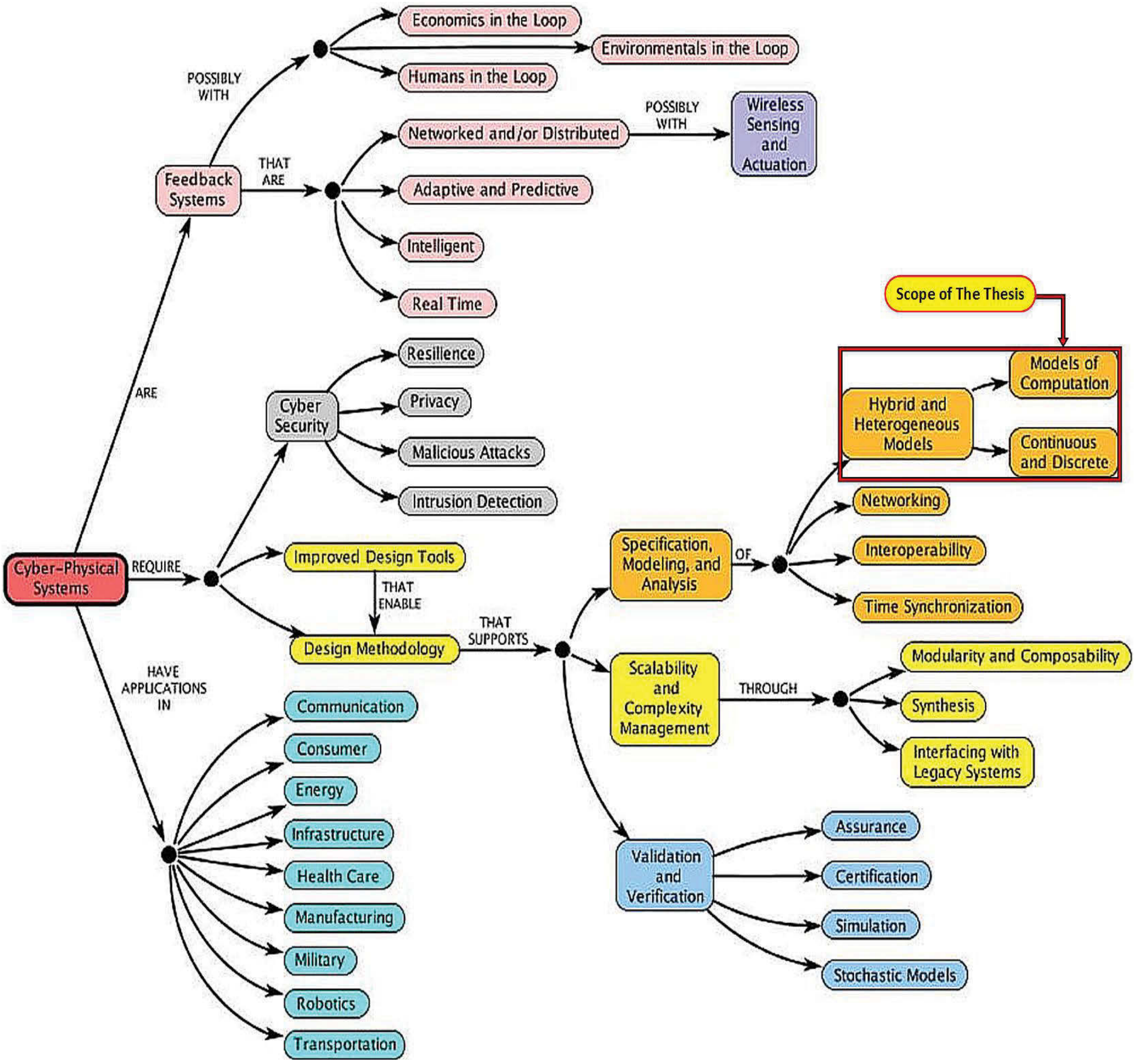


Figure 1.1 Overview of the concept-realm of Cyber-Physical Systems, outlining scope of this research work

The suitability of the chosen research topic in the vast problem scenario of CPS advancements and research objectives, along with the need for a solution devised, is mapped in Figure 1.1 [1]. The *Scope of Thesis* outlines the feasibility and applicability of our research.

1.2.5 Important Terminology

Interdependency: the qualitative correlation between any two variables or functions. They may be active or latent.

Active Interdependency: apparent or quantifiable Input-output, input-system state, or system state-output relations. They form the basis of any architecture, operation or performance study.

Latent Interdependency: those dependencies which may surpass mathematical or pre-defined constraints. They are hence only qualified (characterized), not measured.

Multivariate: any information flow affected by more than one variable, function or index.

Subscale: It refers to the scale of observation within or smaller than the usual input-output, input-state, state-output relationships.

Dimension: Size of the data structure along any particular index or axis.

1.3 RESEARCH QUESTION and HYPOTHESIS FORMULATION

The well-framed Literature review has refined the Gap in the state-of-the-art, objectives and expected advantages. The upcoming subsections enclose these details. They were the benchmarks for devising the proposed solution and implementation methodology.

1.3.1 Research Gap

Modelling Behavioral and Anomalous patterns of multivariate Big Data streams, satisfy the criteria that are listed below. A systematic literature review filtered out these aspects.

- Need to elevate the processing dimension
- Substantial reduction of operational delay, resource utilization, latency, and information volume.
- Modelling states, state transitions and IO conditions based on space, time, or index time windows.
- Distributed data acquisition and cataloging with subscale intelligence.
- Support for scalable interactive interoperability among heterogeneous devices
- Quantify and qualify heterogeneous data streams, uninhibited by mathematical constraints at the same space-time domain.

1.3.2 Hypothesis and Proposed Solution

The research gap led to the framing of the hypothesis, as stated below:

"Model complex dependencies among heterogeneous Big Data streams to validate if it is possible to improve synergistic Big Data management capabilities within a CPS ecosystem."

The developed solution is adapted Tensor Algebra into High-dimensional Data Processing. Tensor Decompositions and Factorizations is proposed as the arithmetic tools to extract the above-explained *Latent Patterns*.

Tensor Decomposition is adapted for High-dimensional Data Processing in a typical CPS environment, for dynamic, multi-modal, multivariate Big Data flows.
They are processed and characterized based on dynamically defined Latent features of Association and Dissociation. Thus system complexities are retained, eliminating dimensionality conversions.

1.3.3 Functional Advantages

`The hypothesis investigates the following statements, which are perceived to be merits of the proposed architectural design.

- For any complex core system, assembling the data streams along multiple scopes will help determine inter-modal dependencies.
- Their quantification will measure the characteristic behavior and response of the CPS setup.
- Total Data crunching time and resources of a complex autonomic system decreases considerably.
- Full processing cycle of a CPS delaying operation mode identification will be bypassed, by prediction through data-pattern matching.
- Recording such patterns can help catalog the overall system characteristics through operational modes represented by abstracted pattern models.

1.3.4 Research Objectives

- Develop customized feature set for the research problem from the data compilation; complicated manual processing for feature extraction and constraint by existing definitions avoided.
- Increased order and dimensionality of the annotated/embedded Latent feature set, qualified through Tensor decomposition models.
- Eliminate low predictive accuracy and false positive rates due to feature redundancy, unbalanced data structure and small sample sizes, through the high dimensions of dataset input.
- Allow detection newer variants of the annotated markers as they emerge in the data flows, and prevent misinterpretations; thereby remove learning bias.

1.3.5 General Study Goals

The concepts of the selected research problem suggested solution and validations also align with the overall development goals of the scientific community: Green computing and communications; designs, algorithms, and math to handle sustainability issues, ICT for sustainable green objectives, Software-defined approaches and systems, etc.

1.4 LEARNING OUTCOMES DURING THE STUDY PERIOD

1.4.1 Subject-Oriented

The aim of this research study (Master of Engineering Research) is to build upon a focused and persuasive problem under investigation and hypothesis, with validation in favor of the stated. This study encompasses the following technical aspects learned:

- Establish a theoretical base and critical evaluation of practical issues faced by heterogeneous *Cyber-physical systems* (Smart Grids, Smart Home networks, etc.).
- Highlight the existence of *Higher-Order High-dimensional Behavior and Anomaly Patterns* in *Big Data* generated within such environments.
- Identify, evaluate and infer the need for reduction analysis *learning models* of Big Data.
- Design a learning model for *Multimodal Latent patterns* using *Tensor Decompositions*.
- Implement the model in a Big Data Analytics tool with significant interpretations and scope for future systemic-intelligence deriving applications.

1.4.2 Practice Induced (Apart from technical expertise)

Aside from the subjective knowledge and streamlined research skills, the following are the learning outcomes of this project, with regards to the general domain of study.

- Demonstrate an understanding of contemporary critical thinking across Big Data Engineering, and modern methodologies in learning and processing real-time (RT) data.
- Acquire professional practice and skill sets in methods for learning and RT Big Data processing.
- Gather advanced practice in such coding platforms and development tools.
- Develop Data characterization and Analysis algorithm for real-world multiway big data streams.
- Appraise and reflect on the principles of ethical practice-led independent research methods.
- Exemplify ideas and knowledge through the preparation of a coherent and well-structured dissertation in a professional context.

1.5 THESIS ORGANIZATION

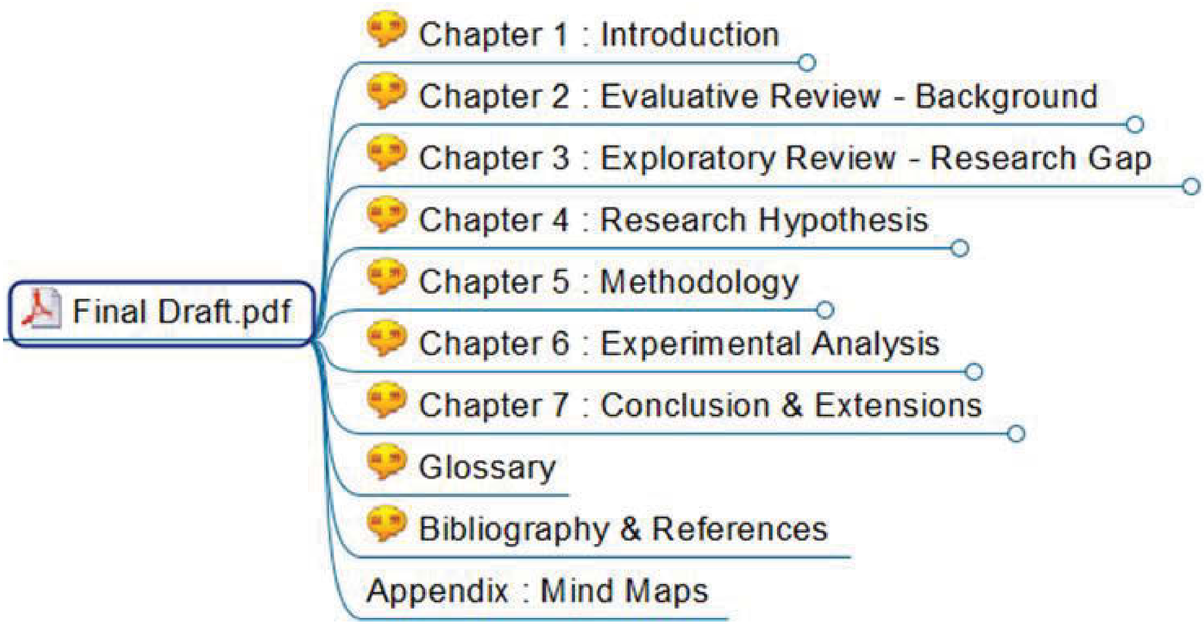


Figure 1.2 Mindmap outline of the Dissertation

Following the Introduction Chapter (current section), the dissertation proceeds to showcase various arguments regarding problem domain at hand, navigating through different stages of Literature Review (Chapter 2 and 3) – Evaluation of the Research area and Problem scenario, and later Exploration of the scope for possible solutions and contributions. These points condense into the Hypothesis and Proposed Solution (Chapter 4). The next section (Chapter 5) performs Methodology Analysis for the standardized Big Data processing scheme for the Research problem under scrutiny, which includes steps like understanding the Big Data set under study and the input parameters, implementing the derived solution In a data engineering solution. The testing and result interpretation is done on sample datasets (Chapter 6). The final chapter will conclude the study (Chapter 7). Each chapter as a brief Outline and Summary section at the beginning and end, sketching the presentation of facts and ideas. Mind maps are also included in every chapter to depict the flow of logic.

1.6 CONTRIBUTION IN PUBLICATIONS

1. A. Rakhi and Z. Chaczko, *"An Autonomic Tensor-Pattern Processor for Multivariate BigData Anomalies in Smart Grids,"* in IEEE Transactions on Smart Grid, Transactions September 2017 **–** *Submitted.*
2. A. Rakhi and Z. Chaczko, "*A Study on Need for Multimodal High-dimensional Tensor Anomaly Detection in CPS Architectures*," IEEE Communications Surveys and Tutorials, Survey and Position vol. 4, 2017 **–** *Submitted.*
3. A. Rakhi and Z. Chaczko, *"Tensor Decomposition in Multimodal BigData: Studying Multiway Behavioral Patterns,"* 5th Asia Pacific International Conference on Computer Aided System Engineering (APCASE 2017), Guilin, ISBN 978-0-9872369-4-4, (e-ISSN: 2300-1933), ePublication in Springer's LNCS – *Accepted*
4. F. Al-Doghman, Z. Chaczko, A. R. Ajayan, and R. Klempous, "*A Review of Fog Computing Technology,"* 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2016). Budapest, Hungary, 09 Oct 2016 - 12 Oct 2016, 'Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2016),' IEEE, Piscataway, USA. 1525-1530.

1.7 CONTRIBUTION IN RESEARCH

The overall contribution of research conducted as part of Masters by Research degree is the modeling of latent-patterns defined in this work using Tucker decompositions. They are then extracted and interpreted, in compliance with reference pattern-trends (actual events occurred) identified from the data source under consideration (SGSC Project Testing phase). The abstracted items are:

- Introduce the concept of Active and Latent dependencies: demonstrate how apparent mathematic relations among system variables only partially characterizes the overall operation.
- Identify possible types of latent independencies between 2 or more information streams in CPS.
- Mathematically qualify those dependencies, for practical Knowledge Discovery, using Tensor Decompositions.
- Characterize them with a multiway high-dimensional arithmetic framework for efficient Knowledge Representation, as a Tensor-based Core Tensor-Factor Matrix (Matrices) system.
- Interpret patterns obtained as dynamically varying Pattern 'images' in multivariate Big Data, based on space-time-index-defined multivariate data structures, relevant to the data source domain
- Examine the scope of applying Multivariate Pattern extraction based system models for Autonomic Intelligence in CPS.

1.8 CHAPTER SUMMARY

A Proof of concept model for connected data-signal-quantity flow analysis for inherently intelligent semi-autonomic systems. The need for full systemic intelligence requires full-autonomicity stemming from proactive and reactive management. This thesis works towards building a new perspective in this problem scenario. The solution devised is to learn predictive features directly from a data feed, making them highly suitable for higher order high-dimensional data processing. Tools in use are the most sought-after solutions for complex Big Data computations of higher orders and dimensions off-lately.

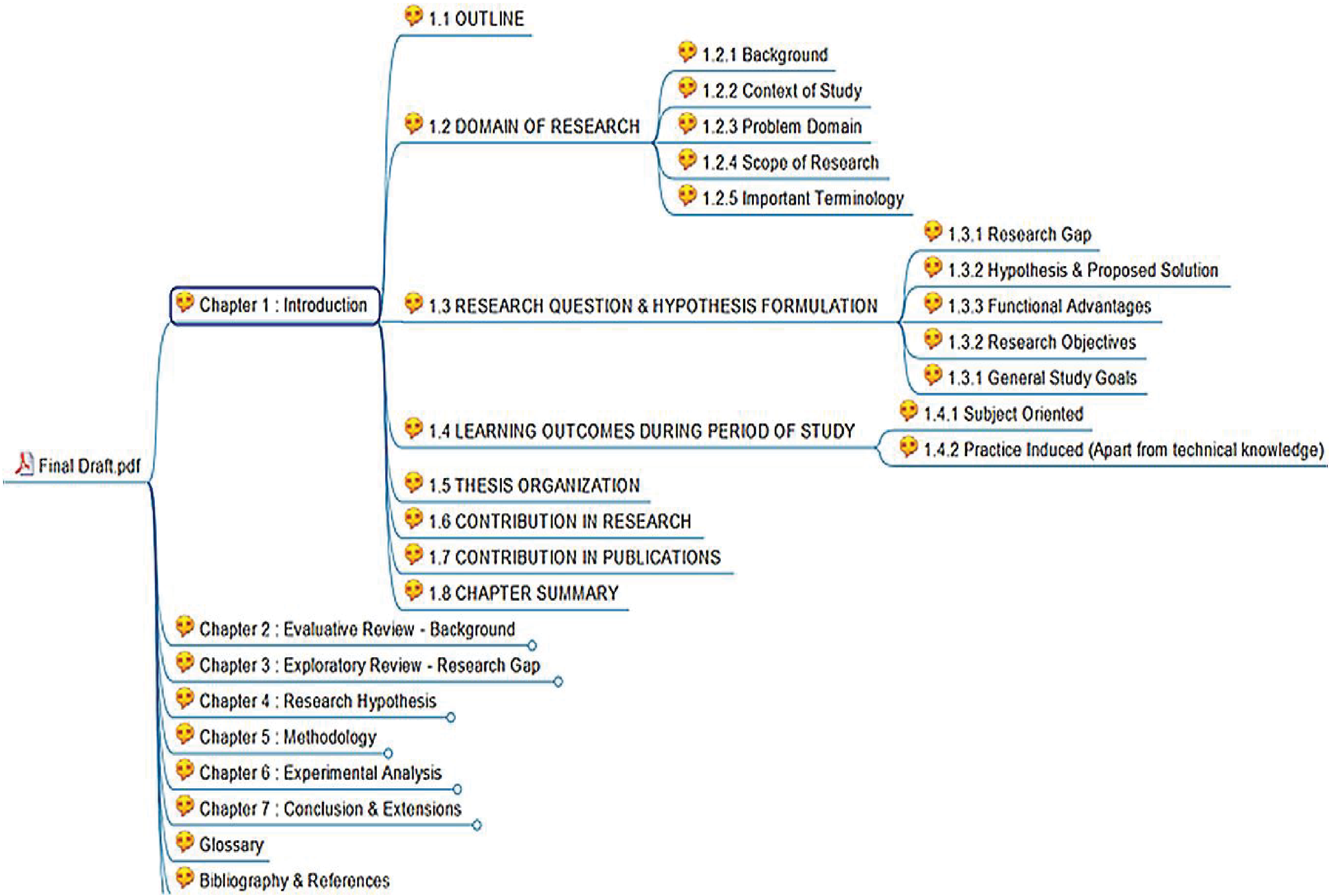


Figure 1.3 Mindmap outline of Chapter 1: Introduction

Chapter 2 : Evaluative Review - Background

2.1 OUTLINE

This chapter summarizes the relevant research literature to this work. A structured, systematic evaluative literature review is composed, through a converging to a pertinent Problem domain. This phase discusses knowledge areas in focus, their coverage and contribution, pinpointing questions requiring immediate attention.

The staged framework of the literature study is as shown in Table 2.1. In all the three stages, primary sources of information constitute the main text, while secondary and tertiary literature are listed.

2.2 Domain of Study

The study purview of this project is complex Cyber-Physical systems that lead to the generation of heterogeneous higher-order data streams deterring real-time autonomic operation.

2.2.1 Background: Cyber-Physical Systems

A System-of-Systems or “SoS” is the cohort of numerous, discrete, independent systems unified under decentralized control to form a larger complex super-system. The computational architecture is always of a higher order since they represent hierarchical, heterarchical and composite functional relationships. Individually, each unit is interacting, interrelated and interdependent.

Table 2.1 Framework adopted for Systematic Structured Review of Relevant Literature

Literature Review Stage framework	Evaluative Review - Chapter 2	Exploratory Review - Chapter 3	Instrumental Review - Chapter 4 & 5
Block 1	Create research context		
Block 2	Study Background concepts		
Block 3	Analyze existing information	Distill links between concepts & state-of-the-art	Review Methodologies
Block 4	Identify current implementations	Survey collected literature for Gaps	List out tools for possible solutions
Block 5	Explore similar research interests	Define Problem domain: problems with existing work	Match results with expected outcomes
Block 6	Study major seminal works	Theorize ideas, merits & demerits & improvement scope	Study requirements, compare & contrast
Block 7		Research problem statement	Match results with problem statement
Block 8		Define Research Objectives	Claims to defend choice
Block 9		frame Hypothesis	Proposed methodology & solution

The distinguishing SoS feature is the ability of each component for standalone functioning, unlike generic complex, hybrid or networked system groups. In such complex systems, a holistic analysis of the information content they generate reveals the interoperation of constituent parts - over time, space and function, within the larger SoS context [1]. Hence abstraction of individual component-data reflects on solid insight into the operation of these super-systems.

A “Cyber-Physical-System” (CPS) is a ubiquitous example of such an organization, dedicated to a particular set of applications and underlying technology. They are the most widely applied complex system structures in modern telecommunication and engineering solutions, and represent the future face of information and communication technology. They carry Multimodal information, which is simultaneous data from multiple sources. The “flat-world” view of most data processing techniques doesn’t suit such intricate system modeling as this [2].

2.2.2 Typical CPS Installations

Applications of CPS are limitless in terms of backbone architecture. As an end-user facility, its most significant manifestation is the Internet of Things or IoT. The highly abstract definition of the Internet of Things (IoT) can be phrased as the hyper-scale, hyper-complex cyber-physical system. A wide range of domains incorporates CP architecture like Real-Time Multivariate Information and Communication systems, Data mining and Information retrieval models, etc. 22, 114

2.2.3 Computational Systemic Intelligence in CPS

Optimizing a CPS organization for best performance is achieved by enhancing “Systemic intelligence” [3]. Systemic Intelligence is merely the ‘intelligence within a system.' Computational intelligence in its purest form is the autonomic adaptability of systems Situations that come out of this “envelope of adaptability” can lead to complications. It will even result in applying emergency fail-safe protocols to avoid complete system failure. The existing self-adaptation approaches still rely on a pre-defined set of strategies not flexible enough to deal with those out-of-norm situations adequately. Tracking changes in systemic behavioral trends, without being delimited by physical constraints need to be applied.

In any system, every information signal represents data on behavioral modularity among separate components and the overall system behavior. So, multivariate data streams combine multiple levels of such interaction features or modes. This attribute is called deep intelligence in a system. However, the numerous facets of the gathered data, contributing to analytic intelligence presents tests of modeling both the data as well as the analytic module. In this scenario, often misplaced are the measures aimed at inducing smartness in the physical or logical CP infrastructure.

The dynamic system characteristics are hardly observable due to disadvantages of traditional firmware design-based implementation. Capturing dynamicity maybe explained on the basis of the terminology described hereafter. While reviewing in detail the computer-based systems-design paradigms, significant facts were highlighted by the comparative study Automated vs. Autonomous vs. Autonomic systems.

a. Autonomous Systems and Autonomy

Traditionally, inventors and researchers run towards autonomy inducing solutions and create autonomous systems and platforms [4, 5]. Autonomous processes, only work on emulation of human processes instead of their replacement [6]. They represent the constant adaptation to changing parameters as a result of time-base or spatial changes. Thus they can only either learn or train.

b. Autonomic Systems and Autonomicity

Goal-directed reflexes embody Autonomic Power Systems, that is adaptive self-management and automated self-* characteristics (* - organization, management, optimization, etc.). They are the cornerstones of intelligent systems.

Autonomic network management is generic and streamlined, achieved by platform-agnostic middleware and information retrieval in a dynamically composed environment. It represents "free will of actions" or spontaneous-governance and so the success likelihood of administrative and technical staff focusing on deeper issues without intervention is high. It is the autonomic design paradigm that embodies the introduction of intelligent ability for autonomic updating of new policies/rules to achieve desirable performance onto devices (programmability).

Pure autonomicity or self-governance is way off from current levels of autonomy, or automated operation. It allows quantifying or characterization of system operation without mathematical or numerical constraints inherent in heterogeneous super-systems. Processing of real-time status and control information to quantify system operation beyond mathematical limitations is vital for the progress of CPS applications.

Such an approach involves expressing them as complex functions of parameters that are arithmetically and analytically heterogeneous. Contextual Awareness in Cyber-Physical Systems is the deeper level of information regarding relations, associations or interactions. It has been applied for quite a few integrative solutions [7-9]. Therefore, the currently lacking *wide-ranging or complete vision,* when solved induces Autonomicity in CPS architectures. Autonomic Intelligence, therefore, manifests as spatiotemporal dependency patterns in data streams.

2.2.4 Technology Scenario 1: The Internet of Things

The Internet-of-Things have emerged as a powerful technology hybridizing numerous innovations. It aims at creating more and more consumer-centric deliverables. The British visionary Kevin Ashton coined the term in 1999. The comprehensive definition is that IoT is as follows: "A network connecting anything with the internet, communicating and exchanging information, using RFID tags, GPS, infrared sensors, laser scanners and other sensing units, supported by the backend protocol, in order to achieve intelligent identification, positioning, tracking, monitoring and management" [10].

The IoT is a complex cyber-physical system (CPS) integrating devices of sensing, identification, processing, communication, networking and other embedded intelligent capabilities [11]. Today the Internet-of-things is a vast grappling web linking every digital entity on the Internet. Visualized it appears as a ubiquitous network of collaborated monitoring systems, actuator units, command modules, scheduler, and databank, centrally managed by an intelligent processing unit. Thus IoT becomes the largest in scale and depth among CPS installations. Deep learning and AI, scalable cloud computing, graphical processing units and High-performance computing solutions automate IoT applications at an unprecedented level.

The existing National Electricity Grid has been undergoing a comprehensive transformation through the application of ICT based coordination and control for quite a while now. The innovative paradigm of Internet-of-Things connects millions of smart devices through the internet. It establishes their self-organizing and intelligent functioning, revolutionizing the approach towards 'electric power' related issues. The union of Smart Grids with advanced machine intelligence induced by the Internet-of-Things will create a consistently adapting, fully autonomous Smart Energy Systems.

IoT serves the Smart Grid hub as an intelligent network based on network technology, database management, communication systems, and so on. Applying IoT will develop stability and security, as well as protect the environment against the generic power grid to enhance sustainable development. Moreover, the power system intelligence levels will be upgraded.

2.2.5 Technology Scenario 2: The Smart Energy Grid

Electric power is the lifeline of every entity, living and non-living, across the world. Rising populations and modernization accelerate the rapid decline of non-renewable energy sources (coal, crude oil, etc.). It also leads to pollution and global warming. The condition highlights the need for a sustainable green energy solution, giving birth to the concept of Smart Grids. Renewable energy sources like solar power, wind, biogas, tidal harnesses, etc. are merged with the modern power grid.

Other elements include smart meters, intelligent electronic devices (smart equipment), automated power flow, time-of-use rates, and much more. The structure is controlled by automated interconnected systems and driven by economic and commercial decision-making modules. It works on the agenda of economic optimization of energy flow using advanced technology. Electric Power sector thus emerges as a Cyber-Physical System requiring Autonomic control and operation. Such systems point at installing the organization with global self-adaptive management and distributed self-management.

The conceptual SG-CPS paradigm may be divided into the following sub-systems from supply-side, demand-side and storage options:

- Substation Automation
- Advanced Metering Infrastructure (AMI)
- Renewable Generator Integration
- Fault Detection, Isolation and Restoration
- Power Network Monitoring or Volt/VAr Optimization (VVO)
- Demand - Side Management and Demand Response (DR)
- Additional Grid Monitoring and Control Technologies

An irrefutable fact is that each of these entities needs to interact within their boundaries as well as across boundaries. The entities that work in both those conditions are Data, Signals and Energy, highlighting why heterogeneous Data Processing is essential.

2.2.6 Main Challenges in the CPS Domain

CPS mechanics, as opposed to generic signal-system setup, has multiple input-output transactions. As such, the main issues when it comes to practical designs of CPS are *heterogeneity, mass data processing, scale and interdependency*. Figure 1.1 had outlined the significant obstacles and characteristics in this domain. Top priority challenges that need to be addressed in this context are:

- Heterogeneous system design preventing platform unification
- Volume of Big Data collected
- Incompatible design specifications

Most complex CPS environments create data structures in higher-orders or modes. Trials, Task definitions, System conditions, Subjects, and groups can occupy the Modes, along with the fundamentals of space, time, and frequency. Sample scenarios are Advanced Multimodal Image-Guided Operating, Human-Machine Interaction, Hyperspectral Imaging, Brain Signal Processing, etc. Parallel data streams thus generated give rise to the Big Data phenomenon. The CPS characteristic of *heterogeneity* directly adds to challenge. Cheap acquisition and storage of such colossal data collection is vital. Hence, the most stringent challenge is processing their Time-Series Big Data streams.

Non-Stationarity is another major obstacle. It refers to the Data streams statistics (mean, variance and frequency) changing over time. Non Stationarity renders Spectral domain analysis more significant. That is how Behavioral Patterns gain importance. They represent the changes over time and phase which is the base for system dynamicity. Hence, the dynamic nature of complex CPS also places an extra constraint of adaptability to changing the environment.

Besides, pushing for faster runtimes, response times and iterations over a smaller time window (as in distributed and decentralized communication networks or hierarchical control systems), will result in the system hitting the limits concerning processor speed, hardware efficiencies, and memory and device communication latencies**.** It is not different than attempts to increase the buffer read or write rate in an 8-bit processor, to match a higher bandwidth, than merely opting for a 16-bit processor**.** Therefore, widening the data processing capabilities will help obtain solutions along with more than two variables during the same time window.

Finally, with the stochastic nature of the system processes, it's only suitable to represent changes and change patterns in the system behavior through multivariate patterns, rather than using mathematical theory or and a wide variety of movement axes.

Hence, a pathway towards extending the depth of adaptability of a processing unit is to derive new operation-rules in runtime to reflect changes in the functioning environment. CPS-generated data sets are growing at an exponential pace matched by hardware volume reduction, require more modern perspectives to examine all those constraints. Bypassing the above-listed restrictions is crucial for more robust data processing platforms.

The discussion and analysis of CPS and heterogeneous complex data streams performed till now will now be conceptually demonstrated as a case study of the Smart Grid environment.

2.3 PILOT STUDY: SMART ENERGY GRIDS

The need to provide a holistic modeling framework for defining and managing CPG, as part of a broader disruption-informed modeling and analysis, has motivated this research.

All Electric power is the lifeline to today's society. The rapid decline of fossil fuels and rising global population push for smarter solutions in the power industry. Legacy electric power systems are made up of massive volumes of grid devices laid out across vast geographic areas. With the recent advancements in Information & Communication Technology, like every other traditional infrastructure, Power Grids also have been transforming. The need to adapt changing technologies, as well as the swift-growing electric power demand is immeasurable. Following points distill out the primary motivations for a drastic electricity system transformation:

- Growing need to decarbonize the electric power sector towards environmental conservation (Grantham Institute for Climate Change and Environment).
- Multi-billion-dollar boom of global solar PV industry led to high targets of emission reduction [12].
- Fast-evolving tech and drastically reducing prices of batteries and electric storage solutions [13].
- Improvement in energy efficiency, across the National Energy Market.

It is apparent in the following statement about Smart Energy Grids, one of the first manifestations of Cyber-Physical infrastructure.

"The implicit assumption in smart grids is that energy may be generated and used anywhere on the grid, the behavior of both generators and users is much more variable and less predictable and that both will continuously react to the other." – Ibid and Distributed Generation and Distributed Storage SGSC Technical Compendium, Ausgrid.

Today, generic power systems have adopted a new face – a super-system or system-of-systems model named Smart Energy Grids, popularly the Smart Grids (SG). Smart Grids are thus the most sought-after technology. It is a hot research topic in Public Services infrastructure domain, with the scope for access to green, sustainable renewable energy, balancing resource consumption and hence economize supply. Being the primary driver of global population rise, energy crisis and environmental concerns, they are vital to our growth and future development. The drive for adopting Smart Energy Grids for our pilot evaluation is the lacking uniform integration, gap in various functionalities and overall layout. The SG Flexible of operation enabled by technology in combination with innovative SG solutions and applications is the motto.

Many technologically prominent countries have already initiated segmented trials and testing of this paradigm. In the Australian context, based on experimental systems, prospects savings margin, standardization attempts and successful deployment strategies are discussed and in [14, 15]. European countries like Sweden, Croatia, and Netherlands, as well as Singapore, the US, China and many more are contending hard to roll out full refurbished Power systems [16-18].

In the Australian context, the proportion of publications on Smart Grid technologies is found to be the highest 33% on Efficiency Improvement (EI), followed by a 17% related to Electric Vehicles

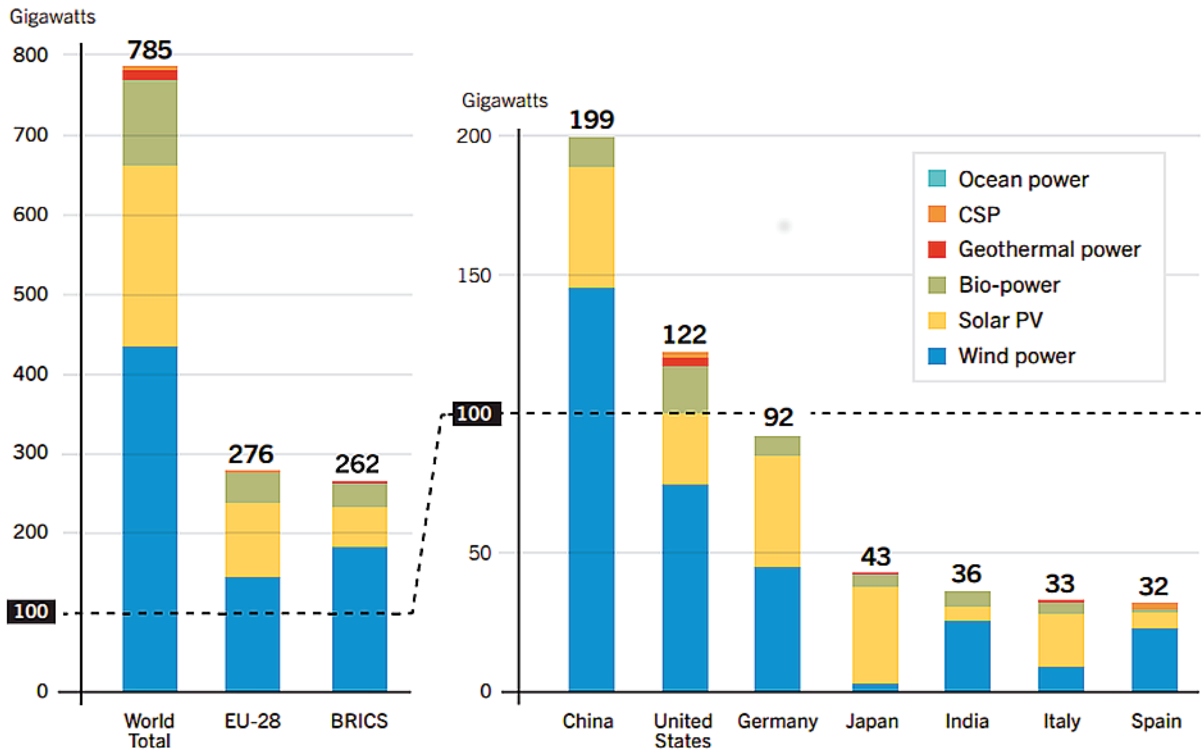


Figure 2.1 Global Renewable Power Capacities by the end of 2015; BRICS countries which are Brazil, the Russian Federation, India, China and South Africa, courtesy #/www.weforum.org

(EV) and AMI or Advanced Metering Infrastructure, but much lesser in other technologies such as around 8% for Distributed Generation (DG), Distribution Automation (DA), Volt-VAr Management models and Self-Healing, while the lesser noted topics include Demand Response (DR), Cyber Security (CS), Energy Storage and User Friendliness.

Research and development in this area have progressed in many directions. However, they all take different perspectives like integrated hardware, smart management using Big Data, Smart Communication, and Metering, etc. Most state-of-the-art systems work along the generic system configurations. The space-time-based integrated behavior is, however, another question.

The pilot study shows how a CPS environment, as is the case of the Smart Grid scenario of tomorrow, will never be devoid of Big Data. A sample system environment as a reference for the research problem and answers will help frame a more precise description. It focuses on a vital development required in these systems - establishing a fully interconnected homogeneous layout for the diverse CPS subsections, establishing convertibility and interaction among them is necessary.

2.3.1 Organization: The Legacy Grid

Physically, the Power grid consists of heterogeneous modules combined like energy sources, loads, storage units, and control systems operating as an interconnected super-system, run by market-related decision-making modules and signaling. A rough categorizing generates three main sectors:

- Power Generation (power plants and production units),
- Power Transmission and Distribution (physical grid infrastructure framework), and
- Power Consumption (primary, secondary and tertiary consumer sectors) systems.

From a logical perspective, overall Energy systems are constitue superposition of three planes of operation:

- Transmission network: comprised of Electrical conductor lines, loads and sources and substation networks transferring energy from point to another.
- Communication and Data Layer: energy communication networks (ECNs) transmitting data and information about energy, and harvesting useful information on the grid components.
- Control systems: elements tied to the grid to initiate autonomous management and smart decision making, like automation systems and metering infrastructure.

The functional planes necessitate external computational routines associated with interactions to the outer world, including consuming or producing energy and participating in management functions. Likewise, they critically depend on certain internal tasks - operate the power system and economically optimize energy usage. This essentially results in two-way flows of power, data, and signals throughout the grid infrastructure.

Different smart grid technologies can be assembled into the following nine categories:

- Electric Vehicles (EVs)
- Distributed Storage (DS)
- Distributed Generation (DG)
- Active Volt-VAr Control (AVVC) [14]
- Wide Area Measurement (WAM) [19]
- Smart Meter Infrastructure (SMI)
- Substation and Feeder Monitoring (SFM) technologies
- Fault Detection, Isolation and Restoration (FDIR) technologies [14]
- Dynamic tariffs (network and retail) and feedback technologies

Every principal domain of the Smart Grid ecology constitutes different spheres of a CPS in one way or the other [20]. The article [21] defines the four pivotal functional segments of a true smart grid system - Integration, Control, Communication and Metering (ICCM) [21]. Integration, simply put, is the interconnection of assorted AC or DC energy sources through the conduction system via appropriate converters.

The future vision is of an entirely interoperable, scalable, adaptable interconnection of the production, transmission, distribution, consumption and management subsystems. It will represent real power, data and signal inflow and outflow (bi-directional) along the physical infrastructure. These

Figure 2.2 Energy system challenges and the role of Smart grids in Response, courtesy #Siemens Electronics.

domains need a hybrid working configuration supporting both intra-unit and inter-unit operation. So, considering the level of advancements, what requires the immediate attention would be a solution to clubbing these together, and the 1st step towards this goal is to ensure smooth physical connectivity. The primary organization of the Smart Grid is as given in Figure 2.3 [22].

2.3.2 State-of-the-Art: Smart Energy System

The global climate change mitigation activities focus primly on the decarbonization attempts on centralized electricity systems - the drive for a consumer-led, low-emission decentralized supply systems. These three aspects translate to the main features of the future vision of energy systems:

1. Incorporate both renewable and non-renewable energy generators (Ubiquitous Generation).
2. Facilitate power flow in both directions across a device or a conductor, facilitating full-scale balancing of supply-demand (Distributed Power Flow).
3. Enable all grid-connected entities with access to non-renewables to contribute to power generation along with consumption (Prosumers in Decentralized Generation).
4. Transform all grid infrastructure activities like pricing, distribution, storage, maintenance, generation and consumption onto Real Time Control (Autonomic Operation).

On a technological background, Smart Energy systems are 'metasystems operating under collective management strategies exhibiting cooperative strategic functioning.' Metasystems are different organizations with common concerns collaborating to determine how they will approach specific issues. Hence Smart Energy Grids are a SoS managing the elemental sub-systems within their super-system environment, as explained in Figure 2.2. The SG framework roles are as follows: (a) Modernize power systems through self-healing, autonomicity, remote control, etc. (b) Update and instruct smart decision making in consumers and stakeholders; (c) Expedite safe, secure, reliable integration of smart grid, and above all, (d) improve energy management efficiency.

Over a short time span, the progress made in the SG arena is quite significant. The article [23] demonstrate extensive research in the integration of WSAN and SCADA along with their applications in CIM. It presents a joint integration scheme using a gateway and Web services on a Web-based SCADA, creating an Internet-operated integrated platform. An appropriate use case and a dependable protocol stack were also developed. In early 2016 [24], established how a worthwhile solution to the performance problems of existing DR implements is the integration of dynamic energy resources. The work created integrated DR (IDR) for a Smart Energy hub. IDR program was articulated as a potential game and verified via Nash equilibrium uniqueness. Cognitive Radio is a relatively newer wireless

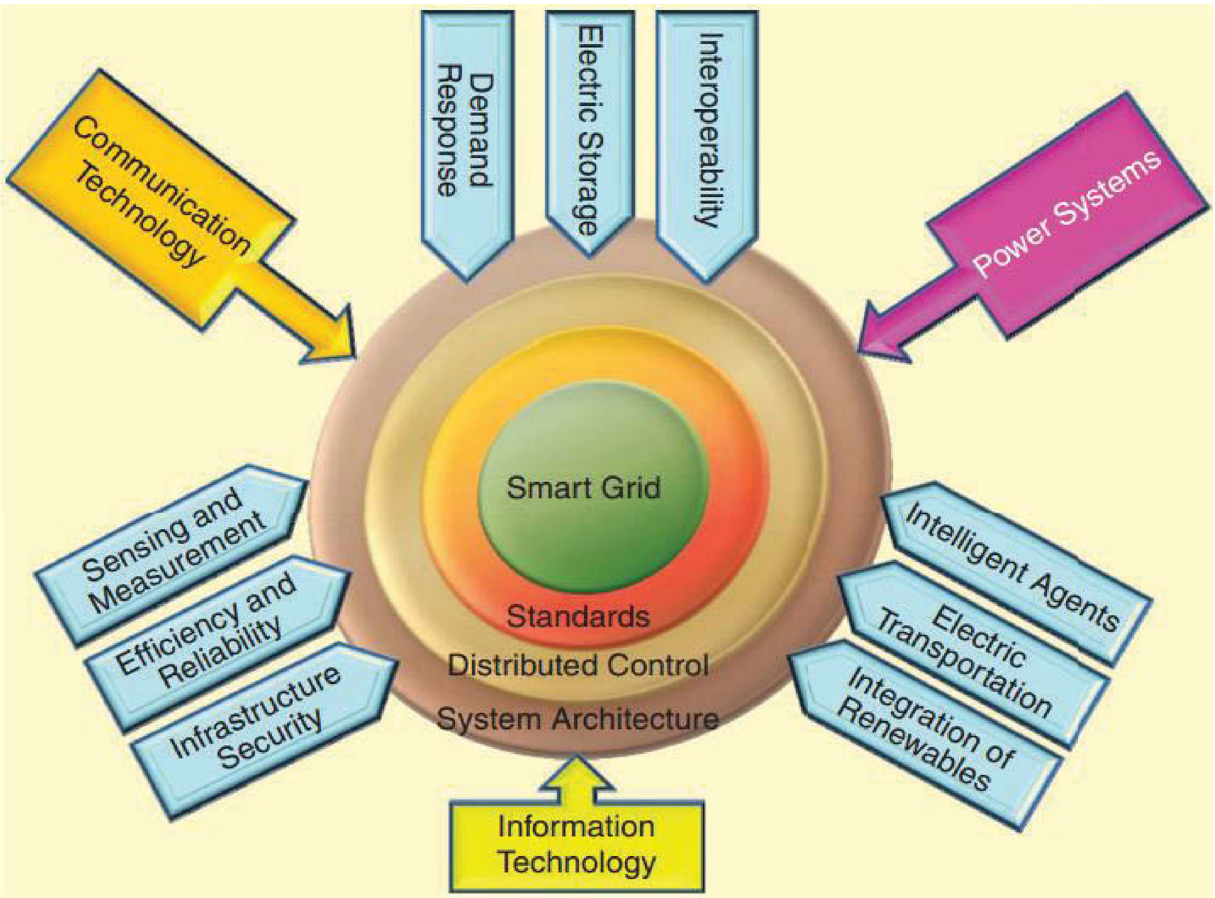


Figure 2.3 Basic smart grid ingredients

network technology applied in SG augmentation of sensing, communication and computation systems [25], where ICA in combination with robust PCA recovers data from the simultaneous smart meter transmissions corrupted by high wideband interference. Similarly, integration of PLC, substation automation and grid sensor networks have also been implemented.

Other than the innovations in the Power System major sub-domains, many review and survey articles covering a wide range of issues entail attention. A different research direction observed is the Microgrid - a miniaturized grid. Microgrid islands allow analysis by drawing parallels with the integration-interoperability problem. In [26], a third-party controlled group consumption and constrained multi-objective optimized load-scheduling is developed. A distributed inductive engine learning local dynamics of the micro-grid element is the output. Mean-field methods apply feedback laws updated from current micro-grid status, for resilient reaction to anomalous events and online control laws to bidirectional discontinuous nonlinear loads. BEMSs or Building Energy-Management Systems for building automation and control represent another SG-CPS integrated environments.

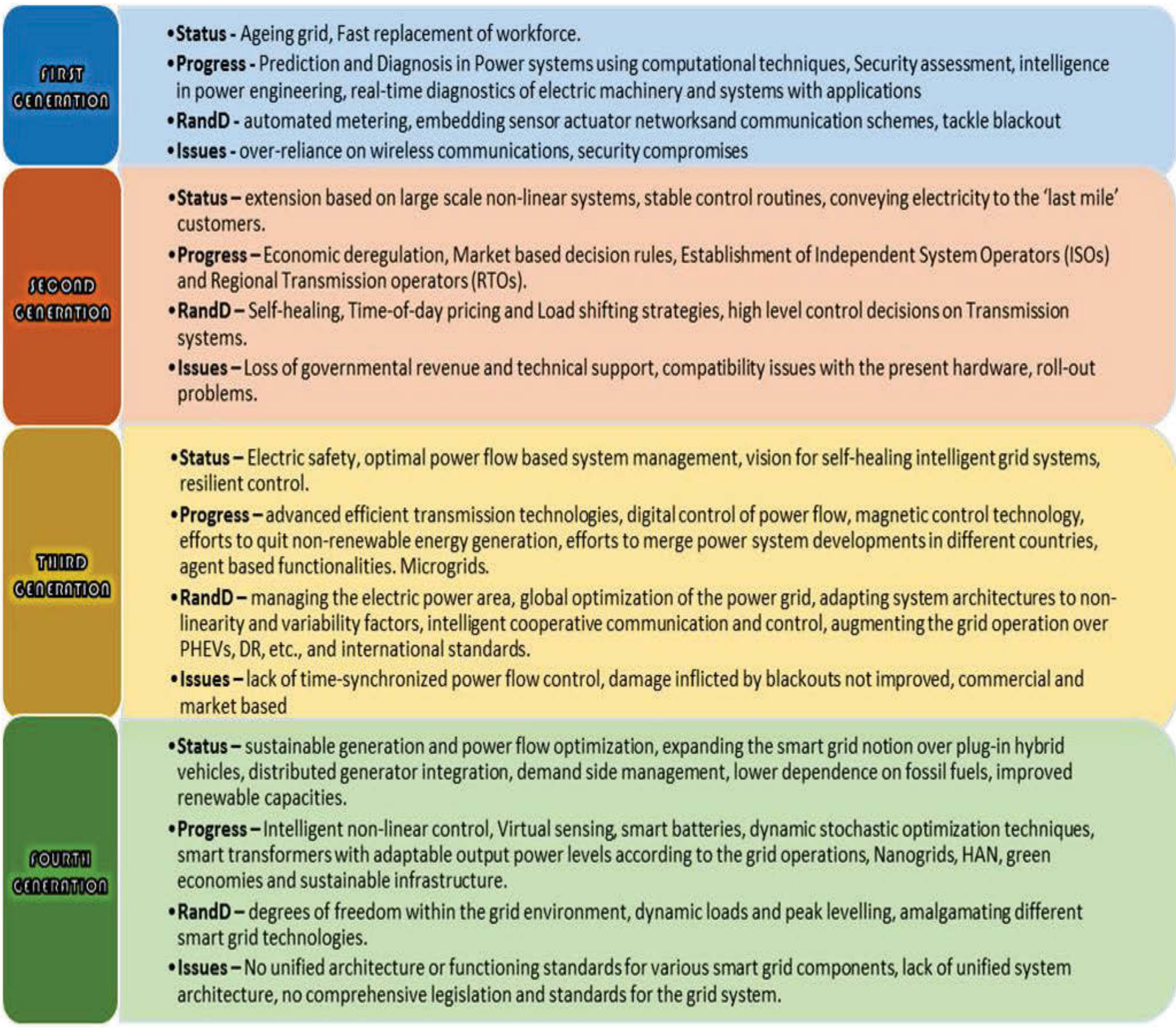


Figure 2.4 Overview of the growth and evolution of Smart Grid Technologies

A Computational Intelligence approach using Artificial Neural Networks solves the research question, with the scope for further improvements listed [27]. Existing research prooves HEMS efficiently reducing grid operational cost, residential peak demand, energy wastage, and household occupant intervention, leading to better eco-friendliness and resident well-being [28]. The use of Computational Intelligence in Heuristic solutions like Artificial Neural Networks to achieve component- integration in power systems are also potentially powerful solutions. Prominent methodologies include Stochastic Control, Adaptive Critic Designs, Swarm algorithms [27, 29], Neuro-fuzzy, Fuzzy-GA, etc.

Though the monitoring, measurement, and control equipment have undergone tremendous changes in the recent, the basic configurations of the grid infrastructure have not changed much over the years. The advancements studied so far provides broad outlooks on the integration-interoperability research theme. The essential Smart Grid ingredients that constitute the heterogeneous elements of the SG ecosystem are given in Figure 2.3.a [22].

2.3.3 Shortcomings: The Need for a Smarter Grid Ecosystem

The impetus for an SG stems from the need to attain increased productivity, lower carbon fuel consumption, enhanced customer experience, and improve utilization of generated power. Notably, the communication system perspective shows drawbacks within standing power grid networks, such as listed below:

- Lack of two-way communication-adequate bandwidth.
- Difficulty in handling increasing data traffic from smart devices.
- Lack of inter-operable infrastructural components.
- Fragmented architectures.

Many standardization attempts, such as Smart Grids Standards Roadmap from Standards Australia for example, has also quoted why “the development and documentation of a common data architecture” [30] is essential - a constrained and boundary-less representation is pivotal for the improvement of SG functionalities. The power grid structure of tomorrow consists of billions of connected devices and decision makers. Simultaneous coordination and operation with wide-area objectives requires decentralized management structure.

The need for integrating Smart Grid with a Data-Networking infrastructure to contextualize ‘Smart Things’ was explained contextually in [31, 32]. So does the integration of all energy and storage options [27]. However, all the standard approaches towards development of Smart Energy Grids are effectively short-sighted. Most of the Energy Analytic studies performed so far were along the perspectives of Smart Metering, Building monitoring, Smart Home Sensor networks, Online Pricing, etc. The works point towards how Interoperability is given relatively high importance than pure augmentation, amongst the modern power grid domains. This fact highlights that various cyber and physical grid subsystems may operate toward specific goals but never interact directly. It is due to the mismatch of physical and logical implementations of these units. The convergence of different smart grid elements and standards is achieved only when this gap is overcome. The problem is further elaborated in the next section.

The present-day grid infrastructure suffers from severe disadvantages like less flexible structures and generator-to-consumer unidirectional power flow, analog front-end modules managed by digital backend support systems Optimization of ICT-enabled contemporary grid requires a great deal of effort in certain ignored areas. The objective behind development and innovation in Smart Power systems are shortlisted below:

- High energy efficiency
- Integrated heterogeneous systems
- Low carbon emission
- Two-way energy flow and prosumers
- Increased renewable power generation and compensate fossil fuel consumption
- Dynamic real-time changes induced by weather equipment, and usage
- Real-time pricing and demand side response moderating production and consumption.

Achieving these features requires a lot more advancement along a multitude of technology scales. The heterogeneous CPS composition complexes these demands. In a Smart Energy Grid, main elements include pure power producers, Prosumers and consumers, grid transmission and distribution infrastructure, storage elements and tertiary components like PHEV, CHP plants, etc. Their featured information streams vary from voltage and current magnitudes to load factors, online price variables and even Demand Response vectors. Obtaining dependencies amongst these data streams means overcoming barriers based on grid interconnection, signal dimensions, analog or digital systems, information or signal transfer, etc.

A close inspection of the state of the art electricity sector reflects on a saturation of the technology scale. Concerning the ideal smart environment characterization, contemporary energy grid communication networks have evolved to a mismatch condition. Intelligent end devices (entities similar to PCs and servers) interact with dumb intermediary devices including substations, and Microgrid interfaces (analogous to switches and routers). To overcome this hurdle, inducing intelligence within the SG framework is of paramount importance. Intelligent traits will alleviate management discord between the elements. The result is the synchronized parallelized operation of the super-system component systems. The underlying problem may be stated as below:

“Interoperability is the preferred path of advancement for Smart Grid stakeholders. Although stated effective, efficient and economical in most recent literature; it still doesn’t add up to the true system-of-systems definition. The real scenario should be of Integration, where any component of the Smart Grid, irrespective of different physical or logical configurations, should be capable of data, info and power signal exchange”.

But the primary requirement is to realize the original vision of “Intelligent Grid systems” – enable the system to learn, understand and deal with new or trying situations, apply that knowledge to manipulate operational environment. The power grid analogy of this definition is as framed below:

“The capability of information gathering sensors and communication agents and actuators, capturing grid behavior and state data along various dimensions, and translating them to critical data and control signals.”

This optimizes performance against various grid actions or events that compromise efficiency. Nonetheless, the very capability stated above places significant pressure on existing Electricity grids by the diversified needs of customer and market domains interacting with existing CP infrastructure.

Autonomic or otherwise, an efficiently designed communication network involves in-depth analysis of the communication system necessities, fitting protocol design, selection of appropriate technological base, and a subsequent heterogeneous network management system [33]. The chief technology gaps hindering the full deployment of Smart Energy Grid organizations were studied in various publishing [34]. Out of which, grid integration facilitating intelligent interconnection of the grid elements and bidirectional communications were identified keystones to overcome both physical and cyber-infrastructure deficiency. The following grid scenarios explain the need for interconnection:

a. *Demand Side Management or DSM:* The most significant efforts would be what the DSM takes, as illustrated at this moment. *A) Complexity*: DSM involves different entities of varied functional requirements, like sensors, appliances, and controllers, which need combined data collection and communication for network formation [35]. *B) Scalability:* the number of users are increasing day by day, demanding more data rate and bandwidth to avoid the latency [36]. C) *Dynamic Consumer behavior*: random behavior of consumers is the most terrible barrier in Load Forecasting and DSM. Reduction in peak load and an increase of reliability and stability of the system mandates capturing information on consumption patterns of users. D*) Lack of an efficient communication infrastructure*: the main obstacle to the deployment of SG in a real system to be solved efficiently, with the least investment and efforts, is through the use of a heterogeneous network [37, 38].

Dynamic DSM, a keystone to initiating accurate intelligence within the grid foundation, has been sidelined in research because of the complex dynamics of consumption, random behavior of consumers and missing computational abilities [39]. Such issues mainly occur because, irrespective of the efficiency, clarity or overall structure of the model, the mathematical statement and framework for those three attributes are incapable of capturing their full significance and implications. The most important feature of DSM is, however, the dynamic power usage and constant load signals, the corresponding dynamic and constant price vectors and demand variables, resulting in the similar behavior of supply and feedback information.

b. *Distribution Network:* These components are active modules deployed for sensing and control. They include smart meters, PMUs, power line monitoring, smart devices, employed through IoT based systems. The information generated include energy consumption and generation profiles, power quality outages, network conditions, conductor health and interfacing connection points. Also, considering electric power dynamics and the billions of user and device entities within the network, all this content take the form of real-time Big Data streams.

b. *Home Energy Management systems (HEMS):* Survey article in [40] explains the significance of observing electricity usage and energy conservation patterns at the consumption premises of the HEMS. The overall system was builton a centralized controller interfaced with various intelligent units in the residential sector. It was also a good motivation towards adopting an interoperable grid interface named the Point of Common Coupling.

c. *Substation Automation:* It embodies the integration of renewable plants, virtual power plants, and microgrids that impose great stress on the physics aspects of the grid dynamics. The stress mainly derives from voltage regulation and conversions, thermal runaway, junction current, impedance matching, etc. So far, efforts regarding grid modernization were pulled in by these aspects that the generation and coordination of ICT for such events lags behind [37, 41].

d. *Intelligent storage units and Electric vehicles (V2G and G2V technologies):* They represent autonomous and dynamic floating loads as well as pseudo power supplies, so time-dependent coordination of grid activity coupled with autonomic management of the two together takes the problem interconnected integrated grid structure to an entirely new dimension [42-44].

e. *The Prosumer abstraction:* This highlight feature of SG portrays an economic motivation based on expense minimization, which is explained here. 1) Power consumption, production, and storage. 2) Operate or own a small or large power grid section or entity, and hence transport electricity. 3)

Optimize economic decisions regarding its energy utilization, with a minor contribution to intelligent decision-making.

A decentralized hub of "Prosumers" may sell their surplus energy as well as buy from the power network, represent the installation of distributed energy resources across a network, balance self-healing Microgrids, organize disaster recovery as well as promote network optimization within the complex SG network. Such hubs point at the added dimension in all the signaling and communication involved amongst the grid elements. The work [45] explains how the Prosumer network notion is the turning point for future energy networks, and right now having such a Heterarchical dynamic network entity interacting with an equally complex central grid management system is still in the conceptual form. Enclosing all the data and communication information in such a scene is what requires further improvement.

f. *Distributed Generation:* DREGs require higher percentage penetration of renewable energy generators. It mandates significant quantities of conventional backup power and vast energy storage subsystems. Only then can the natural variations in power generation statistics depending on the time of day, season and other factors like amount of solar radiation or wind can be compensated, and the variability be handled [46].

These scenarios illustrate how the SG domain suffer from *"Platform Fragmentation"* (in terms of mobile devices technology, IoT systems, Connected Vehicles, etc.). The term refers to the situation where abundant varieties of hardware, software, middleware and framework implementations exist, with a lot of variations and differences, within and amongst sub-systems of a super-system. It makes the task of developing consistent intra- and inter- ecosystem applications tedious.

The situation above explains why information analysis, Smart management and intelligent architectural directions for SG are critical. Information will lead to the convergence of utility-controlled and consumer-oriented energy subsystems. Such fusion of new technologies onto mission-critical control system implements is very challenging [40]. The hard real-time constraints, continuous run-time of the system and resource-constrained legacy devices block commercial deployment of successful experiments.

This assessment summarizes that *bridging the gap* or bringing all subsystem technologies together is essential. Roll-out of supportive infrastructure will establish newer SG components, while spanning to existing elements stagnated as prototypes. Stagnations are partially due to fractions of SG morphing into other trendier concepts, such as the Internet of Things (Smart homes), The Energy Internet, DREDD technologies, etc. [47-49]. The condition represents Ecosystem Entrapment - the overall power system environment getting sectored and each sub-sector leaping forward in progress by themselves. The global super-system framework is left behind without catching up. Various other works showcase similar aspects based on several extensive surveys and studies, as listed in Table 2.2.

2.3.4 Technology Gap: Autonomic Intelligence in Smart Grids

The futuristic innovations in Smart Grids and Smart Power Management has thus shown to be biased. Advancing depths of consumer-driven functionalities and applications orienting toward power consumption optimization, energy savings, and the green drive is way ahead of others. Meaning, in

Table 2.2 Survey and Review articles that helped the framing and defining of the Research Problem and statements

No.	Reference	Synopsis
1.	[6]	Survey of the smart HEMS system configurations with wireless networks, smart home appliances, digital citizen services, and smart sensor technologies.
2.	[21]	Different ICCM methods and tools, and the need for integration of different such technologies as renewable generators, storage, AC and DC subsystems, active and reactive power management and smart communications.
3.	[52]	Review of security challenges and existing best practices for improving and maintaining security, along CP integration between SG and IoT along SCADA architecture.
4.	[53]	Potentials of cyber-physical (CP) integration of next-generation WECS.
5.	[54]	Survey of DR potentials and benefits in SG, enabling technologies like smart meters, energy controllers, communication systems for coordination of efficiency and DR in a smart grid, with reference to real industrial case studies and research projects.
6.	[55]	Overview of Multi-Agent Systems, HVAC-specific examples, potential benefits, and potential challenges; MAS development tools and techniques; multi-agent HVAC control systems survey.
7.	[56]	Study On WASGM and WAC as future Wide Area Systems, via Smart Meters, SSNs, PMUs, PDCs, SCADA and EMS Systems, for wide area control and stability; lack of stable control, smart monitoring and management structure, and intelligence.
8.	[57]	Contemporary smart grid communications, open research issues, potential advantages and research challenges of the smart grid.
9.	[58]	SDN based Integration solutions.

comparison, measures aiming to induce smartness in the physical or logical grid infrastructure (as the definition suggests) is very less-featured. Many vendors and market utilities have launched apps, GUIs, and platforms, to ease and smoothen our interaction with energy providers and related infrastructure.

Unfortunately, they all tend to work around the traditional Power-Data-Control flow dynamics. Expansive review of published works conducted study of various approaches towards different building blocks of the Smart Grid ecosystem. For instance, a study of the ICCM methods and tools for smart communications and power management shows how dynamic system characteristics were not observable due to the drawbacks of legacy measurement systems designed based on energy flows. The current headway of SG paradigm needs application of the corrected sense of autonomicity. In SG environment, establishing Autonomic intelligence properties was given preliminary investigation in position papers, as well as design for an autonomic power system [6, 50, 51].

However, certain paradigms need to be observed and worked upon to derive the exact direction in which autonomic solution design should progress. A comprehensive literature study spotted the missing all-inclusive vision, correlated with the integration problem. Hence, a plausible solution is to ingrain the cyber-Physical Smart Grid with Autonomic Intelligence. In reality, like most technological ventures, the emerging power grid has obstacles ahead of and behind its direction of progress. *"Realization of Intelligence in its true sense"* is a long way to go as has been hinted a couple of times by now [43, 59].

As stated by the International Renewable Agency, *"A sustainable energy system is a smarter, more unified energy system*". This explicitly calls for more advanced intelligence levels. Through the course of this review, the plane-of-interconnection of heterogeneous grid elements is hypothesized. The actual design and implementation of an Integration Plane for the Energy Grid system is a much less complicated task. The various parameters of the Smart Grid which will act as markers for Autonomic Information processing in an SG super-system is as shown in Table 2.3.

2.3.5 Literature Survey

The current progress levels of the Smart Energy grids paradigm are far from full-fledged implementation. So far, Smart Power grids have been described as autonomous systems, autonomous agent based infrastructure, System of Systems, Multi-Agent system, etc. [6, 58, 60]. Most of these facets of the modern power grid are subject to extensive ongoing research and innovations. The present day initiatives and approaches span across a vast range of subjects and concepts, dealing with the same or different kind of issues from similar or contrasting perspectives [55, 61-71]. All inconsistencies pointed out by the Pilot Study section concludes how existing power grid does not

Table 2.3 Autonomicity Inducing Parameters of The Smart Power grid

SES ELEMENTS	AUTONOMOUS FUNCTIONS				INPUTS (PREDICTIONS, VARIABLES & PATTERNS)
NREG	Regulate supply	Charge backup storage unit.	Monitoring power flow	Variable pricing	Predictions, variables & patterns: renewable output, regional demand & load, stored power level
DREG	Predict output power levels.	Smoothening supply.	Load profile regulation	Grid Health	Meteorological data, reference inputs, dynamic pricing, demand management, stored power level
Transmission Network	Islanding	Switching grid segments.	Prosumer services - industrial loads.	Fault tolerance	DREG & NREG performance, industrial customers, grid islands, area-wise observations, Microgrids
Distribution Network	Dynamic island configurations	Autonomous management of end-user elements	Prosumer services - commercial loads.	Price updating	Distribution grid sectors' Power flow, Commercial & Domestic customers
Prosumer	Adaptive Load scheduling.	Selling power to the grid, cluster or Microgrid	Power flow switching - battery, grid & renewables	Billing	Connected devices, Stored power level, dynamic grid price, self-demand, cluster statistics., renewable forecast, grid status – islanding loading, etc.
Consumer	Adaptive Load scheduling.	Power flow switching – battery & grid.	Demand management	Billing	Connected devices, dynamic grid price, self-demand, cluster statistics, grid status – islanding loading, etc.
EV & PHEV Charging Hub	Floating Load capacity.	Storage charging / discharging.	Intelligent EV positioning.	Demand response	Regional load profile for demand management, demand response.

meet the needs of the twenty-first century. The increase in power demand, complexity in managing the heterogeneously built power grid with generation and capacity limitations has not yet been properly balanced.

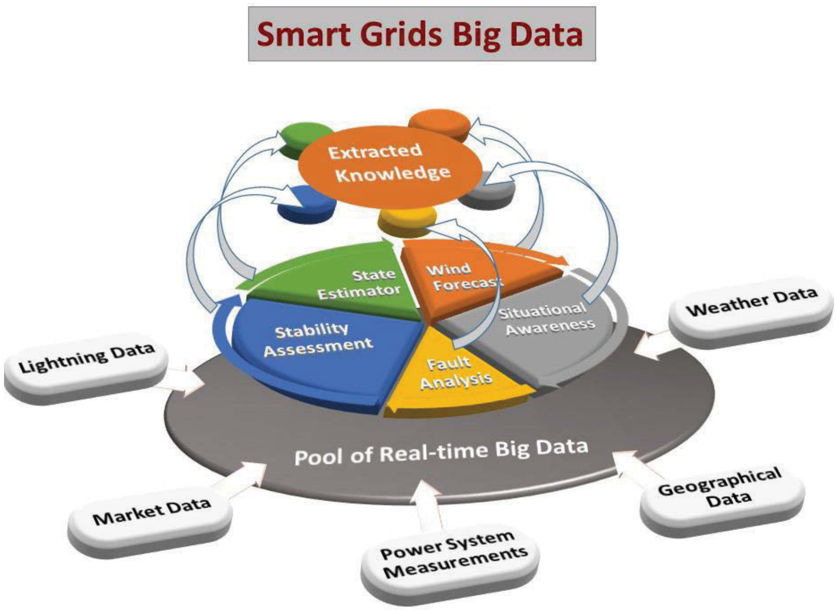


Figure 2.5 Various forms of Big data volumes generated within the SG environment (Source: Texas AM university)

2.4 CHAPTER SUMMARY

To sum up the discussion, as energy systems are less-centered and more decentralized and connected, integration becomes the focal point. Likewise, heterogeneous systems being the subject of investigation, data oriented system management is essential. Thus, a complete makeover is required for bridging this gap, as in update, enhance production, reduce any wastage between generation and consumption, avoid blackouts, outages, grid failure, grid overloading, power theft, etc., leading to higher efficiency and better restoration. This is very crucial for the widespread acceptance of SG functionalities and componentry to be deployed commercially. Data or information exchanged between heterogeneous devices is the basis of all questions.

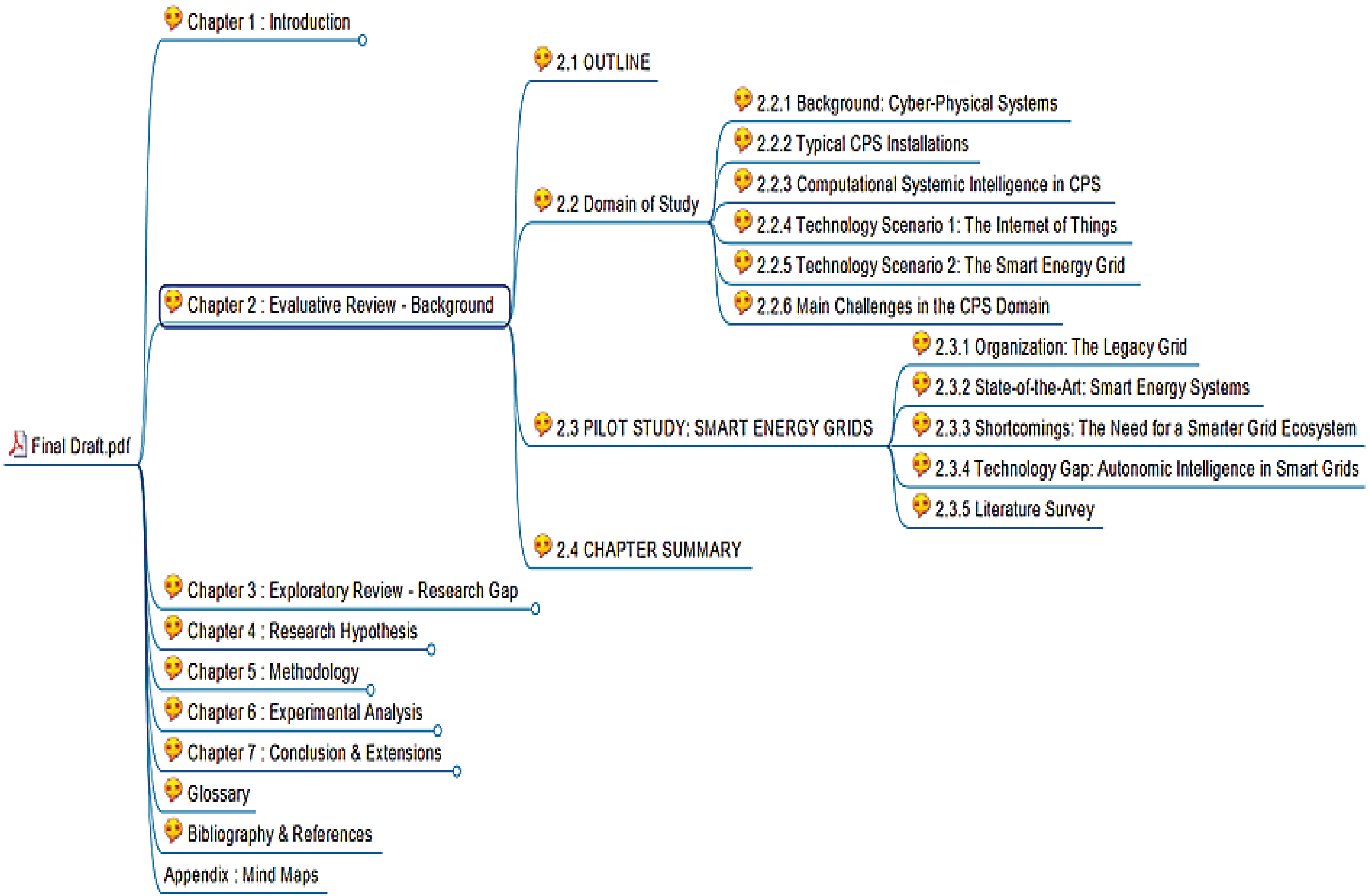


Figure 2.6 Mindmap outline of Chapter 2: Evaluative Review - Thesis Background

Chapter 3 : Exploratory Review - Research Gap

3.1 OUTLINE

Chapter 3 of the Thesis is the exploratory study of the gaps and deficiencies in the contemporary technologies in the set context, culminating in definition of credible Research question and Hypothesis. The gaps are sketched for existing technologies in terms of theory, empirical and practical evidence, research methods etc. The established research context in in Chapter 2 is aligning with the research gap through a viable solution satisfying the study objectives. Hypothesis, Proposed Methodology and validation are framed and explained.

3.2 PROBLEM DOMAIN: MULTIWAY BIG DATA IN CPS

The standard CPS configuration is heavily reliant on connected devices. The ubiquity of sensors and mobile devices demand scalable computing platforms (every day 2.5 quintillion bytes of data are created [72]). The estimated volume of devices to be networked by 2020 is 50 billion [73]. This calls for particular consideration on the aspects of storage, transportation, processing, and access of vast volumes of data produced.

The grid elements have demonstrated the need for more than what is existing now. The depths of smartness and autonomy need major revision, coupled with human system administrators and technicians dealing with the high level issues (with or no intervention on operations like configuring service, provisioning and assurance filing). In the SG domain, considering the level of tech advancements to the individual components, there is less than 20% efficiency in integrating them with the existing power grid. This fact draws the stakeholders' attention to the need to adapt better computational solutions. It will contribute to solving resource consumption, latency and delay issues (performance classifiers) inherent in time-critical grid-data processing. Other resource management criteria to be resolved are accuracy, time distribution and availability, network re-configurability and reconfiguration time.

In different time-variant CPS-systems, similar issues and benchmarks restrict full efficiency. Data from device monitoring collected over time is necessary for rendering higher efficiency, effectiveness, predictability, and portability of processes. However, analyzes of data has not always been carefully screened. This may result to misleading results. Thus, the representation and quality of data is the first and foremost demand prior to any form of analysis. When the complex data streams generated from heterogeneous super-systems are used in aggregation, they have the potential to interact underneath the standard scale of system design. These interrelations may not be quantifiable, but may be qualified, hence visible and estimable.

Analyzing data flow in many mature CPS architectures is heavily constrained. The sheer size and modality prevents extracting vital information beyond our cognition. This signifies the need for more efficient data processing methods. The requisite is thus to capture the simple yet highly descriptive pattern models from large Data troves. It aims at enabling synchronized recording, structured storing and visualized interpretations of multimodal data. The approach adopted in this article is devising a method to overlook the actual operation of a central processing unit under data inflow.

3.2.1 Multiway Big Data in CPS

CPS subsystems interactions represent complex phenomena as combinations of underlying latent components and inter-related variables. Their data streams have to be broken down into factors to extract hidden information generated by those variables.

Most successful prior implements on Big Data processing focus on specific points or attribute pairs, such as applying two-mode relations. Typical 2-mode relations in data mining algorithms are implemented column-wise, and newer ones run along planes. This depicts the common approach in Data Processing applications is to “flatten-out” the data. The realization is done by considering a variable pair at a specific point of time and iterating the process over different aspects, to construct attribute models.

The most persuasive Data Mining algorithms applied in research community were experimented in many subject areas. Some examples are C4.5, k-Means, SVM, Apriori, EM, PageRank, AdaBoost, kNN, Naive Bayes, and CART. They individually or in combination implement Analytics (typically look forward to model the future variables based on past conditions, prediction included) and Analysis (looks backwards over time, identify behaviors and trends of the info inflow to formulate more relevant rules). Processing of Big Data along these rules, are still ridden with problems.

The operationally efficient constraint for arbitrarily large data sets is that the envisioned data sets cannot be resident on the device. This led to the birth of the Cloud architectures [74] and other remote processing-storage technologies like Fog and Edge computing. Following this, Cloud or Fog based solutions [75] were derived, as well as Data Abstraction, which may be achieved by a multitude of techniques. In spite of being greatly advanced, and deeply optimized for latency and resource consumption, they still induce a processing delay of data transfer across. So, alternate visions which eliminate the need for transference and storage of Big Data are welcomed.

Thus, need for newer outlooks into Data Engineering has been established.

3.2.2 Higher Order Big Data Processing

Multidimensional data that arises from the non-linearity of signals reflect ON multiscale structures in higher-order array data structures. Extracting underlying information across modes and dimensions in Big Data is heavily computationally intensive. The massive volumes of real-time continuous data present the difficulty of distributed subscale hidden patterns (subscale refers the perspective into within the given scale of operation). Extracting such patterns thus materializes as the subscale problem of time-series blind source extraction, which helps ***discover***, and not define new modes or axes. Hence, specific pre-defined components need not be defined; instead data is allowed to determine the components in a problem or domain specific solution space.

In the case of complex systems, data streams can either be 1 or more metric or measure from a source. Each stream contains interdependent information on the actual complex system interaction. Analysis of multiple data sources correlated amongst themselves needs various modes and features (signals and sources) distinguished. Estimated parameters play the role of features in a classifier to perform classification.

There is a multitude techniques for exploring and mining deep patterns in series data. Two of the most common and most widely applicable notions are Component Analysis and Spectral Analysis.

1) *Component Analysis:* A statistical domain representation of the data is utilized instead a time or frequency domain counterpart [76]. Data is projected to a new set of axes derived from statistical criterion, denoting independence. Prominent methods included here are PCA or ICA oriented transformations. Such methods are affected by the structure of the data being analyzed.
2) *Spectral Analysis:* This kind of operation utilized data representation in a spatial domain. Here, data projection is onto a set of axes represented discrete frequencies [77], where the independence is assumed. Fourier transform uses projects a data segment onto basis components called Fourier components which are fixed.

The fundamental principles behind standard component analysis techniques hold a strong limitation. Spectral methods such as Fourier techniques work on the assumption that the projections drawn onto each frequency component are independent of the others. In PCA and ICA, on the other hand, a set of axes which are independent of one another in some sense are developed. However, there are cases when the dependency status of the basis components is unknown, or whatever prefixed information necessary to evaluate the right set of independent components. Also, PCA and ICA even try to reduce redundancy by scaling dependencies down to the single features, even when there are higher order latent associations. That is when such techniques fail, highlighting the research problem of this project.

3.2.2.1 Multiway Big Data Processing: The SG Context

The elements involved in the SG domain do not have congruent operating characteristics. They keep varying depending on Data patterns and Anomaly profiles. To match heterogeneous complex system outputs with processing objectives, what is required is constraint-free operational flexibility. This is the main objective behind Integrative Interoperability. Hence what should be done is enabling Data Processing to work around sensitivity measures of an operation routine or a subsystem. Engineering newer models of Data Processing ensures coherence of technological frameworks. Reference independent parameters and dependent variables generated as their complex functions is represented by the parameter data streams and their interdependencies. The streams are studied individually or associated, in time-series format. A typical example is CEP (Complex Event Processing), an event processing technology combining data from manifold data streams that lead to event or pattern conclusions recommending more complicated circumstances [78]. Sense-making and Situational Awareness is another approach [43].

The alternative is to merge multiple streams expressive of multiple sources and variables as one flow prior to processing. Then a multiway arithmetic tool may be applied to extract interdependencies amidst constituent streams of the flows. The analyzed dependencies and associations signify parameters and variables that define the system-of-systems. This proves that Data gathering and analytics is the key solution towards integrating advanced intelligence traits, especially in scenarios of Big Data. Analytics for efficient data handling are collapsing the information gap in all leading technologies, giving rise to complex autonomic systems. Ample support from technology trends like Artificial Intelligence (HMI, HCI etc.), Real-time analytic solutions, and Predictive optimization for converged SAN also aid in achieving the objectives. The data processing problem for instance by allocating/deal locating dynamically processes, and the extraction of useful information for example by leveraging on specific solutions from the Big Data context.

Therefore, an understanding of the Data that is generated, exchanged, and administered into the Power infrastructure is of utmost significance. It is the foundation to any functional or operational attribute, irrespective of the innovation or technology utilized for the implementation. Hence, the background of this research is the ambient intelligence embedded in CPS symbolized by the Big Data flows generated within the CPS environment. The application range for such aspects is too wide to be condensed into specific domains. Typical examples include Smart Energy Grids, Smart Home applications, Smart Cities, AR, VR, Software Defined Networking, Network Virtualization, Real-Time

Streaming applications like Social Network Data Mining, Hyper Spectral Image Processing, and Traffic Pattern Analysis.

3.2.3. Next Phase: Autonomic Middleware for CPS

In many full-fledged established CPS architectures, constituent elements are connected through software interfaces called middleware [79]. A middleware is an agile, but dependable software designed to dynamically support diverse constantly varying necessities, typical of CPS applications.

Messaging-oriented middleware technology, supporting many communication protocols and domains, will progress to accommodate various types of systems and component to efficiently share information [80]. The role of a middleware is to guarantee that components do not strive for subtasks in the larger system and deliver messaging services for them to communicate. The basic requirement of such a platform for heterogeneous CPS is the ability to process multiway data. The higher the size, the better.

Research into middleware architectures or platforms for realizing CPS applications is still in its infancy. The connected component offers time series data which can be fed into a context-aware communication network. They may be stored as user or device preferences in a time and space-aware manner using a distributed cache. This allows the middleware platform full suitability for heterogeneous grid infrastructure. This will also contribute to Pro-reactive control or Strategic control, (partly proactive and partly reactive).

3.2.3.1 Point-of-Interconnection Oriented Middleware Design in SG

Anomaly detection through pattern identification, classification and analysis is a relatively new approach from SG perspective. The SG domain suitability is described by [81], though it is relevant in 1 and 2 dimensional data systems only. These techniques need a lot of modifications and perhaps alterations before being applicable on Big Data based systems. In case of modelling Smart Grid data anomalies, either a rule based or event based approach may be utilized.

But this achievement still has to progress a lot more, with obstacles hindering the design development and realization of innovative solutions for a green sustainable energy infrastructure. Many national, regional and international standards have been developed across the world, for ICT based transformation of the legacy grid structures. The SGCG has unified the important ones under a sole framework to instill standardization of effort [30, 82]. But they all address the various Smart grid domains more on an individual but interoperable approach than an integrated one. Hence, a major research question to be addressed in this domain is the *system integration* of various technologies, disabling of barriers of regulations, consumer commitment, and technology ripeness, by coordinating the legacy grid environment with contemporary tech solutions and related innovations, which will be the crux of this research initiative.

Data modelling and Communication infrastructure design satisfying Power Grid stability is crucial for successful implementation of Smart Grids [31]. Real-time analytic engines enable deep systemic intelligence through the study of the power network to control the current state of the components and forecast what may happen, so as to frame an action plan [83]. It was for long a viable industrial interest. Processing real-time status and control signals will quantify operational details bypassing mathematical constraints. This is done by defining such signals as complex functions of parameters that are arithmetically and analytically heterogeneous. Once the intelligence in such sources is symbolized, it may be mapped onto high level architecture and extracted, implementing autonomic operation. It is then modelled, simulated and tested using a nominal quantity of power system elements and infrastructure standards and models. This technique will revolutionize the various energy system architectures and layout for potential mass scale. It will lead to development of a uniform collaborative systems ensemble, scalable for extensive SG Integration.

A stable Smart Grid ecosystem requires omnidirectional mission-critical data processing functions, in synch with the physical infrastructure, to keep up with the temporal operational variables. As such, pre-defined network protocols that do not flexibly configure or adapt to dynamicity are barely useful. This brings into light the real question – the transfer, processing, storage and application of the voluminous data harvested from smart energy systems.

The need for data sources integration has been solved partially through PLC SCADA which are operational sources and smart meters which are the nonoperational sources. The increased use renewable energy generators and changes in the energy supply system are also key factors driving it. Implementing this aspect in the Smart Grid subsystems is as explained below:

- *Generation capacity* of power sources as a function of external conditions (prevailing weather, system health, etc.), storage status, prosumer load variables, load and power flow patterns and distributed generation efficiency.
- *Communications technology* as a purpose of the grade of grid connected devices including generators, prosumers and electric equipment like IoT modules, grid line status, storage and load variables, patterns of generation and consumption, input and output forecast models, price index.
- *Control technology* as a function of the IoT connectivity, prosumer supply / demand status, Load and Demand management, Online Price Variable, decentralized scheduling etc.

To be more specific, some of those functions are listed below.

- Transmission and Distribution Infrastructure integrity as a function of parameter values breakdown across the physical setup (grid networks, substation equipment, SMI etc.)
- Demand management as a function of consumer behavior, online pricing, or demand response.
- Anomalous event occurrence probability threatening grid efficiency predicted as a function of patterns in measurements from connected electrical, cyber and solid-state devices.
- Operational Maintenance optimized as functions like passive heating and cooling, component lifetime and maintenance.

From a basic material substructure of the power system, the point of power inflow or injection into the system is also known as *the Point of Interconnection* or *Common Coupling*; abbreviated as *PoCC* [84]. It is the physical point in the electrical system formed by line inter-connections to multiple electrical loads, generators, storage units Prosumers, or Microgrid islands. A connection must be able to be made at this point, with no limitation on physical or network statistics, like voltage, data type, etc. Management of the interconnection profile in the cyber-physical electric infrastructure, especially at the PCC, is an issue requiring immediate attention. IEEE-519 [85] defines this point's accessibility by both the utility and the customer for direct measurement. The PoCC where utility and customer/generator assets are connected; is often, but not necessarily always, metered. So, designing and setting up a distributed-networked set of PoCCs will elevate the legacy power systems to the high level Smart Energy systems.

As an initial step, designs put forward to establish voltage-balance at heterogeneous load interconnection is to be studied. The last four years, fortunately, have seen budding research attempts towards this problem. A Smart Distributed DSM of LV Distribution Networks using Multi-Objective Decision making, post analysis of voltage unbalance sensitivity and stochastic evaluation for PVs installation and penetration as well as Grid-to-Vehicle and Vehicle-to-Grid arrangements [84]. The use of droop-based APC for overvoltage prevention in LV feeders with high penetration of PV was set-up and validated in [86]. Apart from these, white papers by commercial vendors like Siemens, governmental research position articles etc. [87-90] speak of why this issue should be dealt with. The International Renewable Agency has made all-inclusive statements on this issue [89], and that certain countries are already progressing rapidly. Denmark's Seamless Wind integration system, Puerto Rico's Smart grid rollout initiative, Singapore's Intelligent Energy System and many more like in Mexico, Netherlands, Korea, Canada, Austria, California (The US), Sweden and Jamaica were described. All such enterprises are pilot or demonstration projects which could handle only 2 or maximum 3 SG entities integrated more using interoperable standards than end-to-end synergistic operation.

While theorizing integrated system characteristics, difference between a ripple, parallel processing and flat architectural frameworks is noteworthy. Also, the incorporation of observing and online analytics, require the analytic tasks to be finished locally at the monitored-data acquisition point [90]. Care must be taken that an overall global operational mode, with respect to fault tolerance, QoS and the other characteristics should be built into the system.

Therefore, the scope of this pilot study is to establish a prospective pathway to induce autonomicity in Smart grid environments – by hypothesizing Multimodal Big Data Patterns as the key to the solution.

SG Interoperability panels dictate a whole batch of standards that dictate such interdependent implementations.

IEC 61850 [37] focuses on substation automation, for data transfer between primary equipment and IEDs through Ethernet network, therefore eliminating the need for the controller unit to directly abstracts information from the grid connected smart devices; *IEC 61499* [37] defines a generic architecture and offers rules for the usage of function blocks in distributed industrial-process measurement and control systems; *IEC 61970* [37] is a semantic model that describes the components of a power system at an electrical level and the relationships between each component; *IEC 62541* [37] is commonly known as OPC Unified Architecture and is a standard widely used in industrial automation– based on SOA to allow end-to-end communication among all members of the power network; *DPWS* standard uses Web Services to enable inter-devices communication, defining two fundamental elements: the devices and their hosted services – based on SOA to allow an impeccable communication among all elements of the power network. The main purpose of having these standards around is to ensure interoperability.

The huge volume of data and information flows involved require advanced data handling processes. To realize all this, what is needed is an efficient multi-way array based autonomic operations.

3.3 RESEARCH GAP

Based on the problem domain analysis, the research gap has condensed into “learning features from raw data”. Motivation is from the fact that more complex, high-dimensional, and noisy real-world time-series data cannot be defined with analytical equations derived from parameters to solve since the dynamics are either too composite or unidentified [10]. Also, traditional shallow methods, which contain only a small number of non-linear operations, do not have the capacity to precisely model such complex data [3].

3.3.1 Need for Higher-Order High Dimensional Pattern Analysis

The present-day real-time analytics solutions push for faster runtimes, response times and iterations over a smaller time window, as in distributed and decentralized communication networks or hierarchical control systems. It will result in the system hitting computational limits with respect to processor speed, hardware efficiencies, and memory and device communication latencies. It is not different than attempts to increase the buffer read or write rate in an 8 bit processor, to match a higher bandwidth, than simply opting for a 16 bit processor. Customary methods used in the modeling of sequential data comprise of the estimation of parameters from a presumed time-series model.

Therefore, widening the data processing capabilities will help obtain solutions along more than 2 variables during the same time window. Finally, with the stochastic nature of the system processes, it's only suitable to represent changes and change patterns in the system behavior through multivariate patterns, rather than using numerical theory or and a wide variety of movement axes.

Multiway analysis represents learning of active and hidden patterns from large scale data flow, by overlooking the dimensions. It works by capturing Time-Frequency Patterns of Intermodal-behavior. The time series analysis of multiple modes is framed on a trade-off between accuracy and processing volume. Thus the approach based on pattern extraction guarantees merits one way or the other. This work illustrates modelling such events and patterns within time-varying Data streams. It is the underlying data obtained from the hidden information; so can be described as the next levels of information depth [91].

3.3.2 Multimodal Pattern Imaging and Detection

Pattern recognition may be pointed to in two different perspectives: 1). Template matching and 2). Feature detection. A template is defined as a pattern based on which items of the same proportions are produced. The template-matching theory suggests that the incoming stimuli are considered in comparison with templates in the long term memory. In case of a match, the stimulus is identified. An improvement to the state-of-the art concepts is to enhance this operation by not limiting the extraction to pre-defined or well-known feature sets. This thesis suggests that dynamically varying Pattern 'images' in multivariate Big Data, based on space-time-index-varying information streams capture every possible state of system components.

The thorough inspection of the Problem Scenario, now deems the identification of an apt solution. The motivation to proceed at this point stemmed from a theory in existence for long:

"What falls short of Words is better described through Images".

It intends to make sense of what is actually perceived in the environment, by the system as well as human elements, through numbers. The conclusive conjecture on this situation is:

1. Find out the “*interdependency patterns”* between behaviors of different grid-connected elements across a period of time,
2. Ascertain if there are variations from the norm, and
3. Apply these patterns for further processing.

The results may be mitigated, or ‘smoothened’ to stabilize the operational features. This is what essentially Homeostasis control is.

3.3.3 Anomaly Detection in Higher Order Data

In today’s cyber physical realm, it is more imperative than ever to change the course of the system operation as soon as possible. This feature is defined as the sensitivity, to adapt to the rapidly changing real-time processing requirements. A review of the state of the art Data Engineering problems validates that it is more imperative than ever to adapt the system sensitivity to the rapidly changing real-time data flow requirements. The challenges clubbed with it are:

- Define indicators and depict and changes in the input flow itself, before it actually gets to manipulate the processing platform.
- Develop Statistical and analytical techniques capable of flagging worrisome data flow inclinations, in time series flows; before they can produce a response in the system.

The fundamental question behind this investigation infers that:

In any complex core system operational dynamics, the difference between event occurrence timestamp and ingestion and processing times, is much different (large). For these same systems, response time per operation cycle is often quite long as well compared to the pre- and post- processing stages. Real-time data processing thus encounters the largest setback in terms of processing delay. It also gives rise to out of order events (Gavin Whyte, KPMG Australia, 2017).

In this context, related incongruities on computational procedure operating on real-time and near-real-time systems violates their base purpose.

For any hybrid or standalone device, its internal technical dynamics materialize in correlation with the input data spectrum, measurement range, and operational characteristics. In lieu of such specifics of a Data set, a much significant aspect to define the systemic behavioral trends is an Anomaly or Anomalous event. They can gauge variations in all kinds of operational specifics, which throws ambient light on the system dynamic variations and hence the operational intelligence.

Any kind of unusual behavior in a systemic environment may be termed Anomalous with respect to the regular behavioral trends.

Occurrence of an anomalous incident is not only the outlier-output values, but also manifestations of other factors such as local environmental changes, coefficient of multiple deterministic hardware events, etc. In case of massive high dimensional data, identifying the exact timestamps of abnormal behavior amongst millions of measurement / reading entries is a taxing process. Not to mention the computational, time and space load on the processor.

The hardships involved in such objectives come with needs to handle large time-dependent variedly characterized data, which may involve accurate realizable forecast procedures, analytics relevant for autonomic system operation and administration, etc. When developing systems, however, it is often impossible to explicitly design for all potential unusual situations up front and frame corresponding strategies. Anomalies, irrespective of occurrence as noisy points or continuous events, are reflected on the metadata structure.

In this regard, defending the system against Anomaly attacks is also an important aspect of engineering Big Data solutions. Autonomic Operation of any heterogeneous CPS than then be improved then by using analytics to predict when system behavior might turn anomalous.

This means establishing vector measured as an outlier [92] (doesn't fall within the desired measurement range), leading to *Autonomic Anomaly Detection* (outlier/error vector outside the desired range/classes).

Anomaly detection analytics in Time series data thus improve supply network efficiency, net energy and resource consumption and reliability. This in turn is adapted into performance characterization, risk management, system up-gradation and so on. Hence proven that it is imperative for all constituent power system units to work in a holistic framework managed by autonomic control.

Anomaly Detection has so far held a different perspectives in the various applications. This promising field needs further optimization to be deployed in the SG environment. It turns handy mainly for traffic monitoring and related security features like Intrusion Detection, Worm Infections, Filtering, etc. [93].

Evolution of latest technologies - Super - Computing, increased Storage and Data Memory Power, Fog applications in Computing and Networking, Secure Cloud Computing and Virtualization, Data Science and Digital transformation – all these help in handling multi-dimensional and complex computing and processing requirements that are inherent in integration and interoperability protocol design.

3.3.4 Research Question: Defining Latent and Active Features of Associations and Dissociations

Features are independent of the dimensionality of data. Likewise, Irrespective of the order number of Data streaming, the properties along any one axis holds. In fact, assembling inter-data-point dependencies along multiple axis will actually help in determining dependencies across different modes or variables. This visualizes as a trait similar in essence to cross-correlation measures.

Therefore patterns capturing such dependencies may be thought of as metadata visualization in terms of variations over one or multiple variables. Metadata is an abstraction of a large dataset from a source that captures patterns of variations, and averages/generalizes well over new small data sets. In essence, they are contextual information that describes the entire data set characteristics, thereby representative of the system macro-behavior.

Different patterns and groups of patterns in the information status obtained from various equipment, defined based on the application-specific parameters of the CPS like output levels, component signaling, etc., correspond to system-response spectrum. They in turn outline the procedure for management of the components. Tracking and chronicling such patterns will also help us determine information regarding the system behavior and enabling conditions, which in turn leads to system defense against Anomaly attacks. Hence for any form of autonomy to be introduced along CPS computational fabric, it's imperative for the signal flow to be monitored and modelled, for generating predictive communication and control routines. The descriptive modelling task of this kind can be related to pattern discovery is used to identify frequent associations within data. Pattern discovery is often used for market basket analysis on transactional purchase data.

The features derived from the knowledge discovery capture dependencies equivalent to metadata visualization of the various information modes as variation maps across one or multiple variables. They signify Contextual information local to each feature's span of variables, in essence their correlation. Extracted features are independent of the dimensionality of data. Likewise, changing the order number of Data streaming, only results in mixing the pattern elements, holding the properties along any one axis.

- *Active associations*: first degree relationships showing input and output (intermodal) variation dependencies.
- *Active dissociations:* first degree relationships showing input and output (intermodal) variation independencies.

Active interrelations are refined out of system response and operational characteristics. For e.g. in the

Smart Grid CPS context, power generated by distributed generators qualify as input, while customer and storage units are the output subsystems.

- ***Latent associations:*** subscale dependency patterns across different data streams, from single or multiple components, irrespective of contributions to input or output variations.
- ***Latent dissociations:*** subscale independency patterns across dimensions of multiple modes, irrespective of contributions to input or output variations.

Table 3.1 Survey and Review articles that helped the framing and defining of the Research Contributions.

No.	Reference	Synopsis
1.	[31]	Investigation of the state-of-the-art information subsystem, communication infrastructure, design, and architecture and system model of SDN enabled SG, along with an abstract business model and common architecture.
2.	[39]	LF and dynamic pricing schemes in SG environment to realize demand side management through statistical and AI based approaches demonstrate the temporal processing techniques and resulting scalability issues.
3.	[53]	Fuzzy based models are extensively used in recent years for site assessment, for installing of photovoltaic/wind farms, power point tracking in solar photovoltaic/wind, and optimization among conflicting criteria – provides realistic estimates.
4.	[58]	Analogy of 'Smart Grid Communication System' to 'Instrumentation Telemetry', its critical applications and parameters.
5.	[94]	5-approximation algorithm to address the generated NP-hard problem, which yields performance-guaranteed solutions, with 3 representative scenarios.
6.	[95]	Comprehensive survey of smart electricity meters and their utilization, key aspects of metering process, stakeholder interests and satisfying technologies, challenges and opportunities due to the advent of Big Data and the Cloud environments.
7.	[96]	About using the data that stems from process automation to help businesses and people make more intelligent decisions.
8.	[97]	Real-time energy consumption data from manufacturing processes can be collected easily, and then analyzed, to improve energy-aware decision-making, relying on a comprehensive literature review
9.	[98]	Phasor measurement units performing their functions as smart sensors in the grid environment and different data streams available from PMUs

Latent interrelations occur in subscale, otherwise undetectable by temporal or spatial data modelling. The mappings generate correlation matrices of the various sources (streams) of component elements along all involved modes. Typical examples in the Smart Grid context would be signify the effect of weather, building occupancy, load variation patterns and other second degree relationships among input and output attributes. They are quantities under study in this thesis.

3.3.5 Literature Survey

The above discussion opened floors for an investigation into various mathematical tools suitable for multi-dimensional conception of multivariate data. With the development of smart grid, physical electrical equipment and data collection and computation equipment are closely interconnected by the communication and control networks. Smart Grid thus becomes energy and information coupled infrastructure, Cyber - Physical Electrical Energy Systems (CPEES), integrating power (sensors, and physical control systems) and information networks (communication and computation and control systems) [11], The present-day grid infrastructure suffers from serious disadvantages in terms of - less flexible structures and generator to consumer unidirectional flow and analog front-end modules managed by digital backend support systems. As depicted, the same observation on varying contexts can be made across the entire Cyber-Physical architectural domain.

3.4 REFINED RESEARCH POSTULATE

This research project investigates the scope for transforming a Time-series block into alternate order and dimensional representation of large volumes of data, to facilitate systematic subscale relations, their interpretation and application. This will help achieve tracking the Stationarity independent semantic associations across space, time and index.

3.4.1 Research Problem in Focus

The complex time-dependent behavior of large heterogeneous data streams, generated as Big Data in CPS sub-systems is first inspected. The aim is to develop an effective solution to capture such relevant active and latent associations or patterns. This will uncover the temporal relationships and dependencies among the components.

This research hypothesizes and proves beyond reasonable doubt that an approach as explained above explicates deeper insight into the data, enabling bypass of manual exploratory analysis and visualization.

It is the pathway to be adopted towards building up autonomic operations induced by deeper systemic intelligence. It will bring better performance to the Data Crunching phase if multiple information facets are integrated together. The importance lies in the need to model Active and Latent Associations and dissociations to represent information flow along multiple modalities, which uncovers Latent intelligence of the complex CPS system.

The various problems identified in the Problem domain of modelling Behavioral and Anomalous patterns of multivariate Big Data streams, that aids in framing the hypothesis are summarized below:

- Need to elevate the processing dimension
- Substantial reduction of operational delay, resource utilization, latency, and information volume.
- Modelling states, state transitions and IO conditions based on space, time, or index time windows.
- Distributed data acquisition and cataloguing with subscale intelligence.
- Support for scalable interactive interoperability among heterogeneous devices
- Quantify and qualify heterogeneous data streams, uninhibited by mathematical constraints at the same space-time domain.

3.5 CHAPTER SUMMARY

Therefore, the scope of this research work is to inspect the suitability of real time anomaly prediction in multimodal Big Data towards implementing these objectives. A robust edge data modelling platform is essential to realize autonomic management of rapidly expanding CP network domains and their distributed services.

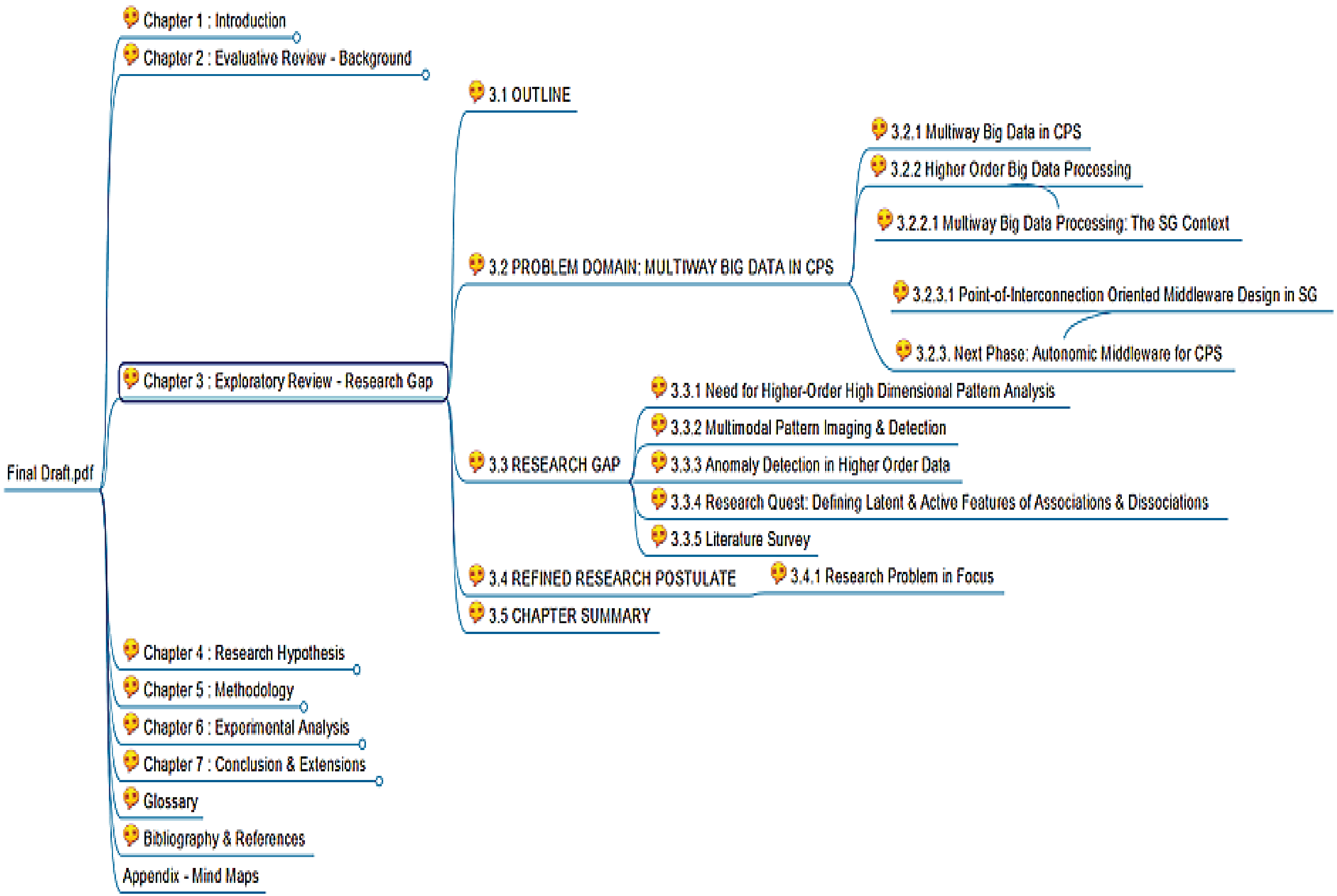


Figure 3.1 Mindmap Outline of Chapter 3: Exploratory Review - Research Gap

Chapter 4 : Research Hypothesis

4.1 OUTLINE

A specific research background, context of the problem investigated and suggested directions have been established in Chapter 2, with objectives exclusively aligning with the research gap. This chapter brings into focus the suitable Solution to the Research problem at hand, while satisfying the study objectives. Mathematical domain of the solution, applicable concepts into the problem, and parallels with other fields of study are reviewed along with discussion of implementation strategies. Hypothesis, Proposed Methodology and validation are framed and explained. Primary sources of information once again constitute the main texted, while secondary and tertiary literature are listed.

4.2 RESEARCH HYPOTHESIS

4.2.1 Core Postulate

Is it possible to model complex dependencies among heterogeneous Big Data streams, in order to improve synergistic management at the SG-CPS interface, with the dimensionality and degree of freedom of the system-knowledge infrastructure increased multifold?

4.2.2 Developed solution

The solution developed in this project adapts Tensor Algebra into High dimensional behavior and Anomaly Pattern Processing. The research contribution is to apply Tensor Decompositions to extract *Latent Patterns of Behavior and Anomaly* defined in Section 3.3.4. The hypothesis thus framed is stated as:

Tensor Decomposition is adapted for High dimensional Data Processing in a typical CPS environment, like the SG ecosystem, for dynamic, multi-modal, multivariate Big Data flows.

They can be processed and characterized, based on dynamically defined Latent interdependency features within the data generating environment, retaining complexity levels and eliminating dimensionality conversions.

The credibility of this method, derived solution and its significance are investigated by analysis and interpretation of proof-of-concept experimental results. The defensible benefits of this approach are:

- For any complex core system, assembling the data streams along multiple scopes or modes will help determine inter-modal dependencies.

- Their quantification will qualitatively measure characteristic behavior and response of the setup.
- Total Data crunching time for a complex autonomic system can be reduced significantly.
- Delay induced by sequential the complete data-processing-window is bypassed for predicting a mode of operation.

4.3 SOLUTION SPACE

Data Integration or Assimilation is the combination of information from various sources or modalities. *Data abstraction* is the concise representation of a particular data body to a reduced representation of the whole [99], in essence eliminate irrelevant information. The generalized procedure for Big Data reduction processing utilizes Learning models and error or anomaly measures. The process advances as given below:

1. Find structured reduced hierarchical component representation of observed multiway data streams.
2. Frame meaningful significance of the data in terms of decomposed factors by detecting patterns amongst the entries in multiple dimensions
3. Find the symbolic representative of the significance, so that for future time-streaming data, patterns maybe be easily extracted based on these representations.

Although work done so far on various kinds of Real-Time Big Data Mining and Extraction techniques produce reasonable information abstraction, latent data trends are still only vaguely perceived. These inherent Big Data patterns can help qualify underlying similarities, and hence the associative behavior of the events or sources, known or otherwise. In this work, they are termed Behavioral norms and outliers viz. Behavior and Anomaly patterns.

Throughout this work, the standard definition for axes of reference and number of data points per axis is as follows:

- *Order (Way)* – number of variables indexing a data point in a tensor. E.g. Vector is a 1-way tensor or 1st order tensor, a data structure indexed by only one variable. Matrix is a 2-way tensor or 2nd order tensor indexed by 2 variables. A 3-way Tensor or a 3rd order Tensor is a cubical data structure (block), called a Block Tensor. This is the data structure used for experimentation in this work.
- *Modes* – the number of varying parameters in a tensor block. These make up the orders or ways of a tensor structure, showing the various information streams.
- *Dimensions* – the actual number of data points with respect to each variable (nth order or way). The higher the dimensions along a mode, the larger the size of the data tensor along that order.

4.3.1 Knowledge Representation

Multi-modality is the high level representation of abstract data about different aspects of reality. Analyzing such data flows helps record different aspects of the real-time event(s) across various indices like space, time, class, hierarchy, etc. It will result in 'Multimodal Fusion' (deeper info about the events under study), or 'Interdependence Analysis' (relationship between the indices) [100].

Complex system data streams can either system outputs, observation parameters, state vectors or input variables. These sources generate data streams which act parallel in existence. Irrespective of the arithmetic units, they all exhibit moving trends and characterize source operation. When all streams originate from elements of the same super-system, their trends hold correlation and can be mapped out across modes or axis of occurrence. These global pattern maps characterize the overall super-system operation. Local features within these global pattern maps offer the simplest dictionary in a classifier to perform behavior grouping of streams or information flows.

Multimodal thus doesn't refer to multiple sources, rather symbolizes the parallel data streams. Knowledge representation in higher order high dimensional data construct is a crucial element in designing CPS. During this process, raw data is modeled with an explicit description of the structured patterns among data. Different types of models are formed in which groupings of data patterns are very important. These groupings exhibit the subtle dependencies among the information sources.

4.3.2 Challenges and Constraints Involved

Data Assimilation can be achieved across dimensions, modes or even blocks of data. Practical constraints that act up are:

Non-Negativity, Semi Non-Negativity, Orthogonality, Semi-Orthogonality, Reducibility, Sparseness, Smoothness and Non-Correlated nature. For high dimensional data, other crucial processing constrains are outlier misclassification and curse of dimensionality.

Constraints such as these will control the iterative non-linear projection (shrinkage) or filtering tendencies of behavior variations. Hence they should be modelled into suitable regularization of the penalty terms to achieve the desired level of performance [52].

Sparseness is the pivotal feature that requires attention while modelling Big Data processing solutions, increasing data dimensions, escalating the computational complexity along higher modes. It affects the local pattern detection accuracy as only limited number of components will exhibit meaningful patterns. Therefore it may be termed the predominant condition affecting Feature selection stages. Sparseness control is Big Data solutions is designed as a cost function in terms of regularization / penalty terms. Non-Negativity is another trait affecting the physical meaning of the time series data, as the options for representing anomalous behaviors narrows down to continuous measures like probability distribution.

4.3.3 Multiway High-Dimensional Data Processing

Multiway High-dimensional data streams originate from complex heterogeneous distributed systems, a typical example being SG-CPS. Such sources necessitate multi-Linear data models which require more compact representations than dimensionality reduction for preserving the knowledge. Those in turn feature denoising, size reduction or compression, and transformations. The main rule formulated is to unravel the dependencies, by defining and detecting between and among various modes. Such an operation can be interpreted as a form of Spatio Temporal Decorrelation. Significant contributions of such an approach are:

- Acceptable performance levels especially for heavily sparse data.
- Rigorous proofs for global convergence
- Context-significant interpretation of latent behavior components (unknown source interrelation).

Multivariate Data Analytics is valuable due to the ability to slice and dice the incoming data blocks in several ways so as to enable anticipation of all possible outcomes. Additionally, comparing metrics over quarterly, monthly, weekly or other time frames that align with the Meta program, is an analysis exercise that can determine the impact of your data input patterns and characteristics. Watch metrics over time and trends will appear, providing even more valuable insight.

4.4 INTRODUCTION TO TENSORS

Tensors embody naturally occurring data clusters that represent numerous data sources and readings at the same time instant. The arithmetic quantity that deals with such data structures of more than 1 mode, indexed by more than one variable are called Tensors. Generalized as Vectors and Matrices for 1 and 2 dimensions, a tensor model proposes richer and likely demonstration of numerous facets than a ‘flat’ matrix. They are referred to as ‘higher-dimensional matrices’. A decomposition or factorization operation on a Tensor is analogous to Matrix factorization extended to higher orders. Tensor decompositions define multilinear models which are more versatile than linear models, yet easier to work with than non-linear models. The main functions are to capture, search, and cluster, classify, assimilate, merge, and process data within an acceptable lapsed time, paving way to new innovative solutions and technologies. Many new decompositions have emerged when using tensors [101]. They enable more compact sophisticated models to capture more complex interactions and manage, visualize, merge, assimilate, search, classify, cluster, and process data in a

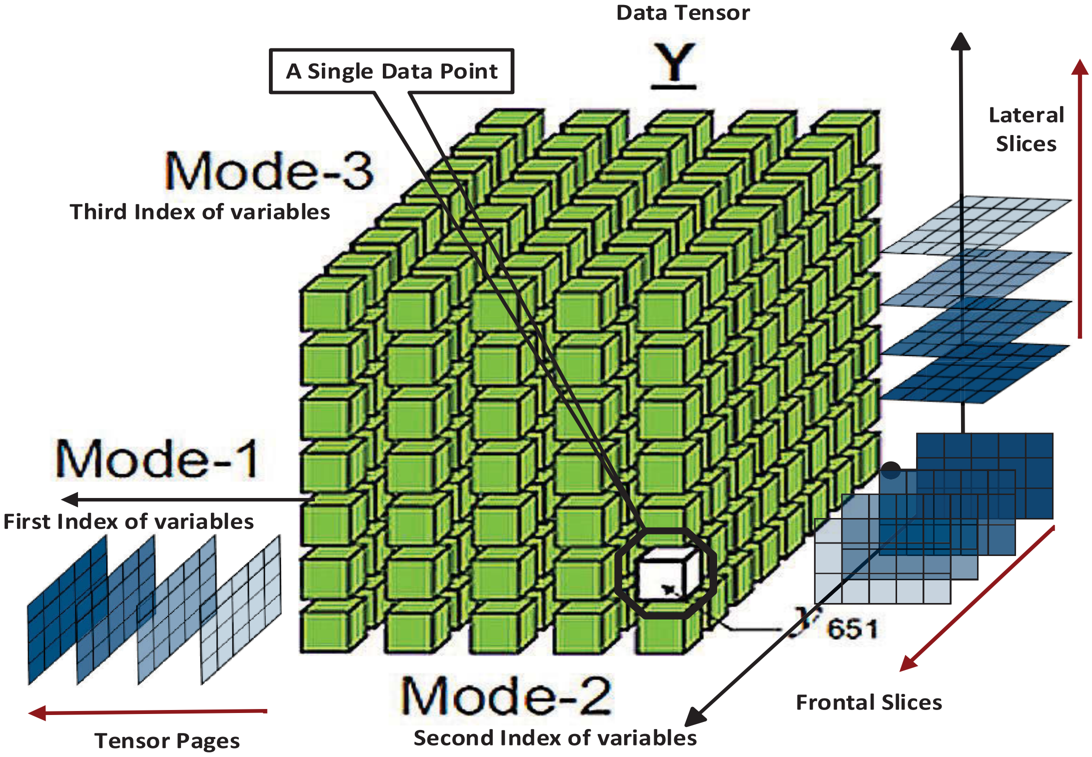


Figure 4.1 Significance of Multiway arrays or tensors in Higher order Big Data Processing

bearable lapsed time, paving way to new innovative solutions and technologies. Many new decompositions have emerged when using tensors [101]. They enable more compact sophisticated models to capture more complex interactions and couplings. Hence, limitations projected by the traditional pairwise interactions created by matrix operations is lifted.

As in Matrix algebra, rank of a Tensor, a rough approximation, is computed repeatedly until error is minimized. "Each factorization / decomposition model gives a different interpretation for Tensor Rank [2]. Large scale processing based on assorted components and device setups can be easily represented in compact multiway array forms, extended into higher orders. It enhances efficiency salvaging computational resources and time. Therefore, Tensor algebra fit well into Multiway Data processing for Cyber-Physical systems.

4.4.1 Benefits of Tensor Algebra based Data Processing

Applying Tensor modelled data to generic networks elevate them by at least an extra order, facilitating *"ambient information extraction"* with the following advantages:

1. Tensors are extensions of the matrix domain, so there are no spatial or locational inhibitions. Purely arithmetic values compute relative behavioral modals. They append multiple variables, and seemingly non-correlated variation trends and remove geographic and displacement constraints.

2. Tensors, being coordinate-independent, relates arrays computed in one coordinate system to that computed in another one, signifying relationships amongst various coordinate structures.
3. A Tensor transformation takes the form of a covariant and/or contravariant transformation law, which determines the type (or valence) of the tensor. They are freed from the limitations of a fixed Cartesian coordinate system, allowing heterogeneous data streams to be appended.
4. Tensorial patterns are also a critical factor impacting resource utilization, so format of constructing tensor patterns for traffic data is very important. They present a higher performance than other intelligence computation methods to a certain extent.
5. High performance and cluster computing, GPUs and GPGPUs, TPUs and more sophisticated technologies are becoming popular with time. This leads to CAPEX and OPEX measures of Tensor-based Pattern Processing being minimized for multivariate multimodal data streams with time.
6. Scalability is another important reason for the growing interest in Tensors. Tensor decomposition algorithms successfully parallelized efficiently in GPUs and Multicore Processing. Recently Apache REEF is also being popularized (a distributed framework originally developed by Microsoft). Initial results are favorable in speed and accuracy measures, and implementations in distributed systems lead to algorithms that scale to extremely large data sets.
7. The most common and difficult-to-solve "Curse of Dimensionality" may be easily salvaged by Tensor Decompositions.
8. But Tensors have an upper hand in capturing higher-order relationships, as they do not produce local optima, under reasonable conditions. Their ability to model multi-way relationships towards global updates helps uncover hierarchical structures in high-dimensional data sets.

Matrix operations such as factorization helps represent the various traits and characteristics of the elements and the global features in planar sources. If big data may be visualized as data streams from spatially-temporally separated multiple sources, paralleling them and processing can also be extended along spatio-temporal dimensions.

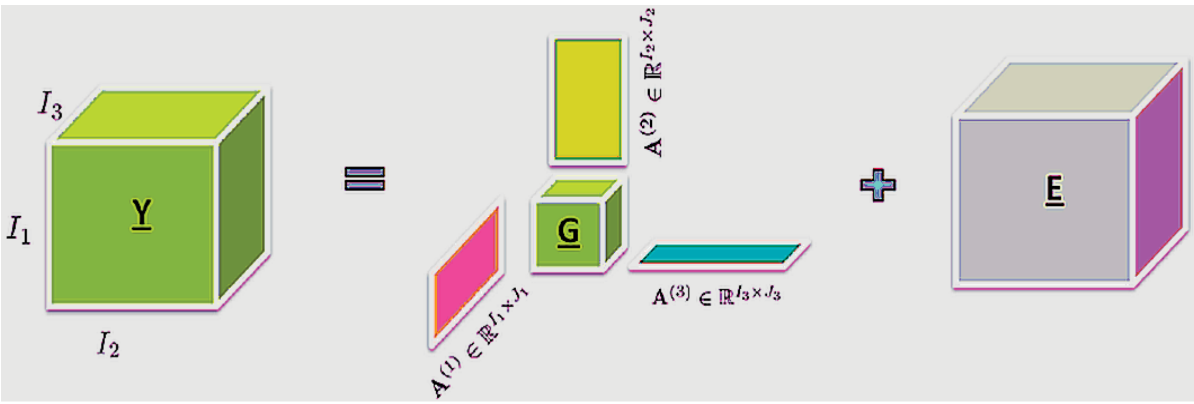


Figure 4.2 Generalized Interpretation of Higher Order Data Decomposition Operation

Anomaly Pattern Detection based Information Processing for Cyber – Physical Smart Grids also *elevates the insitu security of the system.* Examining patterns between any three (or more) dimensions in data sets is ensured. The challenge is not multiway data analysis using Tensors. It can easily be related to a generalized Matrix computation. The difficult parts in designing the computation are listed below, which are scrutinized throughout the Implementation section.

- Preparing the data to fit in with the math structure that will undergo the decomposition.
- Identifying physical significance of the decomposed elementals and correlate them with the physical problem aspects.
- Additional objectives imposed like discrimination, statistical independence, smoothness, dimensionality reduction, sparseness, etc. and their effect on the accumulated data.

4.5 RESEARCH PROPOSITION

Conceptualization of a time series transformation into an alternative representation is a relatively simple task. It is an important step in knowledge discovery as well. The descriptive modelling task is called pattern discovery, to identify frequent associations within data. Difficulty lies in the preprocessing stages. In fact, the success or failure of the entire information extraction process lies on the efficiency of data formatting to satisfy showcasing of different information trends.

The scope of this work is to ascertain the validity of single time-window analysis. Extending the work to Streamed operations will be suggested as part of future directions.

4.5.1 Overview

At any point of time, in a heterogeneous network, the multiple data and signal flow streams represent the operation state or system state. So patterns of variations are rather marginal variation trends. Associations across multiple axes along which heterogeneous data streams are drawn out and Scattered along one or more modes. Packing of this these data streams will be represented by inter-modal factor matrices. The decomposition or factorization of such a structure extracts abstraction of the information content along a single, dual or multiple axes representing a combination of variables. This is precisely the requirement for autonomic data processing.

4.5.2 Multi-Way Pattern Processing: Behavior-Anomaly Detection

Complex, high-dimensional Time-series data has unique properties posing analysis and modelling challenges. Our aim is to extract anomalous behavioral patterns from unconstrained data in smart energy grid context. A spotlight aspect of Tensor Factorizations considered in this work is their ability to *"Map out Anomalies".* With respect to all system metrics, a standard system response and behavior may be mapped to input data characteristics. Any change in such mappings would be recorded as *Behavior Patterns* and if recognized as out-of-norm they become *Anomalous Patterns.*

Semantic associations among the various modes is represented as *modal overlaps*. It qualifies as a similarity index. Hence joint pattern sets of common indexing shows a class of similarity measures, characterized by co-occurrence. However, the numerous facets induce modelling demands, as a result, most of the prior work focus on specific points, like applying two-mode relations.

Instead of the conventional device operation, the system can now subject to behavioral modifications based on these *incoming data patterns*. The classes of patterns corresponding to the *spectrum of system responses* which directly relate to management of the equipment.

4.5.3 Solution Framework

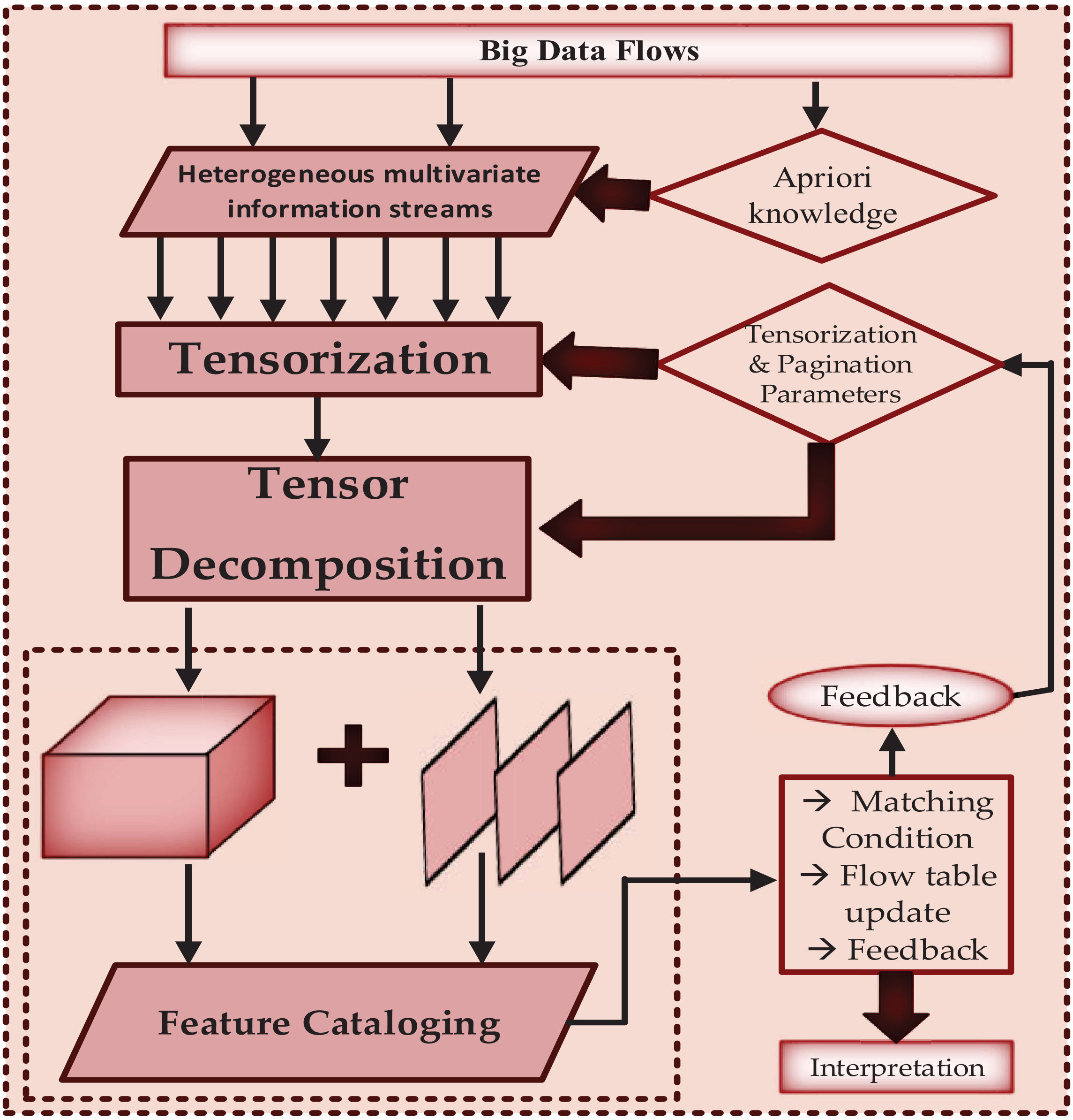


Figure 4.3 High Level Description of the Research Proposition

The proposed solution to the hypothesized question is framed based on Tensor algebra. The crux of this arithmetic operation is to compose data into higher order data structures, packing all the information content as a pre-processed multiway data tensor, and extract patterns and core tensors (congruent matrix core elements and factors).

It is corroborated that decomposed Tensor based Pattern extraction and classification enables tracking and recording Behavioural and Anomaly Patterns. Tensor Factorizations eliminates Large Scale data difficulties: no storing and processing of large data or factor matrices required, as the core tensor comprises of all the essential information that is an abstraction of the entire data input volume.

The decomposition operation extracts linear or higher order factors of varying dimensions. Data Tensors, subjected to normalization and cleaning, undergo factorization and multimodal pattern detection. These can be stored, for inference or further application and test-processing. Thereby, full-scale computations within the system are saved. Apriori knowledge is useful for preprocessing Big Data, to appraise which features or properties may be interpreted in a hidden or latent component. The results can drive a data-pattern processing platform. The tensorized multiway data structure and the High-Level description of the proposed solution are given in Figure 4.1 and 4.2 respectively.

Factor definitions are defined by mathematical relationships of the data streams that are under scrutiny. In this experimental setting, the dependencies under study will be the various *temporal and spatial cross-variable associations,* where a *variable refers to a time-series data behavior from a system component.* Factor matrices are then globally updated with normalization rule at each stage, to allow stream processing and tracking of Behavioral patterns [102].

4.5.4 Modelling Data as Decomposed Tensor

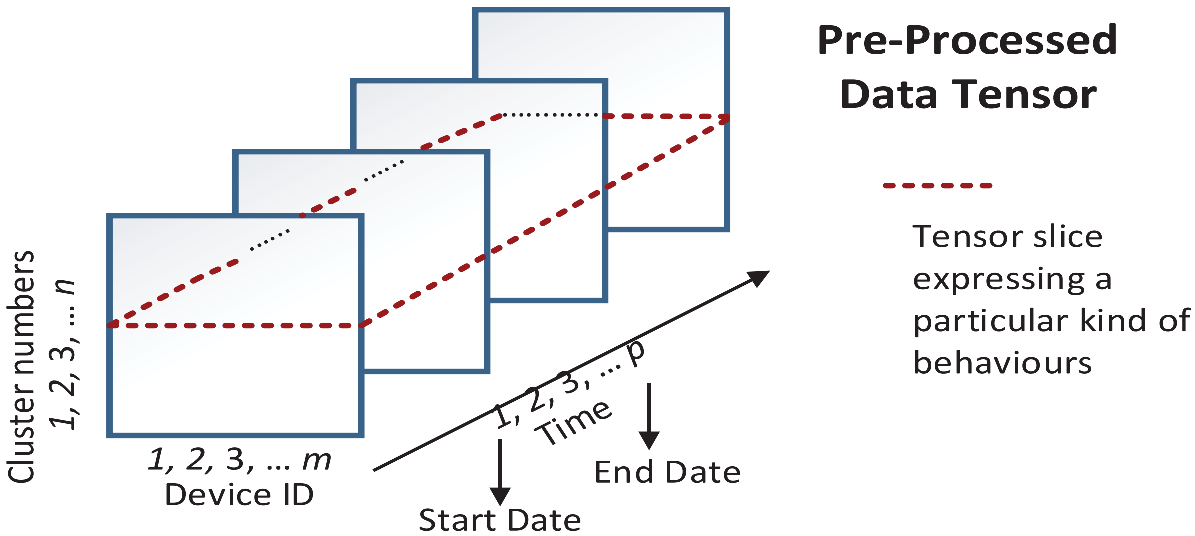


Figure 4.4 Proposed Knowledge Representation to facilitate Multiway Anomaly Pattern processing

Studies so far pointed out that, Tensor modelling of Multivariate data streams is the most effective tool in scaling down time and resource consumption for Big Data processing. Experiments have inferred how chunks of data can be given a 'birds-eye-view' to obtain large scale trend information, bypassing computations involving every single reading [103, 104]. This is the key to framing the research solution of the problem at hand. State-of-the-art Big Data processing in parallel streaming mode, through multiple layers of information, require much more power than standard ML, optimization and heuristic modules. This calls for tools that will go beyond the generic dimensions, which brings this investigation to Tensor Mathematics. It will help achieve tracking the Stationarity independent real-time process.

4.5.5 Anomaly or Outlier Patterns

Behavioral patterns in information (data, signals and system outputs) illustrate system operation in terms space-time-index variant functions. So a deviant pattern or even a single factor element constitutes the anomalous event. Detection of outliers refers to monitoring and flagging unusual events for investigation. In the data mining task of anomaly detection, popular approaches are distance-based and density-based [105, 106], such as Local Outlier Factor. K-nearest neighbor selection method was the mostly preferred distance measure to label observations as outliers or non-outliers. Anomaly and behavior patterns can be parallelized at a logical node of a super-system which represents the state of a particular system model. They can be measured as 'co-occurrence' degrees of patterns. Detecting co-occurrences of particular hyper-dimensional event in terms of patterns plots the trends of unprocessed data streams. The time window of early detection than a series of deterministic positive load compliant streams. Similarity measures made in tensor structures [107] are also inspirations.

4.5.6 Feature Annotation of Anomaly Patterns

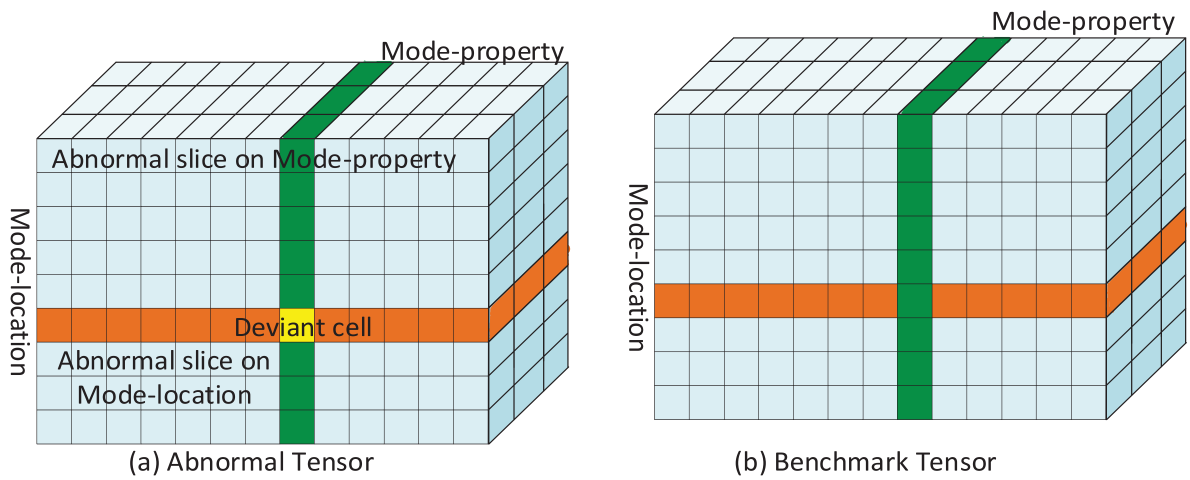


Figure 4.5 Anomaly detection using Tensors as Data Structures

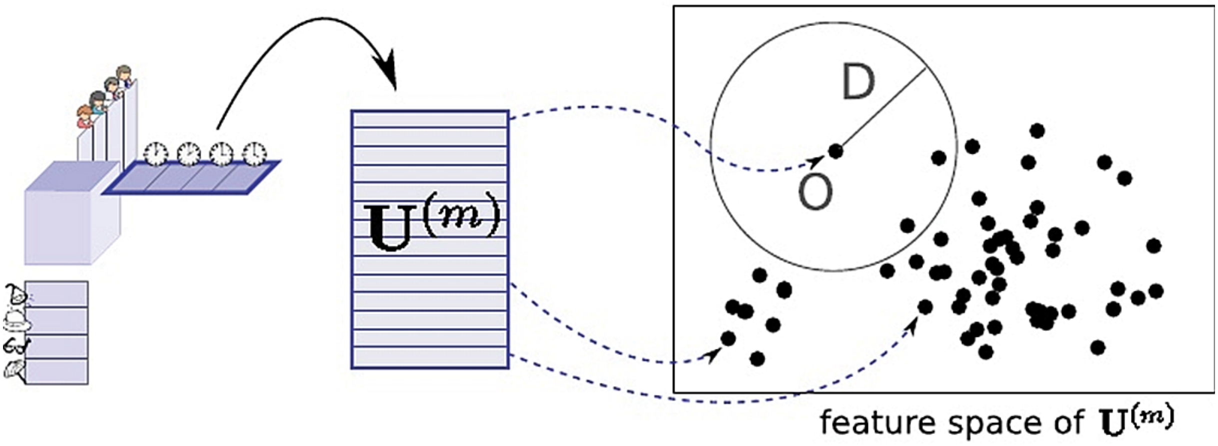


Figure 4.6 Tucker Decompositions for anomaly detection algorithm in hyperspectral images [112]

The best algorithms designate individual instances of an event description whose category is to be estimated using a feature vector of the individual. The vector is derived from measurable properties of the instance. Each property is termed a feature index, also known in statistics as an explanatory variable (or independent variable, although features may or may not be statistically independent). Typical examples are Binary [108], Categorical [109], Ordinal, Integer-valued, Real-valued [110], and Vector based [111] definition. For images as instance, the feature values often are the pixels; likewise if the monitored data is a text document, the feature values will be frequency of occurrences of different words. Certain algorithms operate only on discrete data and require real-valued or integer-valued datasets to be discretized into subgroups.

The proposed method dynamically generates the patterns or feature annotations *'in-situ'***,** that is within the context of the given investigation problem and the data sources. The factors vary in accordance with time-dictated behavior of the CPS elements. The core tensor factor might hold congruent properties but they too may differ in values based on the scale of data behavior variations. Thus a temporally updated or recorded database of these factors gives us the annotated feature set, or reference Feature Pattern Table.

4.5.7 Anomaly (Feature) Detection condition

The Anomaly condition here may be defined as a tolerance value, a mask along 1 or more dimensions, or a particular attribute like histogram. Anomaly Detection will help balance the relatively large processing delay with respect to the event occurrence and ingestion timestamps, and system sensitivity. This is one question currently being worked upon, and is completely application-domain specific. The basic experimentation may be started from statistical vectors or noisy spikes. Details of this approach are covered in the Experimentation chapter.

4.5.8 Applications: Context of the Smart Grid Case study

In our proposed case study, the obtained pattern spectrums classified as normal behavior and anomalous patterns, are the *key enablers of future smart grids*. They make energy efficient integration of renewables, in addressing the energy challenge. Integration points between the component systems which are well defined, standardized and agreed between the designers of the components.

The logical system node can be represented as a 3-layer hierarchy – Data, Signal (Communication and Control) and Physical (the primary system variables, i.e. Power and Load)

Table 4.1 Contextual Anomaly Detection in Smart Grid elements - Survey (C. Alcaraz, L. Cazorla et al. 2015)

	Complexity	Speed of classification	Speed of learning	Handle parameters	Comprehensibility	Accuracy	Learning from observation	Control - interdep. data	Control - missing data	Control - redundancy	Control - noise	Control - subtle changes	Control - drastic changes	Incremental learning	Unlimited networks	Limited networks	Requirements					Control - Corporative net	Control - RTU/Gateway (	RTU/Gateway - sensors (S	Gateway - embedded dev.	Embedded devices (HAN)
	‡[a]	‡	‡	*‡	*	*	*	*	*	*	*	◦	◦	◦			R1	R2	R3	R4	R5	*◦	‡*◦	‡*◦	◦	‡
Classification trees	-	✓	✓	✓	✓	-	-	✗	-	✗	-	-	-	-	✓	•[b]	✓	✗	-	-	-	•	-	-	-	✓
Regression trees	-	-	-	✓	-	-	✓	✗	✓	✗	-	-	-	✓	✓	•	✗	✗	✓	✓	✓	•	-	-	•	-
Association rule learners	✗	-	✓	✗	✓	-	✓	✓	-	✓	✗	-	-	✗	✓	✗	✗	•	✗	✗	✗	•	•	-	-	-
Clustering	✓	✗	✗	✗	✓	✗	✓	✗	✓	✓	✗	-	-	✓	✓	•	•	✗	✓	✓	✓	•	-	-	•	•
Parametric and non-parametric	✓	-	✗	✗	✗	✓	-	-	✓	-	✓	✗	-	-	✓	•	•	✓	•	•	•	✓	✓	•	•	•
Operational models	✓	-	-	✗	✗	✗	-	-	-	-	-	-	✗	-	✓	✓	✓	✗	✗	✗	✗	-	-	-	-	•
Smoothing	✗	-	-	✗	✗	✓		-	✓	-	✓	✓	✗	-	✓	✗	✗	✓	•	•	•	✓	•	-	•	-
Markov chains	✗	-	✗	✗	✗	✓	✓	-	-	-	✗	-	✗	-	✓	✗	✗	✓	✗	✗	✗	•	•	-	-	-
Artificial neural networks	✗	✓	✗	✗	✗	✓	✓	✓	✗	✗	✗	-	-	-	✓	✗	✗	✓	-	-	-	•	•	-	-	-
Bayesian networks	✗	✓	-	✓	✓	-	✓	-	✓	✗	✓	-	-	-	✓	✗	✗	•	-	-	-	•	•	-	-	-
Naïve Bayes networks	-	✓	✓	✓	✓	✗	✓	✗	✓	✗	-	-	-	✓	✓	✗	✓	✗	•	•	•	•	-	-	•	-
Support vector machines	✓	✓	✗	✗	✗	✓	✓	✓	-	✓	✗	-	-	✗	✓	•	✓	•	✗	✗	✗	•	✓	✓	-	•
Rule-based techniques	✓	-	-	✓	✓	✓		-	-	-	✗	-	-	✗	✓	✓	✓	✓	✗	✗	✗	•	✓	✓	-	✓
Rule learners	-	✓	✗	✓	✓	✗	✓	✗	✗	✗	✗	-	-	✗	✓	-	•	✗	✗	✗	✗	•	-	-	-	-
Nearest neighbour	-	✗	✓	✓	✗	✗	✓	✗	✗	✗	✗			✓	✓	-	✗	✗	•	•	•	-	-	-	•	-
Fuzzy logic	✓	✓	✓	✓	✓	✗	-	-	✓	-	✓	-	-	-	✓	✓	✓	✗	-	-	-	•	-	-	-	✓
Genetic algorithm	✗	-	-	✓	-	✗	-	-	-	-	-	-	-	-	✓	✗	✗	✗	✗	✗	✗	-	-	-	-	-
Knowledge-based	✓	-	-	✓	-	✓	-	-	-	-	-	-	-	✓	✓	•	•	✓	•	•	•	✓	✓	•	✓	-
Information-spectral theory	-	-	-	✗	✗	-	-	-	-	✓	✗	✓	-	-	✓	-	✗	✗	-	-	-	•	-	-	-	-
Anomaly-based IDS															✓	✗	✗	✗	✓	✓	✓	•	-	-	-	-
Signature-based IDS															✓	•	-	✓	✗	✗	✗	✓	•	•	•	•
Specification-based IDS															✓	•	-	✓	✓	✓	✓	•	✓	✓	✓	✓

variables. Their sources may be summarized as below:

1. Network sublayer – Network elements and event-topology (intelligent switches, IP sublayer, including automation signals and protocols), as well as logical power flow model.
2. Data sublayer – Data flow maps of the large discrete big data streams generated, from sensors, actuators and automation units.
3. Control sublayer – Computing, computational intelligence, processing control and optimization.

Apart from this comes the Physical Layer where the actual super-system quantity transformation happens. In the smart grid scenario, sensor-actuator units, scalar networks, support systems, generator plants, storage units, pure electrical loads, etc. are the components here moderating electric power flow and Load management.

4.6 CHAPTER SUMMARY

Tracking dataflow helps predict operation phase and system state in anticipation to out-of-norm events. Apart from the actual system generated output quantity, such as power or network data, other component status and control signals maybe monitored through outlier detection in data streams. Thus, it is recognized that a collaborative integration design of the existing smart grid-power system components will create advanced self-sustained configurations. In order to reach there, need for multi-way high-dimensional data processing has also been proved.

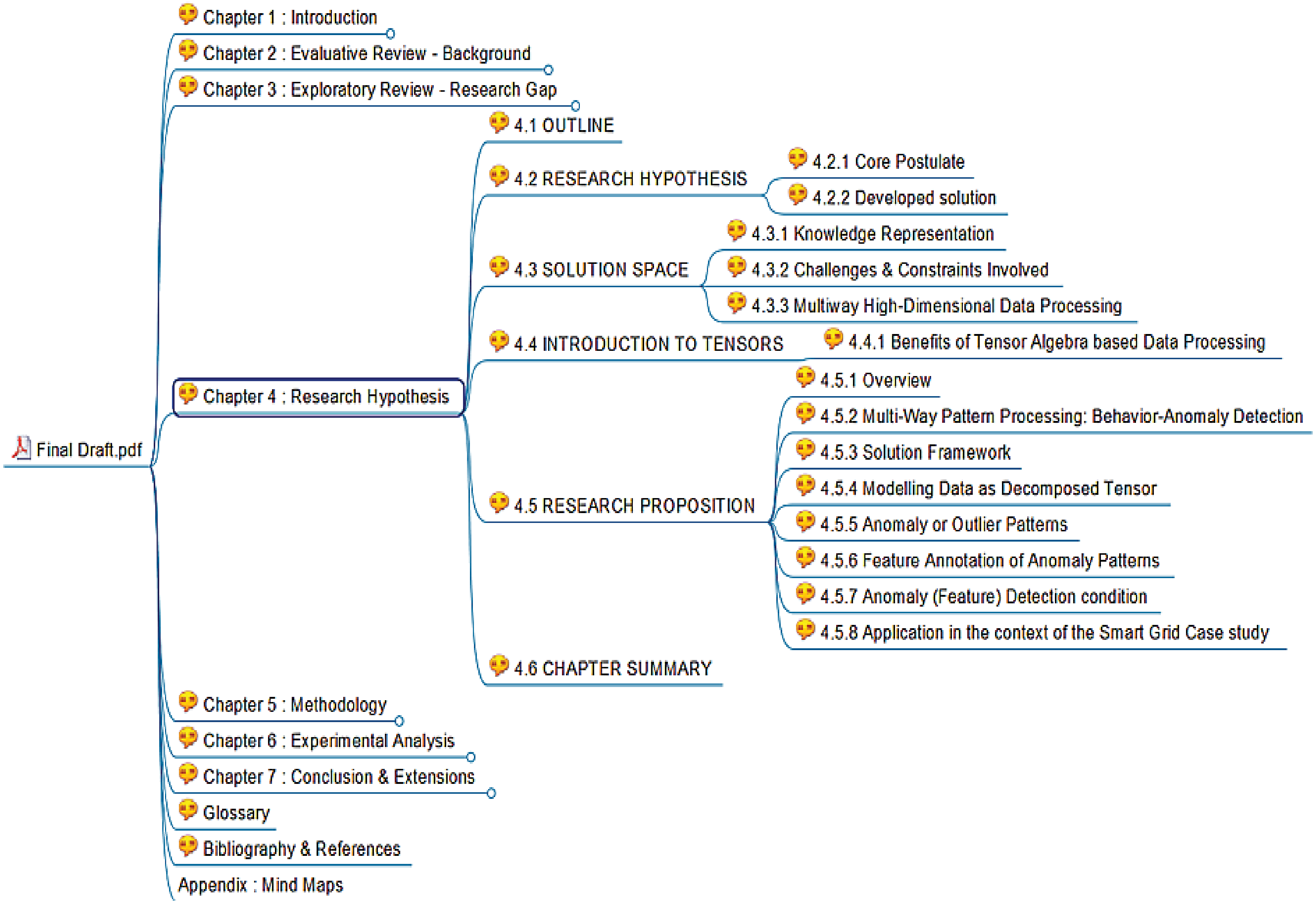


Figure 4.7 Mindmap outline of Chapter 4: Research Hypothesis

Chapter 5 : Methodology

5.1 OUTLINE

This chapter is the instrumental Review to evaluate the possible solution strategies, in terms of related subject areas, concepts and implementation. The various outcomes in parallel from other topics and subject domains are studied, analyzed for effectiveness, and matched with current research objectives. Finally, suggested Methodology and experimental framework is outlined.

5.2 TENSOR FACTORIZATION and DECOMPOSITION

Tensor Decompositions and Factorizations are a branch of emerging scientifically computation low-rank Tensor approximations. This paradigm which had been actively studied upon for a couple of decades now. However, only recently has it gained enough momentum along the Big Data Processing and Analysis stream.

5.2.1 Knowledge Discovery: Significance of TF and TD

Natural data occurring in complex systems are multi-way high-dimensional data structures. Characterizing them and evaluating possible interpretations depends on the order and dimensions of data collection. This means the need to reduce parameters to be evaluated, as achieved by multi-way, decompositions easily. All subscale possible inter-component and inter-modal relationships can then be retained in multi-linear data models. An added plus-point of Multiway Array or Tensor factorization is the dimensionality reduction for Big Data.

There is a multitude techniques for exploring and mining deep patterns in series data. Two of the most common and most widely applicable notions are:

a) *Component Analysis (e.g. PCA or ICA based transformations:* [113] Such techniques are based on finding coordinate axes independent of one another for feature representation; so are unsuitable if dependency status of the basis components is unknown. They also avoid redundancy by scaling dependencies down to the single features, missing out higher order latent associations.

b) Spectral Analysis (e.g. Wavelet Transforms): [114] These methods processes data as projections thrown over individual frequency components that are independent of remaining components.

This means that the above-mentioned techniques fail for interdependent latent basis components, which now highlights the research problem in focus. Tensor Factorization or Decomposition has the advantage of performing elemental extraction along both those aspects minus the assumptions, which ensures their superior performance over both classes of Feature detection

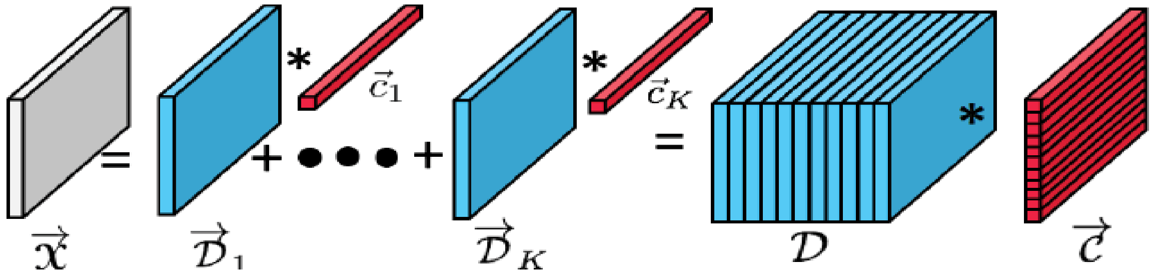


Figure 5.1 a single Tensor signal is a linear combination of tensor dictionary atoms, named tensor fibers or tubes, representing a particular information flow; in aggregation they show intermodal correlation

techniques. The different ways to pack texture set information into tensor structures has been studied based on TPE [115], by separate DC and AC components. Similarly Tri-factorization methods also are significant in this regard (simultaneously cluster runs of rows and columns of the input data matrix X) has been proposed for 3-factor non-negative matrix decomposition [116].

Tensor decompositions also hold importance as energy revealing factorizations. They help optimize number of computations, processing power and even on-time of the processing platform. Tensor models also isolate interrelations not delineated in the space-time (2-way) domain.

Practical TF/TD operations are not always perfect, i.e. they are lossy. 100% FIT is hard for real world data (noise in innate, with limited number components). However, Signal-to-Noise ratio may be improved by multi-layer approaches (decomposed the residual tensor till desired FIT is achieved)

5.2.2 Physical Interpretation

Approximative low-rank Factorizations play a central role in augmenting the data and mining the latent dataset components. The role of TF is to distil out pattern occurrences and activation maps about the feature occurrences. The extracted components are known as Multiway components. The factors hold the significance by Basis components and Activation maps which signifies positions and magnitudes of basis components, the subscale relation-patterns and their weighted occurrence. For a standard factorization model given by the equation:

$$Y = AX + E \;\; or \; AB^T + E \tag{5.1}$$

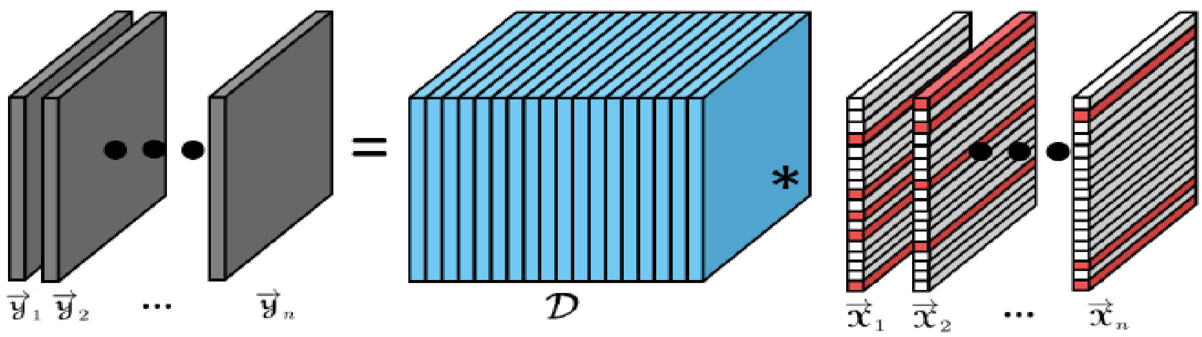


Figure 5.2 Data Tensor represented as tensor columns, a T-product of tensor dictionary and tubal sparse coefficients (Red and white labeling - zero and non-zero tubes)

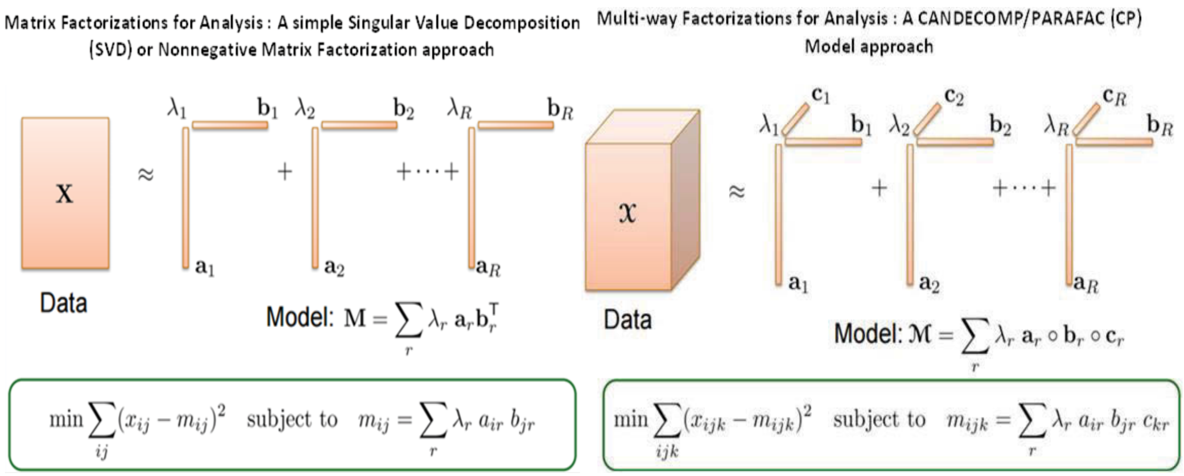


Figure 5.3 Physical significance of Higher Order Big Data Factorizations, as linear combination of basis components (order 'r'; i, j = 1 to r)

Where; $X = B^T$ signifies the basis components which are the hidden dependency sources, and A represents the activation map, interpreted as the weighting coefficients showing the strength of each dependency trait. E is the standard reconstruction error which gives the tolerance threshold for the composite data structure. The obtained Factor parameters and metrics provide significant interpretation in terms of the problem domain. Few examples from literature are: number of devices,

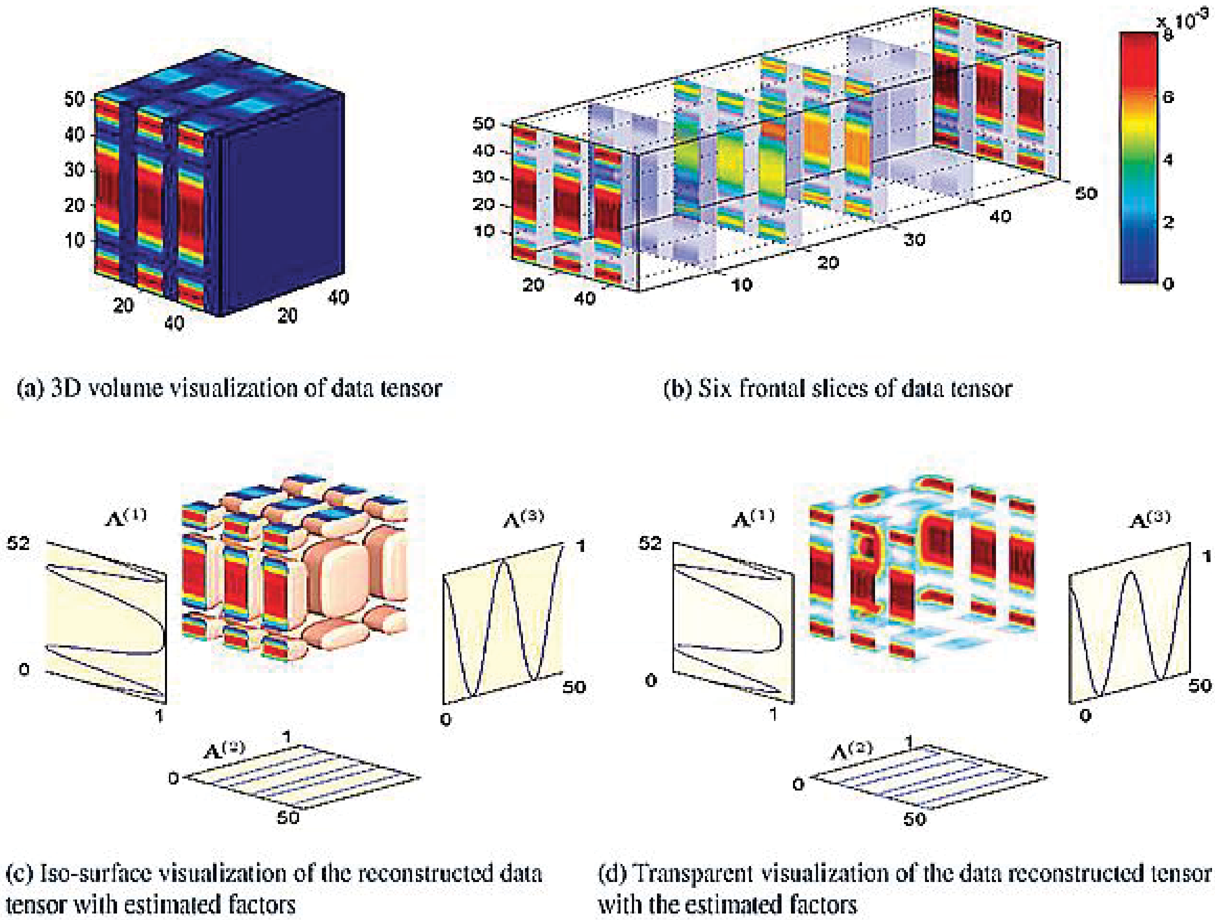


Figure 5.4 Nonnegative Tensor Factorization of a rank-1 third-order tensor, employing different visualization modes

coherence duration, bandwidth measurements or frequency bins, etc.; these are the parameters in terms of which decomposed components help interpret the extracted patterns systematically.

TD/TF thus gives better results in Multimodal Signal Processing, as *intermodal patterns* become enhanced and exposed. It has been proven that by including extra dimensions investigation of time, frequency and index variable patterns is possible in one analysis. Ensuing components describe the Space-Time-frequency signatures and the comparative influence from different indices, by increasing the tensor order beyond 3.

Further operations performed on these components can remove redundancy and attain compact sparse representations. Nonnegative Tensor Factorization of a rank-1 third-order tensor, employing different visualization modes is demonstrated in the article [117]. The physical significance of Higher Order (Multiway) Factorizations of Big Data, as linear combination of basis components is shown in Figure 5.3 [118], and interpretation based on component visualization is demonstrated in Figure 5.4.

Characterizing behavioral anomalies directly contribute to improving sensitivity of the CPS configuration.

5.2.3 Mechanics of Tensor Factorizations

A TD or TF model factorizes a given Nth order tensor into a set of N non-ve component matrices representing the common loading factors. They support partial sampling of a tensor. Standard approaches in general have the following generalized characteristics:

- *Number of observations ~= Number of unknown sources:* Desired algorithm do not warranty estimating all sources. Many methods zero in on the local minima created, especially if not regularized.
- *Number of observations >> Number of unknown sources:* proposed algorithm gives consistent and acceptable results, particularly when sources and mixing matrices are sparse.

This property makes them ideal tools for our problem at hand. The uniqueness of TF stems from the number of components choice. Determining size of the factored data shows the scale of observation for the intermodal patterns along variable combinations.

This technique allows processing on a factor matrix-by-matrix basis rather than the generic vectorial (column-wise) processing. The three most popular decomposition/factorization models for N-th order tensors are the Tucker model and the more restricted PARAFAC model, as explained hereafter.

5.2.4 PARAFAC or CANDECOMP (Canonical Decomposition) or CP Model [2]:

Given a data tensor $\underbrace{Y} \in \mathbb{R}^{I\,x\,T\,x\,Q}$ and the positive index Y, component matrices, also called loading matrices or factors, $A = [a1, a2, a3, \ldots .aJ \in \mathbb{R}^{I\,x\,J}], B = [b1, b2, b3, \ldots .bJ \in \mathbb{R}^{T\,x\,J}]$ $and\ C = [c1, c2, c3, \ldots .cJ \in \mathbb{R}^{Q\,x\,J}]$, which perform the following approximate factorization and its compact form:

$$\underbrace{Y} = \sum_{j=1}^{J} a_j \circ b_j \circ c_j + E = [\![A, B, C]\!] + E \quad \text{Or} \quad \underbrace{Y} = I \times_1 A \times_2 B \times_2 C + E \qquad (5.2)$$

All the vectors $[aj, bj, cj]$ are treated as column vechtors of factor matrices and are linked for each index j via the outer product operator. The aggregated equation, and on an elemental scale is:

$$\underbrace{Y} = \sum_{j=1}^{J} a_j \circ b_j \circ c_j + \underbrace{E} = [\![A, B, C]\!] + \underbrace{E}; \quad \text{And,} \quad y_{itq} = \sum_{j=1}^{J} a_{ij} b_{tj} c_{qj} + e_{itq} 4 \qquad (5.3)$$

The basic PARAFAC model can be represented in compact matrix forms upon applying unfolding representations of the tensor Y:

$$Y_{(1)} \cong A \wedge (C \odot B)^T; \quad Y_{(2)} \cong B \wedge (C \odot A)^T; \quad \text{and} \quad Y_{(3)} \cong C \wedge (B \odot A)^T \qquad (5.4)$$

PARAFAC (Parallel Factorization) or CANDEC (Canonical Decomposition) offers an N-way non-negative tensor to be decomposed into weighted normalized sum of rank 1 tensors.

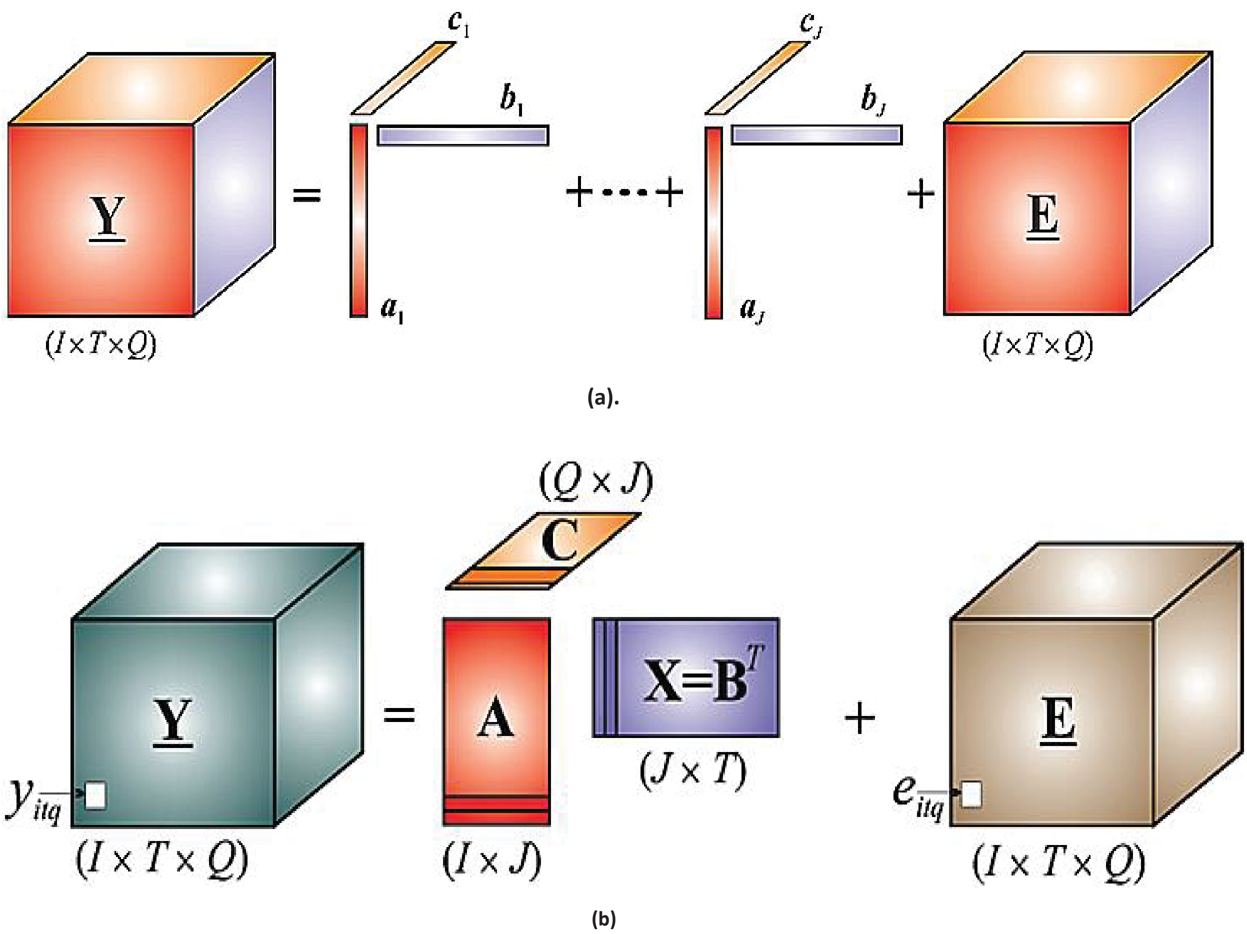
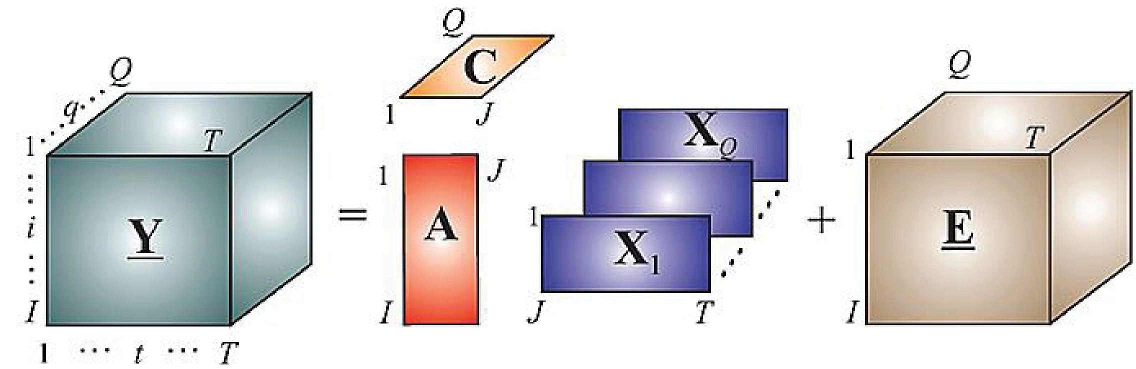
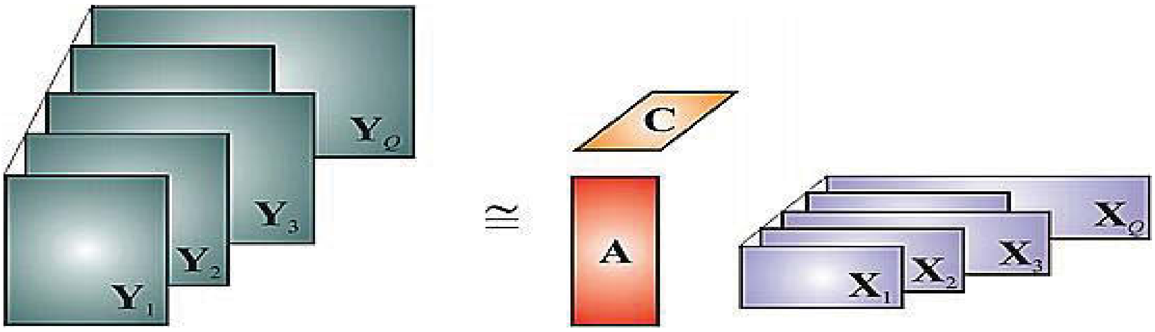


Figure 5.5 3rd-order PARAFAC (a) Standard, and (b) Alternative forms; set of 3 matrices using a scalar representation



(a).



(b).

Figure 5.6 3rd-order NTF1, Standard form; set of 3 matrices using a scalar representation

5.2.5 NTF-1 or Non-Negative Tensor Factorization

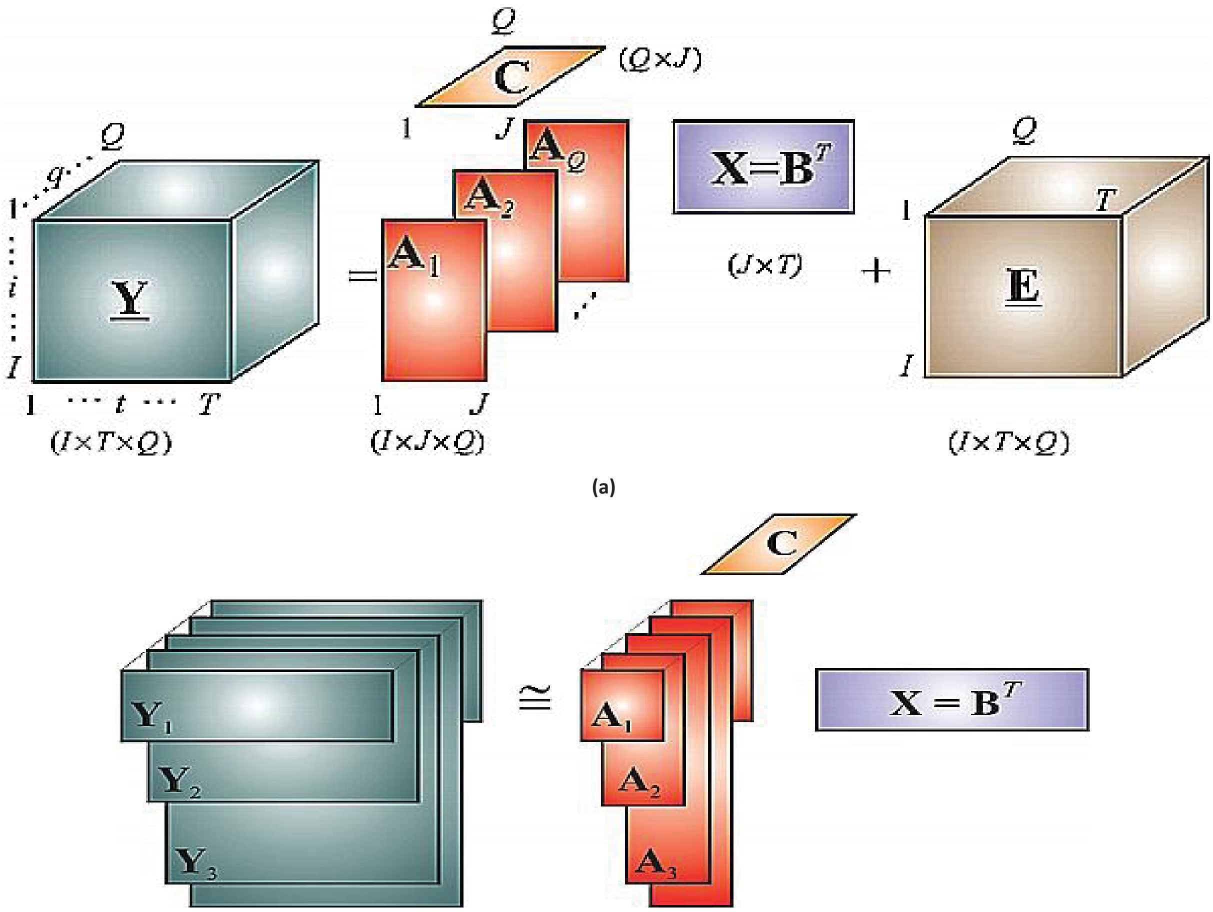
The NTF1 model updates learning rules by minimizing global cost function. NTF is simply PARAFAC model with Non-negativity and Sparsity constraints imposed. It decomposes tensor $\underset{\smile}{Y} \in \mathbb{R}^{I\,x\,T\,x\,Q}$ into a set of non-negative matrices $= [a_{ij}] \in \mathbb{R}_+{}^{I\times J}$, $C \in \mathbb{R}_+{}^{Q\times J}$ $\{X_1, X_2, \dots\dots X_Q\}$, $X_q = [x_{jtq}] \in \mathbb{R}_+{}^{J\times T}$, and $\underset{\smile}{E} \in \mathbb{R}^{I\times T\times Q}$ representing the error content. The Extended NTF1 model for a three-way array is in figure B, applicable where various data subsets have unequal dimensions.

$$Y_q = AD_qX_q + E_q;\ \text{for}\ Y_q \in \mathbb{R}_+{}^{I\times T_q}0;\ X_q = [x_{jtq}] \in \mathbb{R}_+{}^{J\times T_q};\ E_q = E_{::}\mathbb{R}^{I\times T}.$$

The NTF1 problem thus represents a dynamic process by which data analysis is performed iteratively, under modifying conditions, to attain full information about the datasets available, and discuss some minimum structure content in the data framework.

5.2.6 NTF-2 Model

NTF 2 decomposes in terms of lateral slices. The generalized and extended versions are shown in figure, for Matrices$\{X, C\ and\ [A_1, A_2, \dots\dots A_Q]\}$. It extracts irregular frontal slices and 2 components [lateral and vertical slices]. Sparsity is also imposed on the solution by introducing regularization terms. It may be derived from NTF1 by performing Matrix Transposes. Popular NTF applications come under information theory, information geometry, neural computation, convex analysis, pattern recognition, learning, estimating optimization and inferences. Of lately, NTF has been applied to



(a)

(b)

Figure 5.7 3rd-order NTF2, (a) Standard form; (b) Extended model; set of 3 matrices using a scalar representation

several problems from pharmaceutical analysis, food studies, environmental analysis, and in chemistry in general. It offers relatively low complexity for large scale problems. Multiway clustering, Feature selection and Extraction, Multi-sensory or Multi-dimensional data analysis, non-ve neural sparse coding are its chief realizations. However, for tensors of very large dimensions, the above learning algorithm can be very demanding computationally [θ >> 1].

5.2.7 Nonnegative Tucker Decompositions (NTD)

The Tucker decomposition, also called the Tucker3 or best rank (J, R, P) approximation, incorporates 1st, 2nd and 3rd degree intermodal dependencies. Their uniqueness depends on non-negativity and Sparsity. The operation is formulated as follows: For a third-order data tensor labelled $\underset{\smile}{Y} \in \mathbb{R}^{I\,x\,T\,x\,Q}$, and 3 positive indices, $\{J, R, P\} \ll \{I, T, Q\}$; find a core tensor $G = [g_{irp}]$; and 3 component factors $A = [a1, a2, a3, \ldots. aJ \in \mathbb{R}^{I\,x\,J}], B = [b1, b2, b3, \ldots. bJ \in \mathbb{R}^{T\,x\,R}]$ $and\ C = [c1, c2, c3, \ldots. cJ \in \mathbb{R}^{Q\,x\,P}]$, which performs the following decomposition:

$$\underset{\smile}{Y} = \sum_{j=1}^{J} \sum_{r=1}^{R} \sum_{p=1}^{P} g_{jrp}\left(a_j \circ b_r \circ c_p\right) + \underset{\smile}{E};$$ or,

$$\underset{\smile}{Y} = \underset{\smile}{G} \times_1 A \times_2 B \times_3 C + E = [\![G; A, B, C]\!] + \underset{\smile}{E} \tag{5.5}$$

The short notation is given by: $\underset{\smile}{\hat{Y}} = [\![G; A, B, C]\!]$, for $(I \times T \times Q) = (I \times J)\ (J \times R \times P)(R \times T)$

Equivalently in the element-wise form: $\underset{\smile}{Y} = \sum_{j=1}^{J} \sum_{r=1}^{R} \sum_{p=1}^{P} g_{jrp}\left(a_j \circ b_r \circ c_p\right) + \underset{\smile}{E}$ (5.6)

The unfolding approach gives rise to compact matrix forms expressed by the Kronecker products:

$$Y_{(1)} \cong AG_{(1)}(C \otimes B)^T; \quad Y_{(2)} \cong BG_{(2)}(C \otimes A)^T; \quad Y_{(3)} \cong CG_{(3)}(B \otimes A)^T \tag{5.7}$$

And the corresponding vectorised forms;

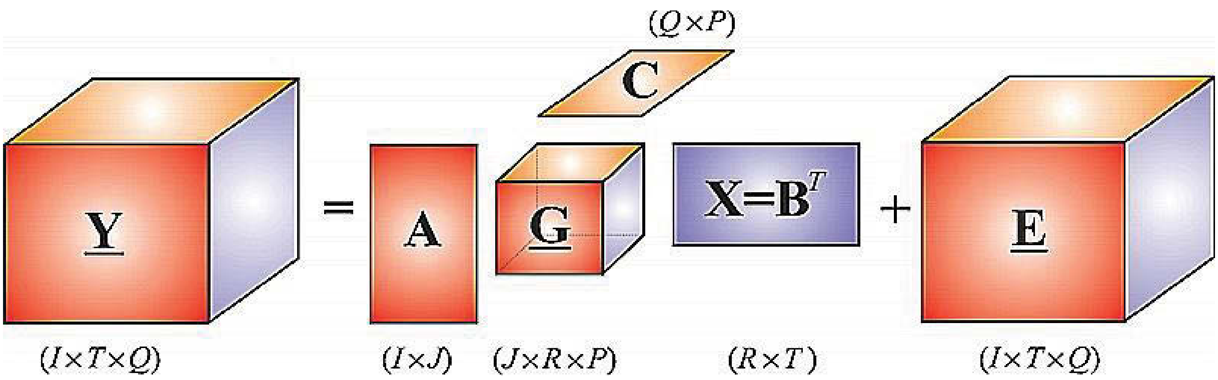
$$vec(Y_{(1))} \cong vec(AG_{(1)}(C \otimes B)^T) = (C \otimes B) \otimes A\ vec(G_1))$$

$$vec(Y_{(2)} \cong vec(BG_{(2)}(C \otimes A)^T) = (C \otimes A) \otimes B\ vec(G_2))$$

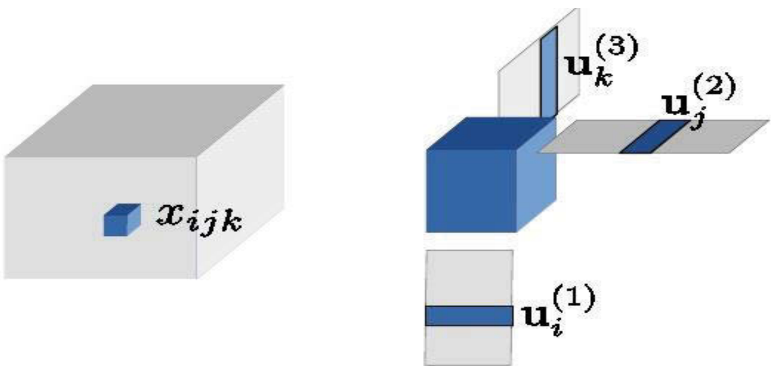
$$vec(Y_{(3)} \cong vec(CG_{(3)}(B \otimes A)^T) = (B \otimes A) \otimes C\ vec(G_3)) \tag{5.8}$$

Summarizing: $\underset{\smile}{Y} \cong \underset{\smile}{G} \times_1 A \times_2$B; short hand notation for Tucker Tensor model.

TUCKER Model comprehends all conceivable linear interactions among components or vectors affecting the data modalities. These interactions are explicitly represented in terms of a core tensor – if core modules are normalized, then presumed to be additive with non-negativity restraints. Tucker allows interaction between modes as well as within each modality, while PARAFAC expresses only intermodal relations [119].



(a).



(b).

Figure 5.8 (a). Overview of Tucker 3 Decomposition model, and (b). significance of the factors

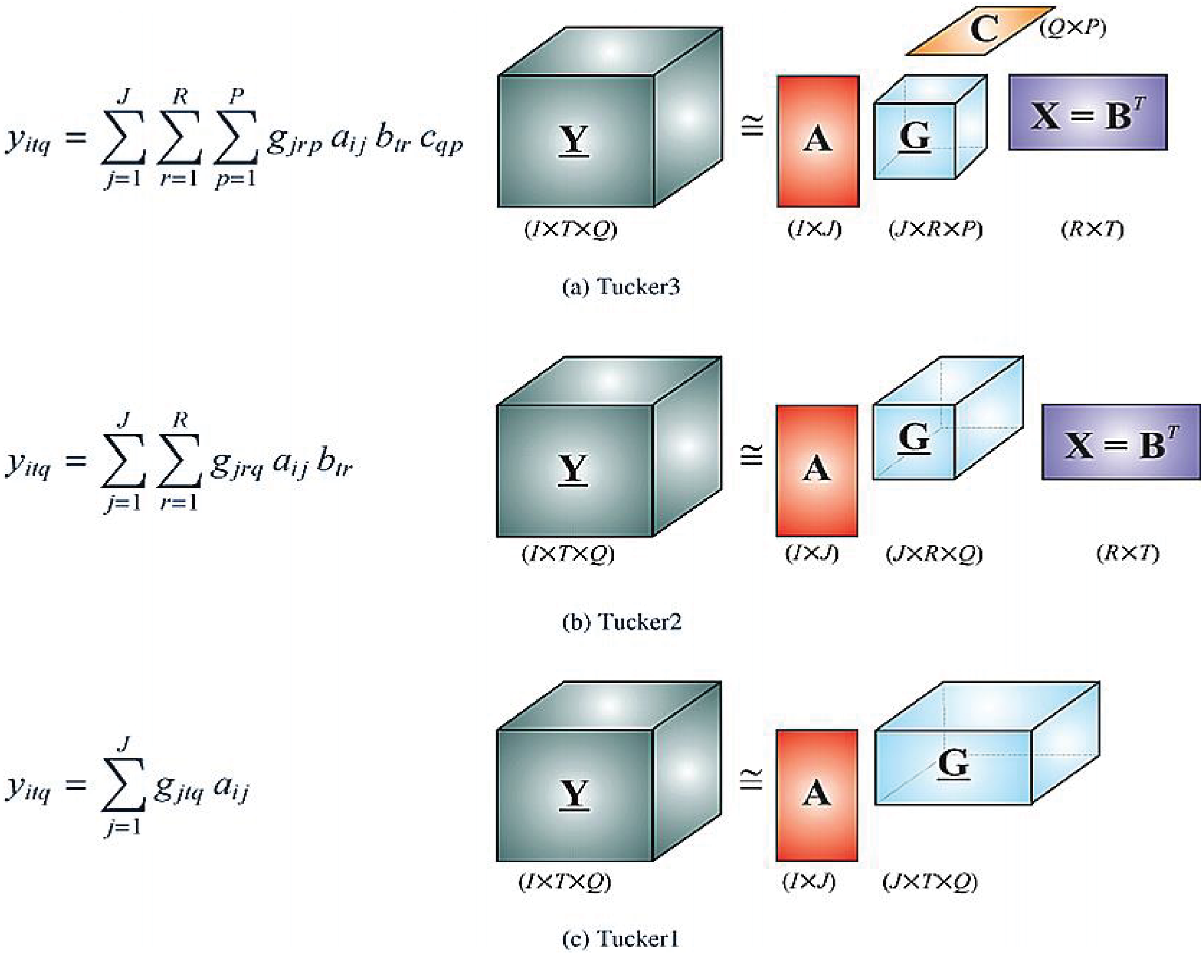


Figure 5.9 Summary of the 3 related Tucker Decomposition models

An interesting fact is the tensor approximation by factor matrices and a core tensor. This helps abridge mathematical processes and decrease computational cost of some operations in multi-linear (tensor) algebra. All components are l2-normalized unit length vectors. Then all coefficients of core tensor effectively express energy of all the original data. Evaluation of interaction amongst the hidden components of the TD model qualitatively defines the subscale patterns of interactions among various sources of information.

The main advantage of TD is how it drastically reduces data samples required for real world data. This is specifically beneficial for time series information in the multi-dimensional dataset compression and sparse representation. Instead for storing the full data, core tensor and the factors now effectively represent the energy content volume and components planes/matrices.

TD is the most effective tool for large scale problems: unfolded matrices which are extremely large requiring to be tackled by multiplicative algorithms and sampling or summations very sparsely (only for some pre-selected entries or tensor tubes). Interactions of a Tucker core are explicitly expressed in a core tensor, and if components $aj\ (n)$ are normalized, then strictly positive for data. This is specifically especially beneficial for time series information, in terms of multi-dimensional data compression and sparse representation. Instead for storing the full data, it can be represented by the core tensor (energy content volume) and the factors (components planes/matrices).

Comprehensive surveys of models and algorithmic aspects of Tensor Factorization (TF) and Tucker Decomposition (TD) has been performed. Other models under consideration are: Higher Order Singular Value Decomposition [117], Sparse Hierarchical Tucker Factorization [120], etc.

5.2.8 Computational Significance

In Regressive vs. Progressive analytical approach, the need to avoid computational challenges introduced by system constraints calls for regressive approaches. It also manifests as overlooking the actual operation of the system under the data inflow, called *System-Off-The-Loop methodology.* If the response time per function cycle of a processor can be bypassed, then most test data sets might as well be matched to the prospective system state-outputs. Actual processing and computation will thus be avoided, leading to drastic reductions in the "ON Time" of the system. Real-time functions are limited to newer off-the-record data sets alone. Thhus, computing performance and efficiency are effectively enhanced through optimizing memory read cycles, execution routines or IO functions.

In short, System-off-the-Loop Computational approach is one where Data inflow, which features the system response and behavior, is analyzed proactively to estimate the response. The prime motive behind this approach is to anticipatory defense against Anomaly attacks. This is the prime motive behind this research: a "reverse engineering" approach, which aims at system behavior modification based on signal patterns. It will work towards realizing the full potential of rapid-growing mobile, network and other cyber-physical applications and services. Instead of a system manipulating data, it is now the turn of Data streams to manipulate the system.

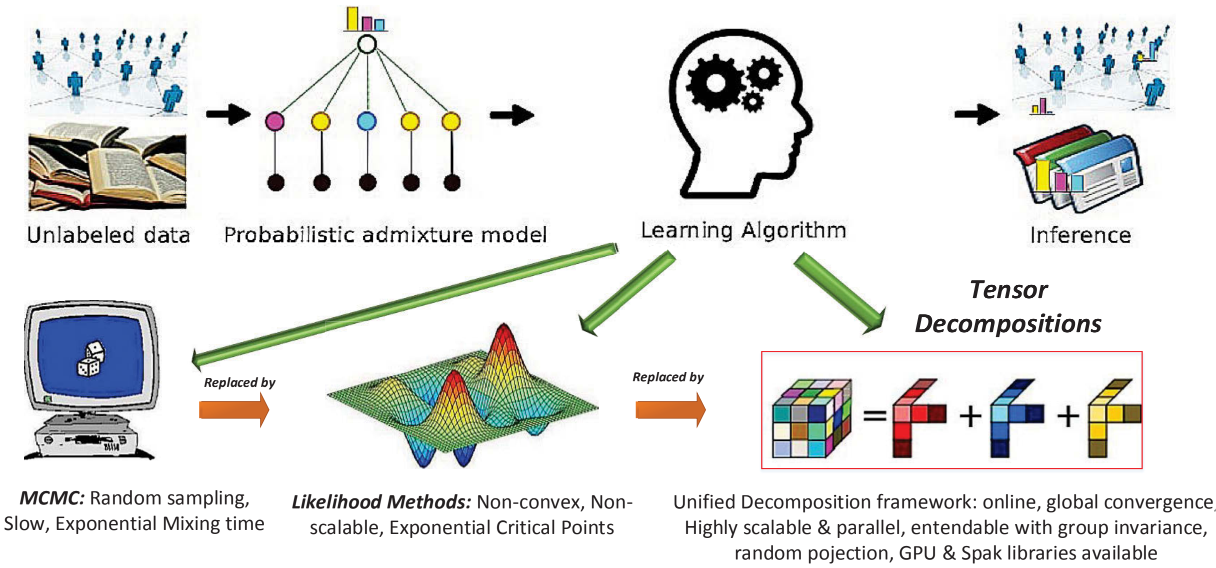


Figure 5.10 Role of Tensor Factorizations and Decompositions as Learning Algorithms

5.3 LITERATURE SURVEY

Popular Tensor processing models in fields such as social network analysis [121] have proven tensor methods to be faster and more accurate than other methods. Being able to capture higher-order relationships proves to be quite useful, especially in the Big Data Scenario of the Internet of Things [122]. Researchers from various research fields are interested in different, usually very diverse aspects of Tensors and their factorization methods.

The most sought after applications come from Multi variate multichannel signal processing, web mining and Social Networking, denoising of multimodal data streams, multimedia compression, feature extraction and clustering of hyperspectral images, EEG data analysis, brain computer interface and video tracking, etc. [123-129]. The standard domains in the context of ICT, smart sensor networks, dynamic RT systems and the IoT environment holds great potential for Tensor algebra [126-130]. Not to mention all the Big-Data relevant scenarios [105, 131, 132].

A recent networking domain application is Tensor computations applied to SDN systems in terms of QoS improvement [133], and tensor element modelling of Big Data Processing in Software Defined IoT, with various case studies [122]. The authors of [2, 134] have conducted many Case study analyses, explained through examples to reveal performance, convergence properties and efficiency of Multiway decompositions for synthetic and real world data. The domains chosen were Data Clustering, Text mining, Email surveillance, Musical instrument classification, Face recognition, Hand written digit classification, texture classification, Raman and Fluorescent Spectroscopy, Hyperspectral Imaging. Standard applications include neuroscience, bioinformatics and Chemometrics.

Three popular operation for which TD – TF are primary solutions are described in the upcoming sections, as part of a series of experiments conducted by the authors in [2].

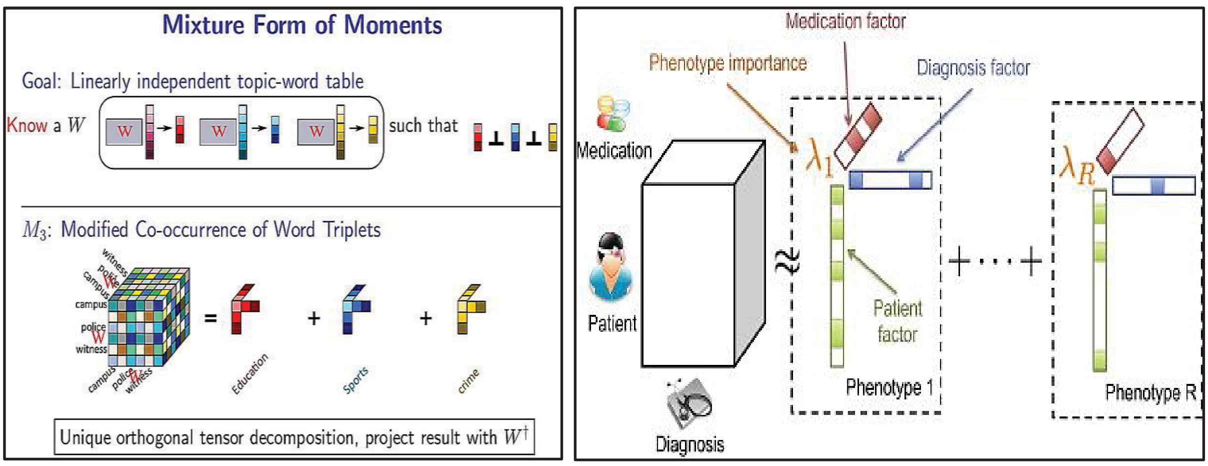


Figure 5.11 Discovery of Latent Factors in High-dimensional Data Using Tensor Methods [135] and High-throughput candidate phenotype generation via tensor factorization [136]

5.3.1 Data Compression

TD dramatically reduces the number of data samples required for real world data description (especially in more than 2 order). Instead of keeping the complete data tensor, the approximate representation is saved (in terms of factors/component matrices and core tensors). The resulting compact approximation is efficient. Factors generated are mostly sparse, since the actual number of samples is much lesser than total number of entries in all factors and core tensor [137].

5.3.2 Data Clustering

Unsupervised classification of patterns into groups or clusters with similar features is termed Hierarchical Clustering or Partitioning. Tensor factor based clustering processes a whole set of data is simultaneously processed in a single update cycle. Hierarchical partitioning, in which a nested series of partitioning with varying degrees and levels of dissimilarities, is a more robust and effective techni-

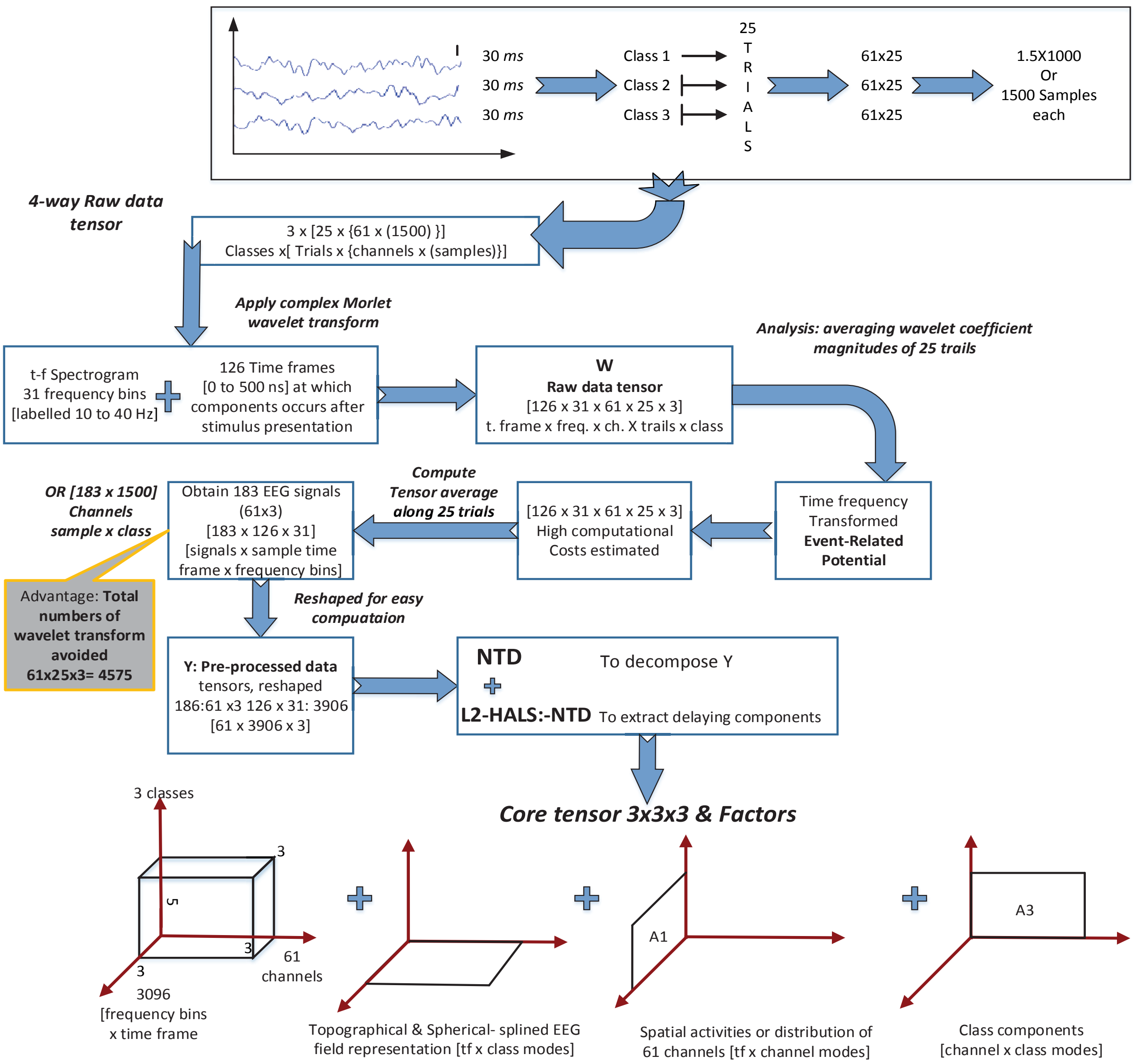


Figure 5.12 Application of Tensor Factorization for EEG Signal Spectral Decomposition for Intermodal Dependencies

-que but offers high computational complexity (gets better for large DS). Tensor models hold interpretations similar to most popular clustering schemes such as K-means and its variants, Mahalanobis distance based clustering, Graph Theoretic clustering, Neural Network Clustering, Spectral Clustering, Fuzzy clustering, etc. Ding et.al observes these characteristics.

5.3.3 Signal Classification

Supervised or unsupervised techniques are applicable in classifying time-series signals. Non-Negative Matrix Factorizations are applied for unsupervised classification of EEG signals, in which the model spontaneously allocates the objects from a DS given into the classes, with no prior knowledge of the classes or features of the object. Other similar applications include music or sound instruments notes class determination, classifying image, textures, genes or neural signals, emails or text docs, etc. Other than standard applications many prominent TD models have been applied into different applications for Knowledge discovery. Likewise, TD based group recommendation model to suggest groups to users in Social Network Data Mining [102], help identify the classes of patterns corresponding to time, etc. was deployed for Data mining, Information retrieval, filtering and storage. A different approach of application was in Fluid Mechanics to model fluid flow. In the majority of these applications, the typical tensor is a three-order tensor of $I\ (batch) \times J\ (measurement) \times K\ (time)$ which usually is unfolded in batch or time mode (batch-wise or time-wise unfolded matrix). Tensor-based batch-processing identifies the uncharacteristic batches or time instants, as is in most of the above explained applications.

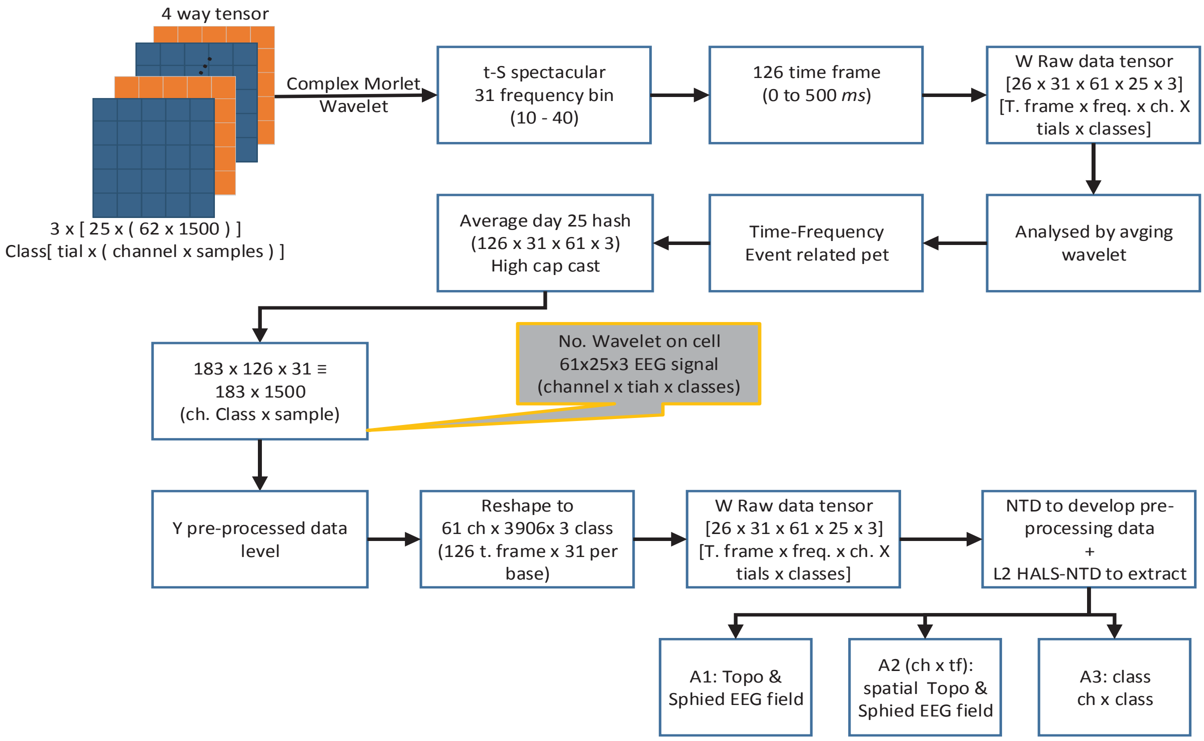


Figure 5.13 Application of Non-Negative Tucker Decomposition for EEG Signal Classification

5.3.4 Conceptual Parallels

Parallels can be drawn between extracting the anomalous patterns from immense data modules and Blind Source Separation. As is widely known, the term Blind refers to the unknown projection axes that have been fixed. Therefore, the sources which are metrics quantifying data behavior are determined without any prior knowledge of the source modules or the system behind it. Properly synchronizing and reconstructing events across remote sources is followed by correlating their semantics. That leads to inferring the system state / situation meaningful for the specific application. That is how detection of Anomalies or Data Pattern Changes in time-series signaling and data flows become significant in CPS scenario. Detecting anomalies helps record and utilize local dynamics of a heterogeneous system.

One of the earliest demonstrated extensive use of dynamic patterns of User and Load stats, and resulting integration attempts was in [138]. The proposition was validated with dynamic response to deviations in energy consumption, demand and grid condition. The design was based on review of two types of pattern or rule learning methods for contexts in single domain and whole domains, with slow learning curves and poor performance, and put forward a domain-based grouping organization to effectively produce rules for dynamic pattern generation through reasoning process. There are 3 main types of anomalies observable in time series data streams, as below: Point, Contextual, and Behavioral [139].

5.4 EXPERIMENTION FRAMEWORK

Anomaly pattern analysis is done by extracting information flow and flow path patterns, in an algorithm. The captured and processed patterns will make up information variables, representing functions and relationships between from the attributes of individual entities. The scope of this data towards proactive management of the parent system is discussed.

5.4.1 Definition of Behavior and Anomaly patterns

The features derived from the knowledge discovery capture dependencies equivalent to metadata visualization of the various information modes as variation maps across one or multiple variables. They signify Contextual information local to each feature's span of variables, in essence their correlation. Extracted features are independent of the dimensionality of data. Likewise, changing the order number of Data streaming, only results in mixing the pattern elements, holding the properties along any one axis. The interpretations for a TD based method will be:

CORE: Metadata is contextual data or information that describes data or its characteristics. The Meta data extracted from a Data Tensor block helps model and track anomaly pattern behavior.

FACTORS: Feature vector. A feature is a characteristic or attribute of an example, which might be useful for learning the desired concept.

- **Active associations:** first degree relationships showing input and output (intermodal) variation dependencies.
- **Active dissociations:** first degree relationships showing input and output (intermodal) variation independencies.

Active interrelations are refined out of system response and operational characteristics. For e.g. in the Smart Grid CPS context, power generated by distributed generators qualify as input, while customer and storage units are the output subsystems.

- **Latent associations:** subscale dependency patterns across different data streams, from single or multiple components, irrespective of contributions to input or output variations.
- **Latent dissociations:** subscale independency patterns across dimensions of multiple modes, irrespective of contributions to input or output variations.

Latent interrelations occur in subscale, otherwise undetectable by temporal or spatial data modelling. The mappings generate correlation matrices of the various sources (streams) of component elements along all involved modes.

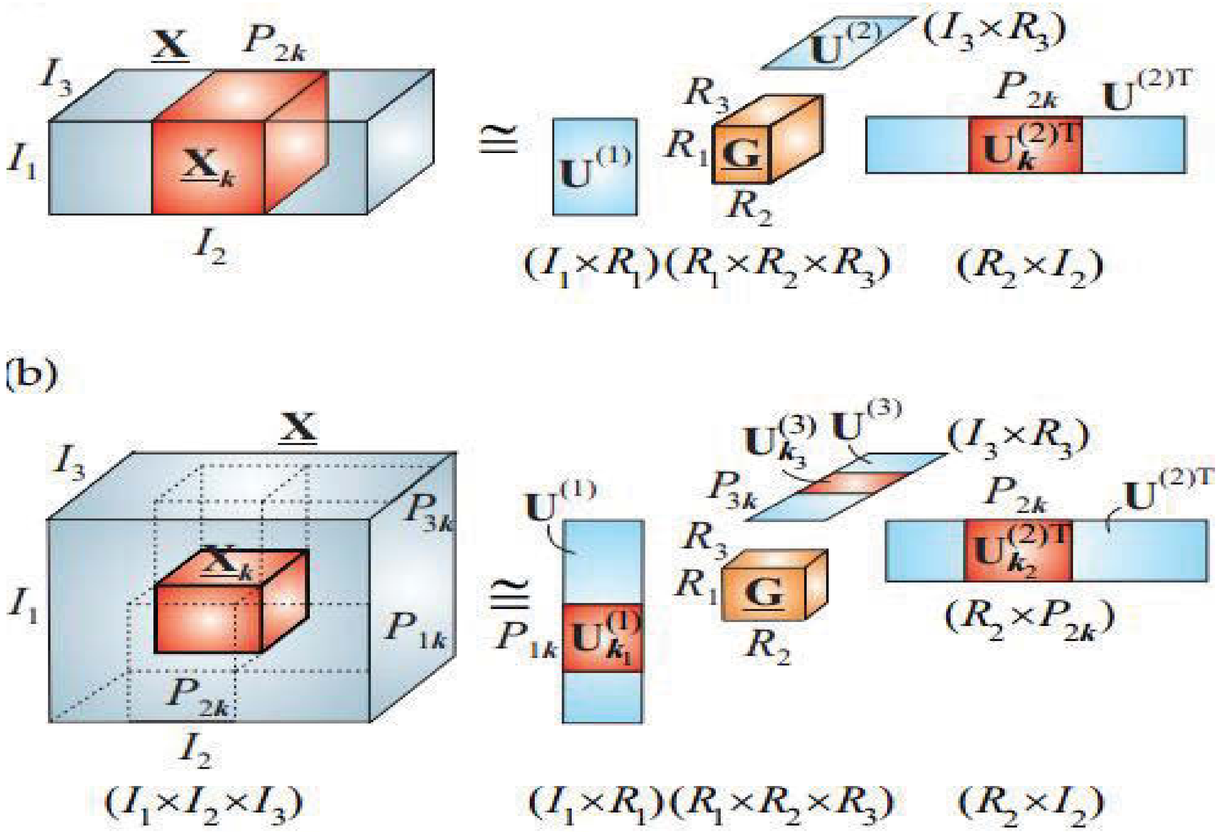


Figure 5.14 Anomaly detection using Tucker model: 3rd-order data tensors divided into blocks (a) along the largest dimension mode (all clusters), b) along all modes (varying time windows, 1 month or more)

Proposed model measures the subscale or latent associations between patterns and pattern classes by considering any 2 attributes of tensor decomposition as Investigative Parameters of Extracted Patterns. Feature Tensors or Feature Tensor Pages show a pattern variation stream being tracked. Features spreading across more than one data stream, computed in terms of cross-correlation between each pair of streams, followed by the best match.

Typical examples in the Smart Grid context would be signify the effect of weather, building occupancy, load variation patterns and other second degree relationships among input and output attributes. They are quantities under study in this thesis.

- **Active associations** – the power generated affected by the input subsystem, power consumption affected by the output subsystem (EIF input and output modules) and similar first degree relationships on input n output variations.
- **Active dissociations** – first degree variables not affecting input and output variations.

- **Latent associations** - signify the effect of weather, building occupancy, load variation patterns and other second degree relationships on input n output variations. Latent associations signify the underlying patterns of a particular data-stream, from single or multiple sources that will be mapped based on a correlation table, and hence be used to estimate information flow patterns.
- **Latent dissociations** – second degree and later variables not affecting input and output variations.

Due to complexity of coding issues, this experiment is limited to investigating associations (active and latent), in terms of modelling and monitoring.

5.4.2 Phase I: Data Acquisition

This stage refers to acquiring the multivariate higher order data flows. The pre-requisites for the processing decide the choice. Investigative parameters include Origin, Authenticity, Periodicity, Noise levels and above all, Information content, measurement time window, long time period of study / observation, sample range, etc. Also, the measurement cataloging standard, like file and header structure, file size and contents, etc. These need to be associated with the devised data tensor formats to comply as pattern variables, to help steps like reshaping, data read and write or comparisons.

Detailed study of the data generator CPS setup is necessary to select the right data streams. The correlation between data streams is dependent on the time of occurrence of events. Hence synchronization between data collection cycles allows fusion of intermodal associations (timestamps, markers + universal clock). Other methods to define time-stamps, although the measurement instants for different sources are different, is to interpolate them over time intervals.

5.4.3 Phase II: Pre-Processing

This stage is about Cleansing information streams, Data optimization or normalization. The Pre - processing stage is necessary to organize and compose data sets in a disciplined structure to match the physical significance of the factorized tensor. Data Preprocessing is compulsory as real world data are generally incomplete or missing attribute values, certain attributes of importance, or even having only aggregate data. It can also be noisy, contain errors or outliers, inconsistent and containing discrepancies in codes or names. This is because not time-variant systems data acquisition methods are often loosely moderated. Data Parsing and Reduction are important operations, performed by projecting normalized data onto the tensor frame (mode n, dimension m). Noise or artifact removal often encompasses data reduction stages (filtering) followed by a data reconstruction stage (such as interpolation, optional though). Data pre-processing thus leads to final training and testing data set.

5.4.4 Phase III: Tensorization or Data representation

Data Representation is about packing the information streams along the various modes, and selecting required dimensions to incorporate the anticipated patterns. A three-mode tensor can be constructed from the usage data, $\mathrm{Z} \in R^{(\mathrm{I} \times \mathrm{J} \times \mathrm{N})}$ Where $\mathrm{I}, \mathrm{J}, \text{and } \mathrm{N}$, are variables A, B and C respectively. Each element z^{ijn} in the tensor denotes component i's interaction with component j and component k. Tensorization refers to procuring the multi-dimensional dataset and transfer it into a TD compatible signal hierarchy to reflect the active and latent multiscale structures.

5.4.5 Phase IV: Tensor Processing

The TD operation required number of components and core tensor and factor dimensions are set. Data Matrix is allowed to decompose based on chosen decomposition rule, such as NTD, α-HALS NTD, etc. When tensor is factorized into components (matrices or vectors), their innate properties like linear combination, superposition, spectral composition etc., will represent the latencies or semantic relations or interdependencies

5.4.6 Phase V: Post-processing: Anomaly Detection in Data Tensors

The factors and core component are inspected, unravelling Latent behaviors of the corresponding element. Then, these specific features or observations are associated with control/communication signals to be sent to an external device. Result visualization identify and distil physically significant and meaningful hidden components, as well visualize their consequence or impact, though various convergence or divergence based interpretative metrics and cost functions.

5.4.7 Phase VI: Performance Examination and Learning Curve

Optimization criteria are set to ensure robustness the results. Different performance measurement parameters include number of iterations, Residual analysis, visual appearance of loadings, smoothness index, etc. They may typically be categorized as Convergence measure or Divergence measures. Optimization techniques (optimized loss / cost terms) and statistical measures used for learning rule update and learning curve moderation for the pattern extractor.

5.4.8 Implementation Environments

Basic setup and experimentation for Tensor based Data Processing is implementable in any multi-paradigm numerical computing environment and programming language. MATLAB is an apt tool for initial testing and trials on data. A whole variety of tool boxes for Parallel and Distributed Computing, Multiway Data manipulation, etc., enable building Tensorization and Tensor computation algorithms. The best part is the number of free range tool boxes made publicly made available for Tensor Factorizations. Popular open-source software resources for tensor decompositions in other low-rank tensor formats include:

- Tensor Toolbox: a versatile framework for basic operations on sparse and dense tensors, including CPD and Tucker formats, by Bader and Kolda.
- Ktensor, Ttensor and Htensor toolboxes: for tensors in CP decomposition, Tucker decomposition, and HTD.
- TDALAB and TENSORBOX: They provide a user-friendly interface and advanced algorithms for CPD, nonnegative Tucker decomposition and MWCA.
- Tensorlab toolbox: It builds upon the complex optimization framework and offers numerical algorithms for computing the CPD, BTD and Tucker decompositions. The toolbox includes a library of constraints (e.g. Non-negativity and orthogonality) and the possibility to combine and jointly factorize dense, sparse and incomplete tensors.
- N-Way Toolbox: it which includes (constrained) CPD, Tucker decomposition and PLS in the context of Chemometrics applications. Many of these methods can handle constraints (e.g. Non-negativity and Orthogonality) and missing elements, by Andersson and Bro.
- TT Toolbox, the Hierarchical Tucker Toolbox and the Tensor Calculus library: They provide tensor tools for scientific computing, by Oseledets.
- Matlab HTucker toolbox: for conveniently storing and manipulating higher-order tensors.
- Tensorflow library in Python: It allows flow based tensor processing without further reductions and pre-processing stages.

The control optimization modules based on Pattern processing maybe prototyped in domain specific environments as Matlab, LabView, etc. Demonstrating the integrated operation heterogeneous systems based on such controls require platform-aware integrated simulation setups. A detailed investigation of the various apparatuses and tools that are being applied in the Smart Energy Grid scenario was performed [140, 141], along with the primary approach types as Co-Simulation, Integrated Simulation, System-of-Systems and Multi-agent systems. It was found that OpalRT and GridSim are ideal packages with MatLab/Simulink interfacing possible, to enable RT-HIL emulation. Matlab will also aid in simulation and preparation of the Data files. GPUmat, the toolbox for utilizing GPU processing with Matlab, allows Data to be fed into the Matlab system as XML docs. The purpose of using GPU is further justified, other than the HPC clustering and parallel processing is the great accuracy levels than by CPU, as well as the steep learning curve.

5.5 CHAPTER SUMMARY

Tensor computation-based pattern extraction model is put forward in this research paper to utilize various latent attributes of the multi-faceted data and represent multi-dimensional relations. Suggested methodology *models Data and Signal flow in a regressive manner (energy-analytic). This will logarithmically scale down various efficiency limiting factors such as processing time, energy consumption, latency and response time of the origin system*. Methodology derived towards realizing these goals is tracking patterns in the data cube based on Tensor properties, like core tensor, or factor matrices, etc. It will allow capture underlying dependencies to build a Multimodal Anomaly Detector.

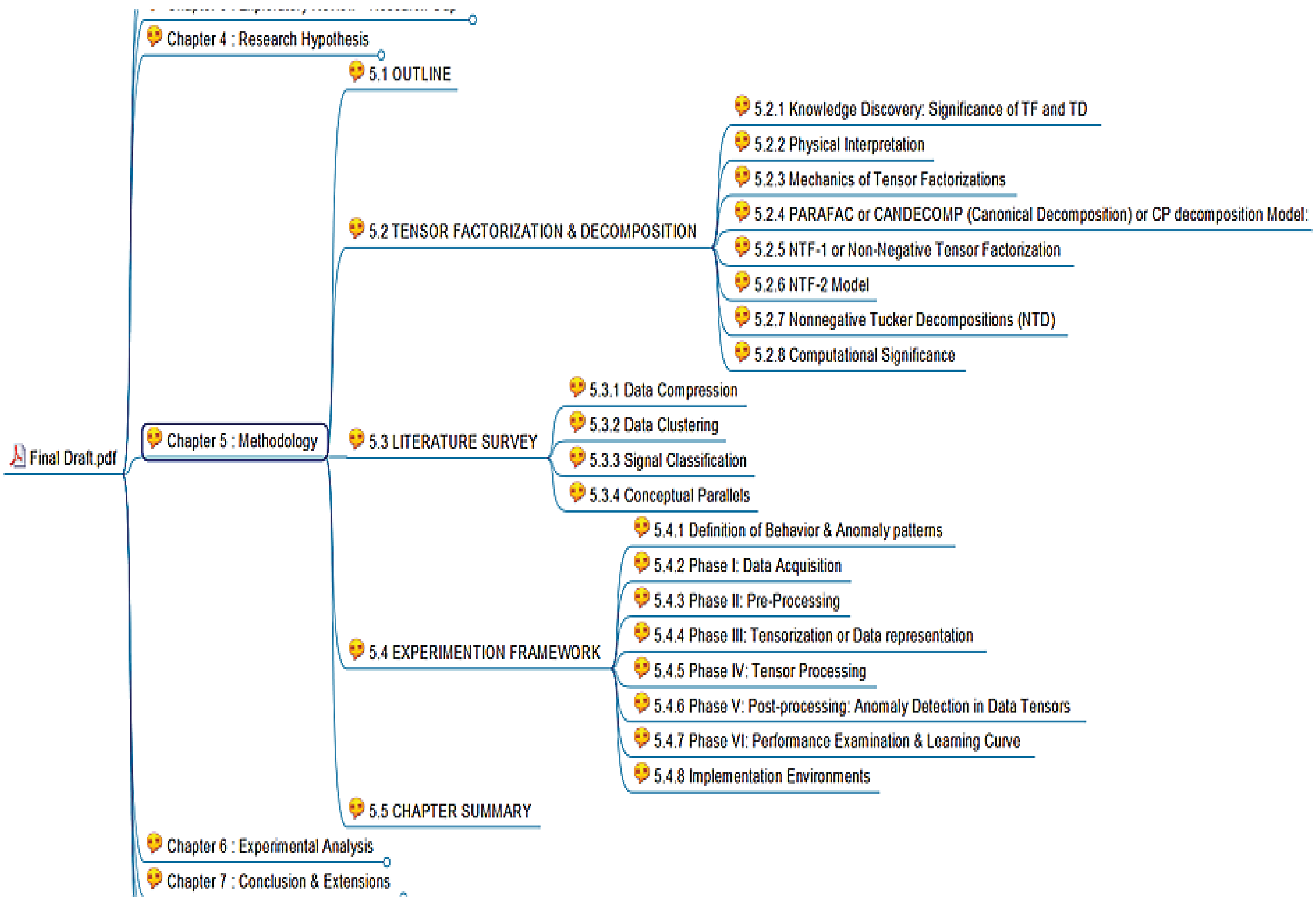


Figure 5.15 Mindmap outline of Chapter 5: Methodology

Chapter 6 : Experimental Analysis

6.1 OUTLINE

This chapter deals with the experimental trials performed to validate decomposition of time-variant multi-channel data stream. It confirms the claim of the hypothesis that multimodal patterns observed in higher order data capture subscale behavioral and anomaly patterns. An outline of the entire procedure is portrayed in Figure 6.1.

6.2 EXPLORATORY STUDY OF DATASET

In order to fully tap into the information content of raw data sets in a database, a good appreciation of the features and contents is essential. Any capability of analysis and interpretation require careful identification and pre-processing. The database in use for this experimentation was derived from the Australian Government's $100 million Smart Grid, Smart City 2014 project which includes data on trial results and feedback on different elements of the cost benefit assessment and final report. This major government initiative was delivered by Ausgrid and its consortia partners for

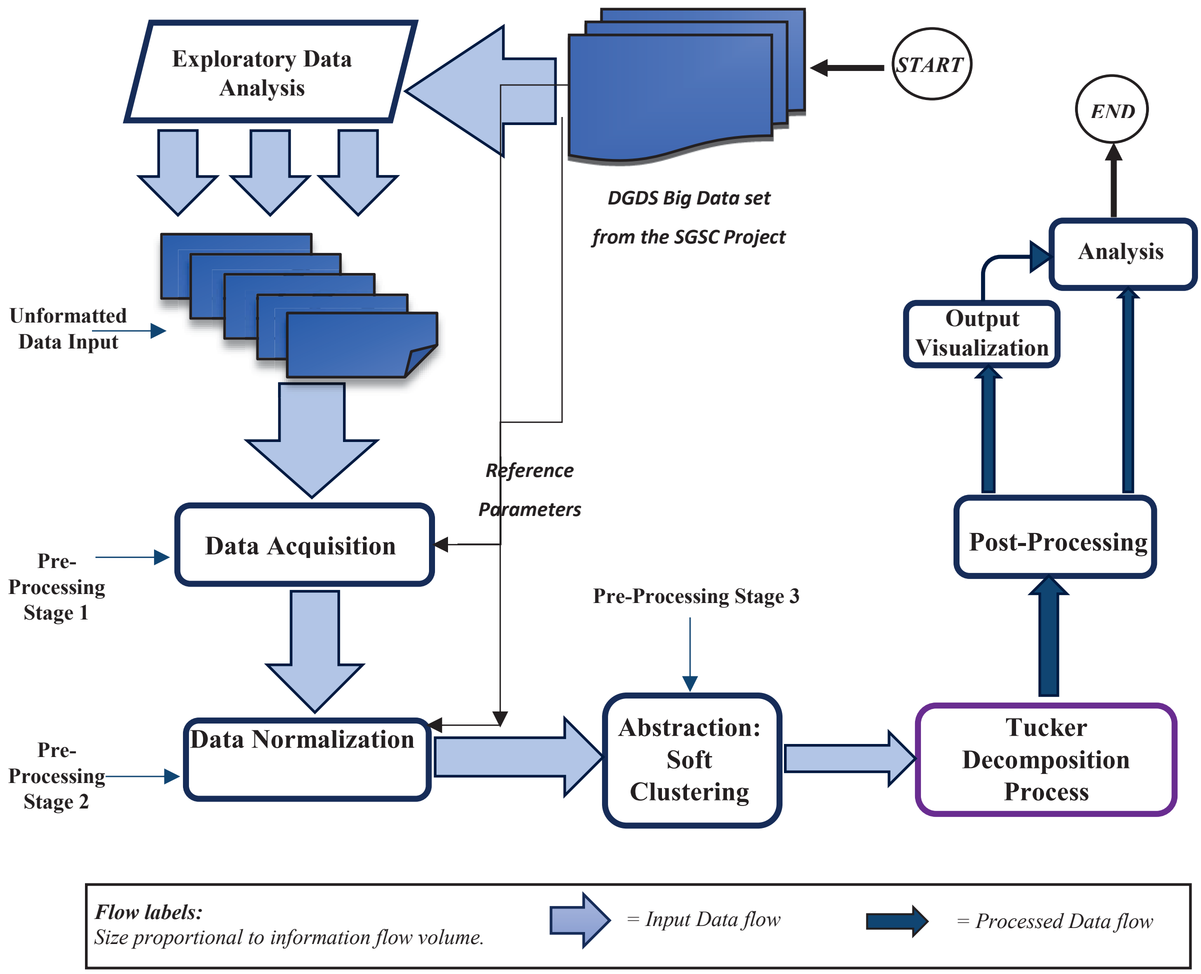


Figure 6.1 Flowchart demonstrating PoC, hence the experimental analysis phase and corresponding functions implemented.

Name of work stream	Description of work stream
Supporting Information and Communication Technology Platforms (Common Platforms)	Investigated the feasibility of various high-speed, reliable and secure data communications network and associated IT systems which integrate with the electrical distribution network. It also examined interoperability with the National Broadband Network.
Grid Applications	Investigated the ability of grid-side monitoring and control technologies to reduce network operating costs and support the future planning and implementation of lower cost networks.
Customer Applications	Focused on residential electricity consumption, reliability, customer behaviour and responses to feedback technologies and pricing models. It also included an electric vehicle trial and investigations into the interoperability of electricity metering with gas and water metering.
Distributed Generation and Distributed Storage	Investigated the feasibility and potential benefits of distributed generation and distributed storage within electricity grids.
Electric Vehicles	Investigated the potential impact of wide-scale uptake of electric vehicles in Australia on the electricity distribution network.

Figure 6.2 Listing of the major subsystems of the SGSC Trial setup

supporting and informing the industry-led adoption of smart grid technologies in Australia. The project trialed Australia's first commercial-scale smart grid [142]. The Ausgrid consortium funded by the Australian Government consists of the government as well as private stakeholders.

6.2.1 Overview

The Program, sought to determine whether there were benefits from the deployment of these technologies for Australia. The trials were based in Newcastle, New South Wales, but also covered areas in Sydney CBD, Newington, Ku-Ring-Gai, and the rural township of Scone. Reports from the project are available under Related Reports below and data from the project is gradually being made available on data.gov.au. A more detailed account of the timelines and deliverables, trial findings, conclusions, lessons learned and recommendations can be found in individual Technical Compendia, Executive summary and Main report documents, which were studied in detail and organized as précis in the following sections.

Smart Grid, Smart City Program has shown that it is possible to more effectively manage peak demand through consumer behavioral change and with the assistance of feedback technologies and dynamic tariffs. The major sections of the project implementation is as listed in Figure 6.3.

The DGDS Project commenced in October 2010 running through a number of phases as shown in Figure 2. The high penetration of solar power was studied. Solar generation profiles and solar connection data by network assets were compared with load profiles of individual feeder loads at various levels of Ausgrid network. They were found to be highly dependent on weather conditions and loading of the network asset. Another possible strategy followed by the DNSPs to obtain value from

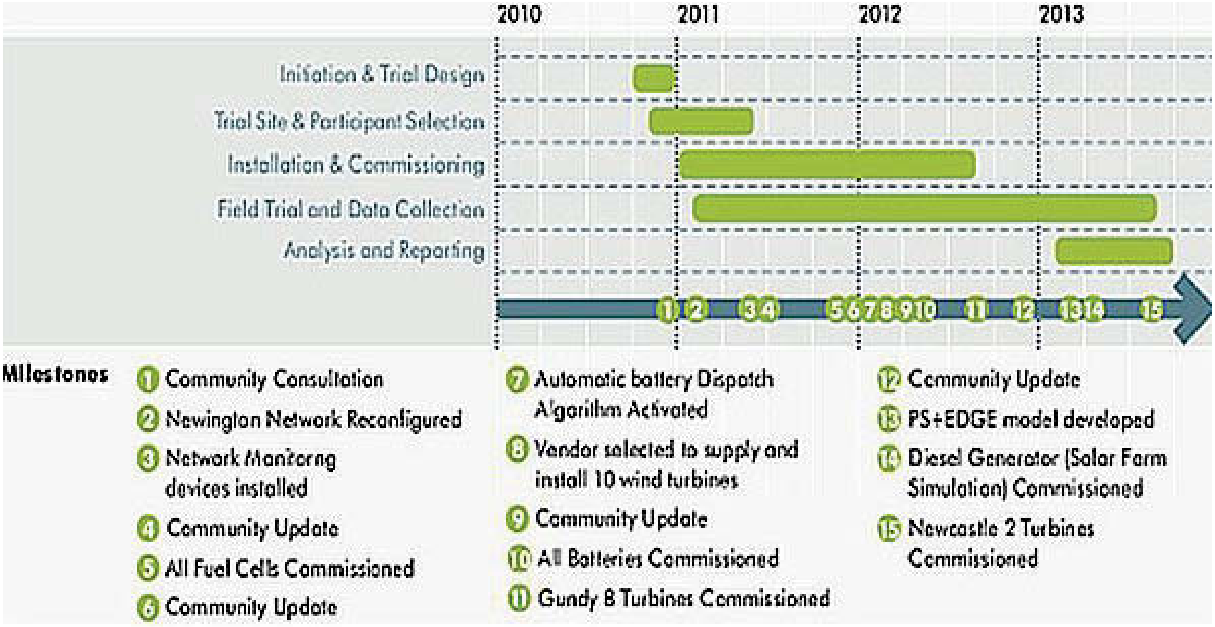


Figure 6.3 DGDS project timeline diagram

distributed storage technologies installed so far was also an important motivation. They were connected to the grid-side of the network where the customer does not receive any significant direct benefits through reduced electricity bills. The possibility of novel control and protection strategies, including a new approach to detecting broken conductors, were also included. The operator controlled trials conducted over Summer (November - January 2012/13) and the automated battery control trials that occurred over Winter (June - August 2013), demonstrate the possible implementation and usage of a Distributed Energy Resource Management System (DERMS).

A list of the major subsystems of the SGSC Trial setup is given in figure 6.2 [143]. The timeline diagram of the various phases of the project are given in Figure 6.3 [144].

6.2.2 Organization

The program operation was designed as field trials, advanced modelling and PS+EDGE simulation trial elements.

- The *field trial results* were used to validate the models and gain insights into the commercial-scale deployment of these technologies.
- The *advanced modelling and field PS+EDGE simulations* gave added analytical tools for different penetrations of DSDG devices which are not commercially and physically possible in a field trial.
- The *trials and modeling using PS+EDGE* investigated the impact of DGDS on power quality parameters including: Voltage levels, Power factor (PF), Voltage unbalance, Total harmonic distortion (THD), and Flicker (PST and PLT).

Three trial setups were established to study the impact of Distributed Generation and Distributed Storage on power quality parameters. Locations where various types and combinations of DGDS were installed:

- Newington Area (suburban area with solar PV; 31 March 2013 to 7 April 2013) [145].
- Newcastle Area (urban area with battery storage, ceramic fuel cells and solar PV; 8 December 2012 to 17 December 2012) [146].
- Upper Gundy Area (rural area with battery storage, solar PV systems and wind turbines; 8 December 2012 to 17 December 2012) [147].

6.2.3 Data Collection and Management

The following data sources are pertinent to understanding the DGDS trials [144]:

- Network models for the Newcastle, Upper Gundy and Newington areas – to understand the interconnectivity of feeders and measurement devices. This allows the connectivity between all the feeders to be understood and all the measurement signals that are available at each point in the network to be discovered.
- SCADA data from all measurement devices in the Newcastle, Upper Gundy and Newington areas, including the zone substation points and the Distribution Monitoring and Control devices,
- Measurements data from RedFlow and BlueGen devices, wind turbines and photovoltaic systems,
- Metering data from the meters through which the DGDS devices were connected to the network.

Voltage and other associated information such as load and power factor can be obtained from the following types of device in the each area:

- Zone Substation SCADA signals, including voltage, current, power factor and load.
- Distribution Monitoring and Control devices in distribution substations.
- RedFlow and BlueGen devices.
- Meters.

All data collected from the DGDS devices is stored in the Ausgrid Historian system. The detailed description was catalogued in the *Common Platform Study* documentation. The datasets of interest is regarding the different modelling inputs formed from SGSC customer trials.

6.2.4 Raw Data Imported for Experiments

SGSC trial stakeholders organized to enroll some customers into Demand-Response program. Datasets were composed to reflect the objectives of the trial and the sequences of demand response cycle. The *'Half-Hour Consumption and Generation'* data provide kWh readings of the actual flow of

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Octo	Nov	Dec
2008												
2009												
2010												
2011												
2012												
2013												
2014				Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil

	1	2	3	4	
2008 2009 2010	JAN2008_TO_ OCT2010	NOV2010_TO_ OCT2011	NOV2011_TO_ APR2012	MAY2012_TO_ AUG2012	
	5	**6**	**7**	**8**	**9**
2011 2012 2013 2014	SEP2012_TO_ NOV2012	DEC2012_TO_ FEB2013	MAR2013_TO_ MAY2013	JUN2013_TO_ AUG2013	SEP2013_TO_ MAR2014

Figure 6.4 Composition and structure of DGDS dataset; dataset timelines for each file and Color coding for timelines

energy in the SGSC dataset. (The specific names of various data objects and fields are made clear using '((item))' font). The main items are:

- Electricity imported from the grid ((General Supply)) - main power source (all applied tariffs),
- Import during off-peak regular or control load,
- Gross or net values of prosumer generation exported back to the grid ((Generation))
- ((Peak Events)) and their details like ((Event Key)), ((Type)), ((Date)), ((Start Time)), and ((End Time)).

The devices data are recorded frequently ((Readings Date-Time)) and ((Time of Day)), while the peak events are also recorded ((Readings)). The type of measurements, e.g., analogue versus digital, and units, e.g., kWh, are indicated as well ((Measurement Types)). This the format of the DGDS subset of the SGSC. The directories from the DGDS dataset were periodic data files over 40 GB each.

6.3 IMPLEMENTATION FRAMEWORK

RHEL (Linux) environment was chosen for the experimentation The RHEL, based on the analyzed literature, offers the following benefits to Big Data Processing:

> Free availability; Easy installation and use; Wide application domains; the power of Bash scripting in coding; Compliance to a large range of software and application packages.

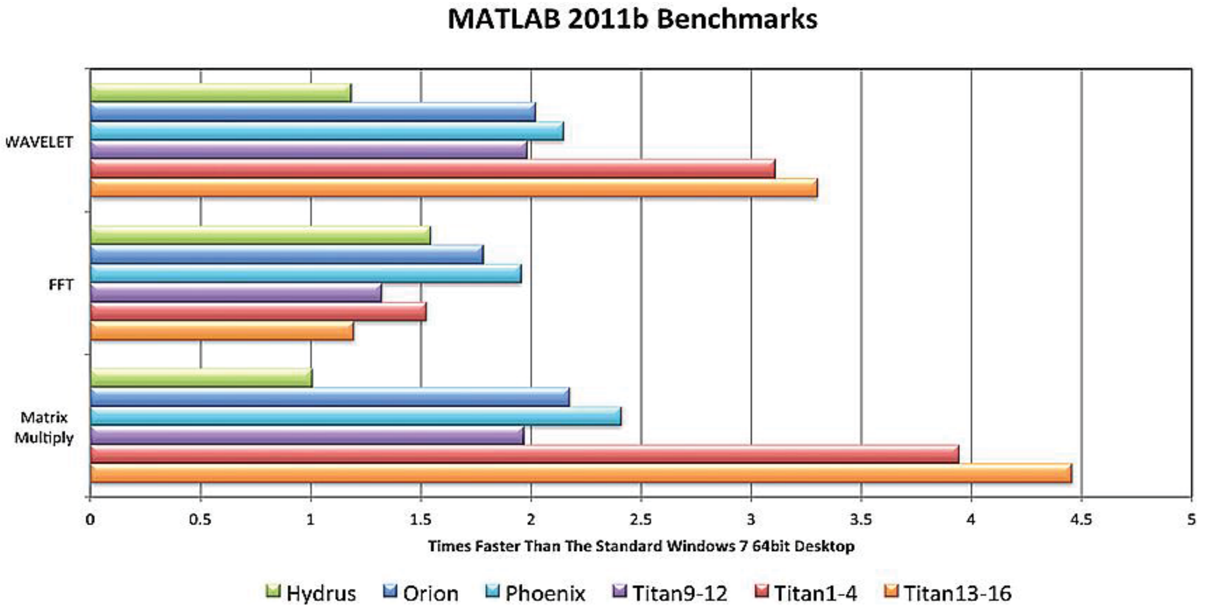


Figure 6.5 Cluster nodes performance analysis based n Matlab 2011 benchmark

6.3.1 Computational Environment

The entire coding, processing and analysis was conducted at the high performance computing facility (HPC) of our university - Advanced Research Computing Laboratory (ARCLab f.k.a FEIT Cluster). ARCLab provides an interactive high performance computing resource for all researchers within UTS.

Table 6.1 Hardware specifications of the 27 HPC Cluster nodes which provided access for computing and analysis

No.	Feature	TITAN 1 to 4	TITAN 5 to 8	TITAN 13 to 15	ATLAS (8)	VULCAN (4)
1	Description	2x 3.1GHz Intel Xeon E5-2687W	2x 3.33GHz Intel Xeon X5680	2x 2.8GHz Intel Xeon E5-2680 v2	2x 3.1GHz Intel Xeon E5-2687w v3	2x 2.5GHz Intel Xeon E5-2670 v2
2	# of Cores	8	6	10	10	10
3	# of QPI Links	2	2	2	2	2
4	Processor Base Frequency	3.1GHz	3.33GHz	2.8GHz	3.1GHz	2.5GHz
5	Cache	20MB L3 Cache	12MB Level 3 Cache	25MB L3 Cache	25MB L3 Cache	25MB L3 Cache
6	Bus Speed	8 GT/s QPI	6.4GT/s QPI	8GT/s QPI	9.6GT/s QPI	8GT/s QPI
7	Max Turbo Frequency	3.8GHz	3.6GHz	3.6GHz	3.5GHz	3.3GHz
8	Memory Types	1600MHz ECC DDR3-RAM, 128 Gb	1333MHz ECC DDR3-RAM, 48GB	1866MHz ECC DDR3-RAM, 256 Gb	2133MHz ECC DDR4-RAM, 64GB	1866MHz ECC DDR3-RAM, 128GB
9	Channel	Quad	Triple	Quad	Quad	Quad
10	GPU	NIVDIA Quadro 2000 1GB Graphics Card	NIVDIA Quadro K600 1GB Graphics Card	NIVDIA Quadro K2000 2GB Graphics Card	NIVDIA Quadro K2200 4GB Graphics Card (GPU) NIVDIA Quadro M4000 8B Graphics Card (GPU)	2x NVIDIA Tesla K20M 5GB Graphics Card
11	RHEL version	6.7	6.7	6.7	7.3	6.7

The facility has been distinctively designed to for users at all levels, starting with beginners at little or no UNIX experience. The hardware specifications of the cluster are as follows, with the resources that were utilized for the research are summarized in the Table that follows.

- 6 Clusters {ORION, HYDRUS, PHOENIX, TITAN, VULCAN and ATLAS} and a couple of private nodes {SATURN GPU Cluster} running Red Hat Enterprise Linux (RHEL). All connected to server with a total of ~350TB of disk storage, designed as a distributed file system.
- Total of 1150 cores and 5.6TB of memory housed in the facility.

A custom developed web portal regulates the access. Users can verify information about the available cluster nodes via the portal and activate a graphical or terminal, local to UTS or offsite, access stream. The standard interactive graphical User Session via NoMachine Web Companion was used mostly during the course of research. Load Balancing by Login allows Queue-less access in this case.

The other options used include Terminal access using SSH, which came of use due to code interpretation complication arising from the massive data size. Also, remote access dictates the use of UTS VPN. Finally, transferring large information quantity uses SFTP client such as Filezilla.

6.3.2 Bash Scripting and Python Virtual Environment

Bash Shell scripts functions can invoke Unix programs, pipe output files, redirect I/O from/to files, control flow, file status check, etc. and hence are widely adopted. Programming languages operating over interpreters Python and Ruby, have added readability and maintainability apart from these listed functions.

Bash, the standard shell in Linux terminal for common users is faster in terms of performance. With credits to SED and AWK scripting data crunching is even achieved in a few lines of instructions. However, dealing with large programs makes Bash increasingly complicated whereas Python does not.

With the above argument in limelight, for Pre-processing stage 1, Shell scripting was chosen whereas the rest of the procedure was done on Python Virtual environment. The fundamental packages and Libraries relevant for scientific computing utilized during the entire process include: Numpy, Pandas, TQDM, Scipy, Cython, Matplotlib, Hdbscan, and Mpltools.

In order to implement the entire processing in Python the Python Virtual environment was selected. According to the documentation, Virtual environment “virtualenv” allows users to install Python modules in a pseudo root space within the base Python installation.

Implementation was done in Python after experiment with different libraries and algorithms. The programs were also optimized, based on the experimenting system requirements. Parallel computing and distributed implementation were also utilized. Matlab was tested initially but encountered the following hardships:

1. Matlab memory requirements, especially for the interpreter were severe. The result was several trial runs either just getting hung up or crashing.
2. Matlab tended to run out of memory, much earlier than Python for dynamic processing more than Write-Save executions. Python did as well, but took much longer to time out.

6.3.3 Python Toolboxes for Tensor Decomposition Scripting

The greatest challenge in using Python environment for TD is that the existing Python libraries for tensor factorization don't have all the required options implemented. Hence parallelized implementation wasn’t as efficient as anticipated, as since the complexity of the TD mathematical operation is just too big. Even parallelizing it wouldn't allow setting a much bigger input data range. Also in order to do it, the decomposition pseudocode and base algorithmic code will have write myself which would be unfortunately much bigger a scope than that of this project. Another hindrance was that the documentations for the libraries were quite limited in detail.

1. Tensorly: Tensorly was stated and tested to be the faster tool box out of the lot. Core size was the primary factor affecting the computational complexity [148]. Hence just lowering the core size but applying on bigger data was recorded to be the most efficient computation. Tucker decomposition via Higher Order Orthogonal Iteration (HOOI) was the main component of this toolbox, along with PARAFAC, Non-Negative PARAFAC and NTD and Robust_PCA.

2. PyTensor: Pytensor also has more than 1 decomposition rule, which are basic Tucker I and Dynamic Tensor Analysis.

3. Scikit Tensor: Canonical / Parafac Decomposition, Tucker Decomposition, RESCAL, DEDICOM and INDSCAL are the components of this library. However, in Python environment, Scikit-tensor toolbox code has been notified to be failing.

4. RTensor: RTensor in Python allows interfacing with R language environments; documentation limits it to freemium open source distribution of Anaconda. R language tensor decomposition package, the RTensor library, is a quite powerful software worth learning and has all decompositions available in it.

5. Tensor Train tool box: Tensor train toolbox has very poor associated documentation, with basic operations required for Tensor processing, but devoid of any methods for decomposition rules [149].

6. Python hosted Tensor toolbox: The documentation was quite lacking but the support or last commit was quite new, but the one previous to it was done couple of years ago.

7. Tensorflow: TensorFlow is the open-source software library for Machine Learning across a range of tasks developed by Google. It is capable of processing multi-way data structures as single variables, and makes parallelizing calculations much easier. However because of lack of implementations of tensor factorization or decompositions made it the less preferable option.

8. Drumichiro NMF and NTF: The sample code to study Non-Negative Matrix and Tensor Factorization was written by Drumchiro, and made freely available in Github. For Python beginners the easiest environment to run these Python scripts was stated as Anaconda. The complete toolbox focusses on denoising.

9. Python Toolbox for Nonnegative Matrix Factorization: This package includes Python implementations (with Numpy and Scipy) of numerical algorithms for computing NMF.

It seems like all of them actually differ in implementation, and were developed based on different pseudo-code (as shown in the different articles they cite). So the choice was Tensorly.

6.4 EXPERIMENT RESULTS and DISCUSSION

6.4.1 Pre–Processing – 1: Acquisition

As customary in standard Big data processing, raw data has to be normalized and formatted, to match the standardized processed input structure. Stage 1 of Data preprocessing deals with Acquisition of the raw input data collection. The aim is Dimensionality Reduction of the higher order expansive documents which would otherwise prove to be critically computationally-intensive input signal.

Development and Testing: The shell script for crunching Big data files into smaller ones takes the read-write size as the number of lines that is required in each file. Total number of device elements or measurement timestamps for which valid entry logs were made is unknown. Also, the particulars of the individual field or simulation trials of the DGDS were not stated folder-wise. Hence testing phase consisted of trials with increasing read-write cycles till the critical level of CPU or memory usage was attained, which resulted in process being killed. So the Big Data Files were parsed into components file collections, after extraction of 300000 lines per iteration.

The required columns 1, 3, 4 corresponding to device ID, timestamp and measurement reading, were also extracted per read size of 300000. It was also imperative to scan the length or the number of lines in each data catalog prior to the chunking. Linux terminal was scripting eases the computational load and elevates the data crunching speed.

AWK programming in the shell helps identify the lines or strings where a particular change occurs, just before or after the change. Shell scripting slices the < 20GB .dat files into csv files of 300,000 lines each. Time and device ID to count the number of rows are compared, and are loaded into the corresponding selected measurements in the corresponding csv file generated and numbered in the increasing count. Parsing the raw Big Data Files into smaller file components require heavier-duty processing, a Bash script was developed using VIM editor, a powerful text editor tool. The testing was performed systematically starting with a small amount of data first and progressed iteratively.

Table 6.2 Timeline and constituent data modules of Pre-Processing Stages 1 & 2

Measurement logs on different Time periods	STAGE 0		STAGE 1			
	Log Entries count	Duration < 1 day	Test Period- 2 days	Duration per folder	No. of subfiles	Analysis Duration - ~ 1 day to 1.5 days per folder
JAN2008_TO_OCT2010	703648772			19 days	2348	
NOV2010_TO_OCT2011	714010236			19 days	2383	
NOV2011_TO_APR2012	700100041			18 days	2336	
MAY2012_TO_AUG2012	801951664			23 days	2676	
SEP2012_TO_NOV2012	699991546			19 days	2337	
DEC2012_TO_FEB2013	745003332			21 days	2486	
MAR2013_TO_MAY2013	828656211			24 days	2765	
JUN2013_TO_AUG2013	710310323			19 days	2370	
SEP2013_TO_MAR2014	391432572			10 days	1307	
	Scan the Length		Parse into components file collections, after extraction			

Measurement logs on different Time periods	STAGE 2				
	Testing - 4.5 days	Folder ID - No. of files generated - Duration of computation [rough count]			
JAN2008_TO_OCT2010		opJanNov08 - 172 - 5.8 hrs	opDec08Nov09 - 306 - 6 hrs	opDec09Oct10 - 14029 - 5.5 hrs	
NOV2010_TO_OCT2011		opNov10 - 13855 - 5 hrs	opDec10JanFeb11 - 15609 - 6.5 hrs	opMarAprMay11 - 14588 - 6.5 hrs	opJun11Oct11 - 18166 - 7.5 hrs
NOV2011_TO_APR2012		opNovDec11 - 18475 - 5 hrs	opJanFeb12 - 19735 - 5.5 hrs	opMarApr12 - 24913 - 6 hrs	
MAY2012_TO_AUG2012		opMay12 - 28236 - 7.5 hrs	opJun12 - 33443 - 8hrs	opJul12 - 39794 - 8.5 hrs	opAug12 - 36659 - 8 hrs
SEP2012_TO_NOV2012		opSeptOct12 - 36519 - 7.75hrs	opNov12 - 36485 - 7.75hrs		
DEC2012_TO_FEB2013		opDec12 - 53315 - 8 hrs	opJan13 - 53315 - 8 hrs	opFeb13 - 41826 - 8 hrs	
MAR2013_TO_MAY2013		opMar13 - 46531 - 6 hrs	opApr13 - 43753 - 5.25 hrs	opMay13 - 46531 - 5.5 hrs	
JUN2013_TO_AUG2013		opJun13 - 54440 - 6 hrs	opJul13 - 46273 - 4.5 hrs	opAug13 - 46358 - 4.5 hrs	
SEP2013_TO_MAR2014		op - 61183 - 12 hrs			

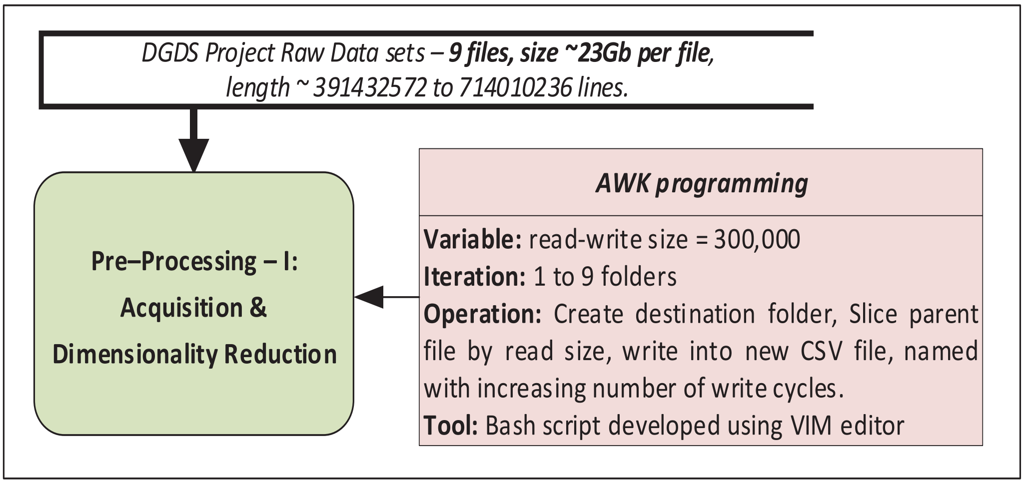


Figure 6.6 UML Representation of Preprocessing - I phase

Output: The reshaped divided files are stored in the 9 folders, and each file has rows starting with the device IDs and subsequent columns giving the time and measurement entries. In each file,

- Columns 1 and 3, (columns 1 & 4) are entries of device ID and measurement values (numerical entries) read directly from the source files

  The 3rd date and time column (string values) read using Text scan and processed further by first converted into a Time-character array, then stored as individual cell arrays.

6.4.2 Pre–Processing – 2: Normalization

Stage 2 of Data preprocessing deals with Data Abstraction of individual Device readings, for the specific time slot for the day (00:29:00 hours to 16:30:25) which was fixed based on the input DGDS data set exploration; trial periods, behavioral stability, unique characteristics etc. observed by the parent project team. From this stage on, Python virtual environment is invoked for coding and processing. The input to this stage comes from the Divided_files folder generated in the previous step.

Table 6.3 Algorithm for Generation of Unique IDs

```
initialize collection of devices; for readings file in readings_dir;
        for entry i = 1 to <Number_of_files>
                For entry j = 1 to <end_of_file>
                        Extract (col1(j)) == x)
                        If x == unique_ID_list (k)
                                unique_ID_list (k+1) = x
                        else
                                return (2)
```

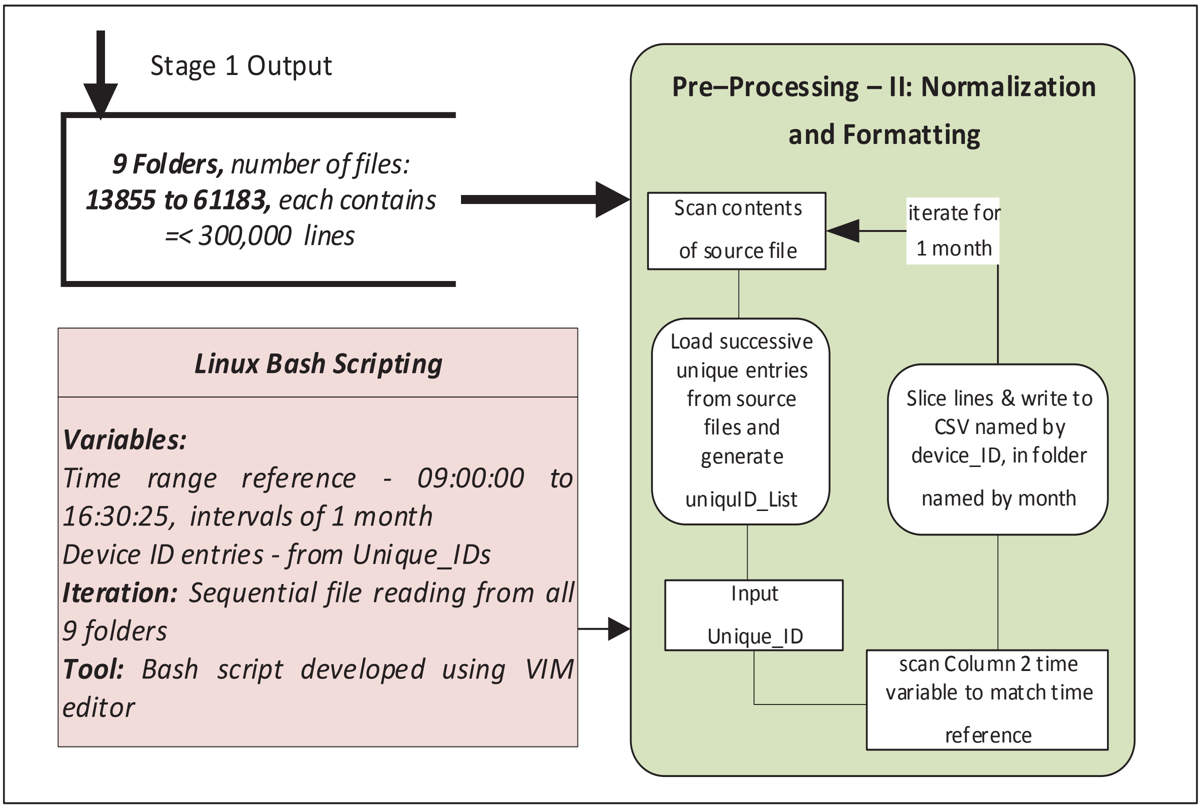


Figure 6.7 UML Representation of Preprocessing - II phase

1. *Generate Unique IDs:*

The necessary procedure to create paged format data for each device is to identify the component devices of a particular field experiment setup by creating a record of unique device entities The input is the sub-files of 300000 entries recorded in each of the main directories. Unique ID list is the output, which serves as a reference index for the next step. Iteration rate was 3.5 to 4 on average, with less than 5 minutes execution time in each case. For example, a total of 60124 unique ID s where identified in the March to May 2013 dataset. A global unique ids list from the 9 unique_ids files were also generated.

2. *Data Extraction into Monthly intervals:*

The time period based extraction process makes an output folder, and save all entries corresponding to each device ID (device ID = column 1 values) in a separate file within the stipulated time period, using the device ID as the file name. Data extraction implemented in parallel computing Python modules is relatively much faster and more efficient than the crunching stage via terminal scripting. But it is also more memory-hungry. This lead to process getting killed occasionally. However, overall speed increase is really high compared to the data crunching stage. Optimization was done based on the Pandas, TQDM and other libraries' documentation, based execution times of write-save-per-iteration or save-at-the-end operations. The ode was developed with regular write-save, due to a

Table 6.4 Algorithm for Data Extraction into Monthly intervals

```
initialize collection of devices; for readings file in readings_dir; enter time_window
        for entry i = 1 to <Number_of_files>;
                for entry j = 1 to <end_of_file>
                        if col2(j)== time_window
                                Extract (col1(j)) == x)
                                If x = unique_ID_list (k)
                                        create unique_ID_list (k)_csv && write col2(j), col3(j)
                                else
                                        unique_ID_list (k+1) = x
                                        return (2)
```

wide date range; execution time elongates but the process will not be killed. Catalogued entries are sliced based on observation time window, split point being Unique device IDs, forming pages of data for the final processed data. The saving part is significantly shorter, with an added progress bar. Trials started from a week, with the results being appended on to destination file after each iteration. Coding followed iterative execution method (appending outputs generated Date period selection made exclusive of set limits). The Average Execution rate during the factorizations was 3.5 to 6.8 seconds/Iteration, while the computation duration was 2.5 to 3 hours for 2000 files.

Output: Normalization via Parsing of Big Data files via Linux Terminal gave the divided component CSV files; indexed based on the timestamp and device ID, with measurement entries. Number of sub-files per period catalog was also recorded.

6.4.3 Pre–Processing – 3: Soft Clustering Devices for specific periods

This stage deals with Meta-analysis of sorted normalized data, prior to prime processing. It is viewed as a form of systematic review based on largely a statistical technique, taking the findings from several studies on the same subject and analyzing those using standardized statistical procedures. This helps to draw conclusions and detect patterns and relationships between findings.

Unsupervised clustering of devices based on data entry (column 3) trends was designed to aggregated data streams based on behavioral variations, to make visualizing tensor patterns decipherable. (1st dimension of the raw data tensor). The output of this phase creates reference indices for reading tensor pages (device ID-based measurement logs), each group consisting exhibiting consistent readings, or rather consistent reading patterns (similarity sets). Two methods were identified after checking the complexity of different clustering methods (unsupervised algorithms are

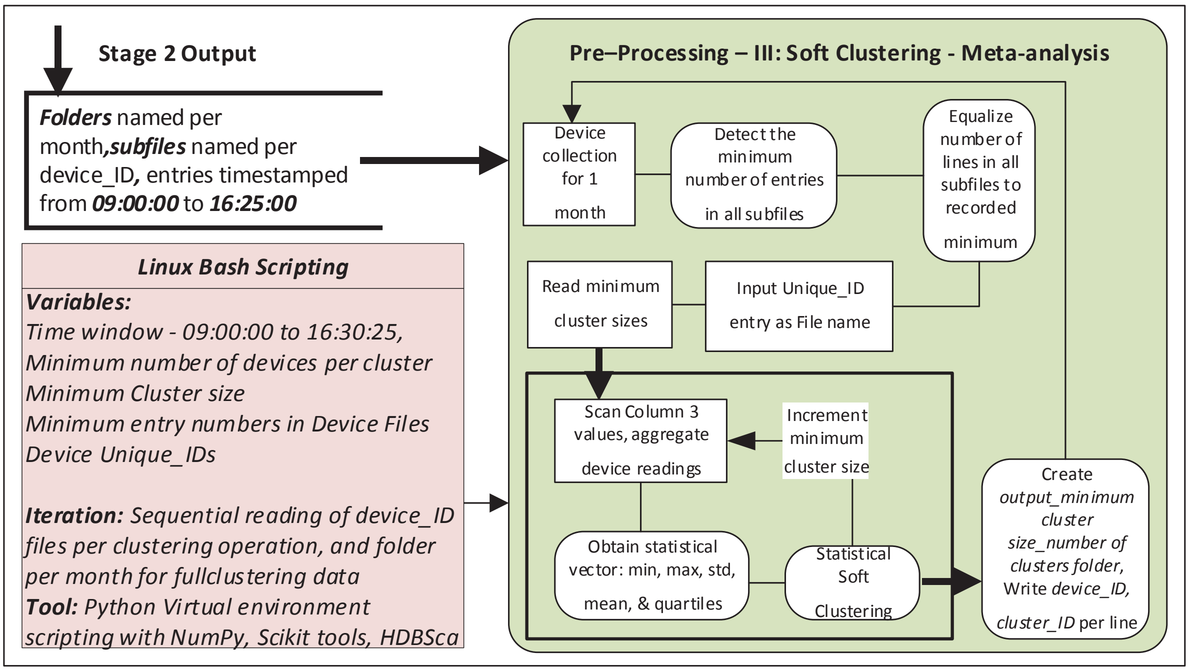


Figure 6.8 UML Representation of Preprocessing – III phase

quite slow). In the DGDS dataset, the nature of individual sets is unknown, although an overall feature description maybe can be commented upon, hence the unsupervised method.

- 1st Clustering Parameter: *Mean Shift:*

Data reduction is done with respect to the device ID or entity with the lowest number of measurements. Trials showed acceptable scaling with the large number of dimensions in the dataset under study, much better than basic K-means. This method was outperformed by the second option.

Table 6.5 Algorithm for Statistical Vector based Soft Clustering

```
initialize collection of devices; for readings file in readings_dir; enter time_window
        for entry i = 1 to <Number_of_files>;
                find i = $i with lowest number of rows = $a
                reduce number of rows to $a by averaging for all i
                for entry i = 1 to <Number_of_files>;
                for all entries j = 1 to <end_of_file>
                        7 statistics (min, max, std, mean, 25%, 50%, 75%)
                exit
        exit
        for entry i = 1 to <Number_of_files>;
                for minimum cluster size to maximum cluster size:
                    save clustering result of hdbscan(cluster size, devices) to csv file
                exit
        exit
```

- 2nd Clustering Parameter set: *Statistical Vector-based:*

This methods computes the measures of central tendency from each file and uses them as the clustering indices along with number of entries and non-zero entries, so the values for each file are now means from those chunks. HDBScan toolbox was utilized for scripting, and optimization for lower memory requirements, better speed and clustering efficiency on test data of a 10k devices over a week's readings.

Metrics: *Mean, SD, Min, Max, 1st Quartile, 3rd Quartile and total and non-zero* number of rows.

Output Visualization of Clustered devices

Clustering outputs for each set was saved up as an index list of device ID and probabilistic membership ID and saved as 'output_a_b.csv, where *a = $minimum_number-of_devices_per_cluster, b = $number_of_clusters (*e.g. output_5_155.csv.*)* The cluster indexed '-1' holding outlier measurements appear flat. While the training phase (extracting core energy of data from processed noise-free tensor), outliers that dampens the significant information are avoided. However, the final pattern detection includes data tensor with outliers, but core obtained from pure tensor without them. They are represent the anomalies. Minimum to maximum cluster sizes were 5 to 100. Visual indicators based analysis was performed on the plots of clustered devices in each clustering trial. Sample plots for clustered division output_75_23 and output_45_49 saved in the respective

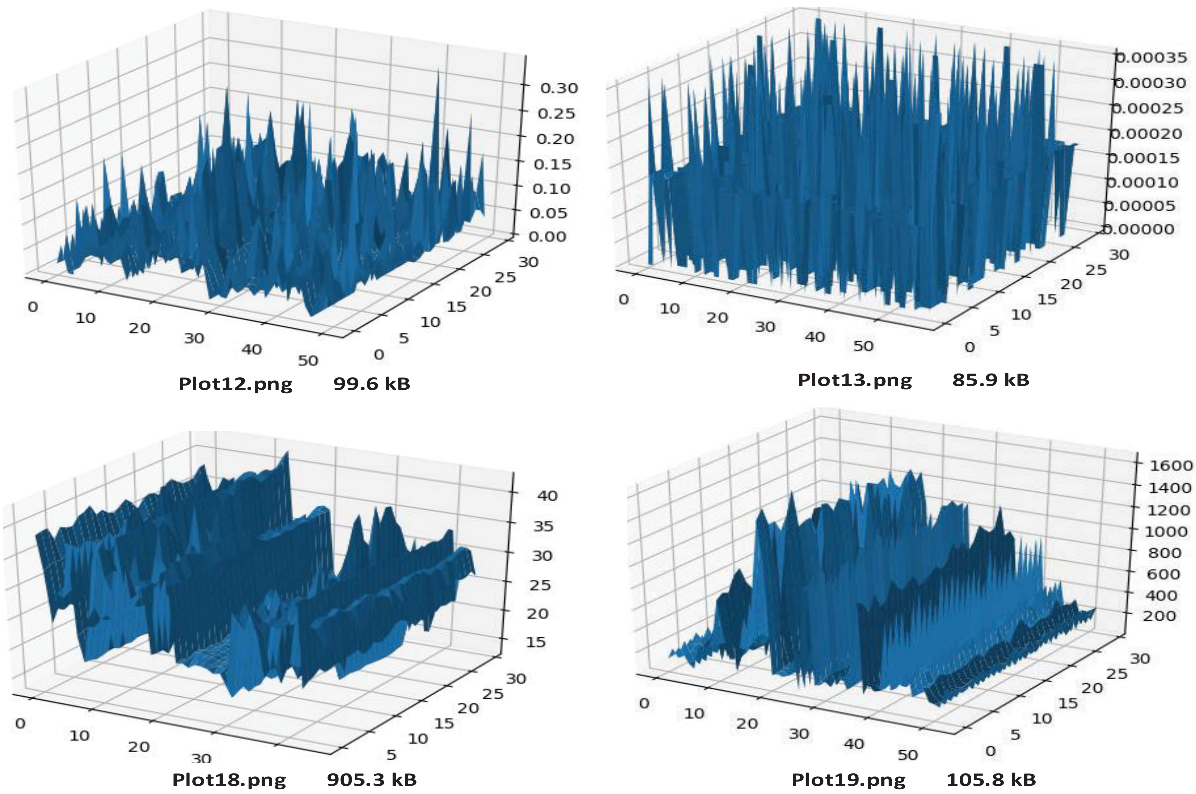


Figure 6.9 Device measurement plots for the clustered group output_45_50, MARCH_to_MAY_2013

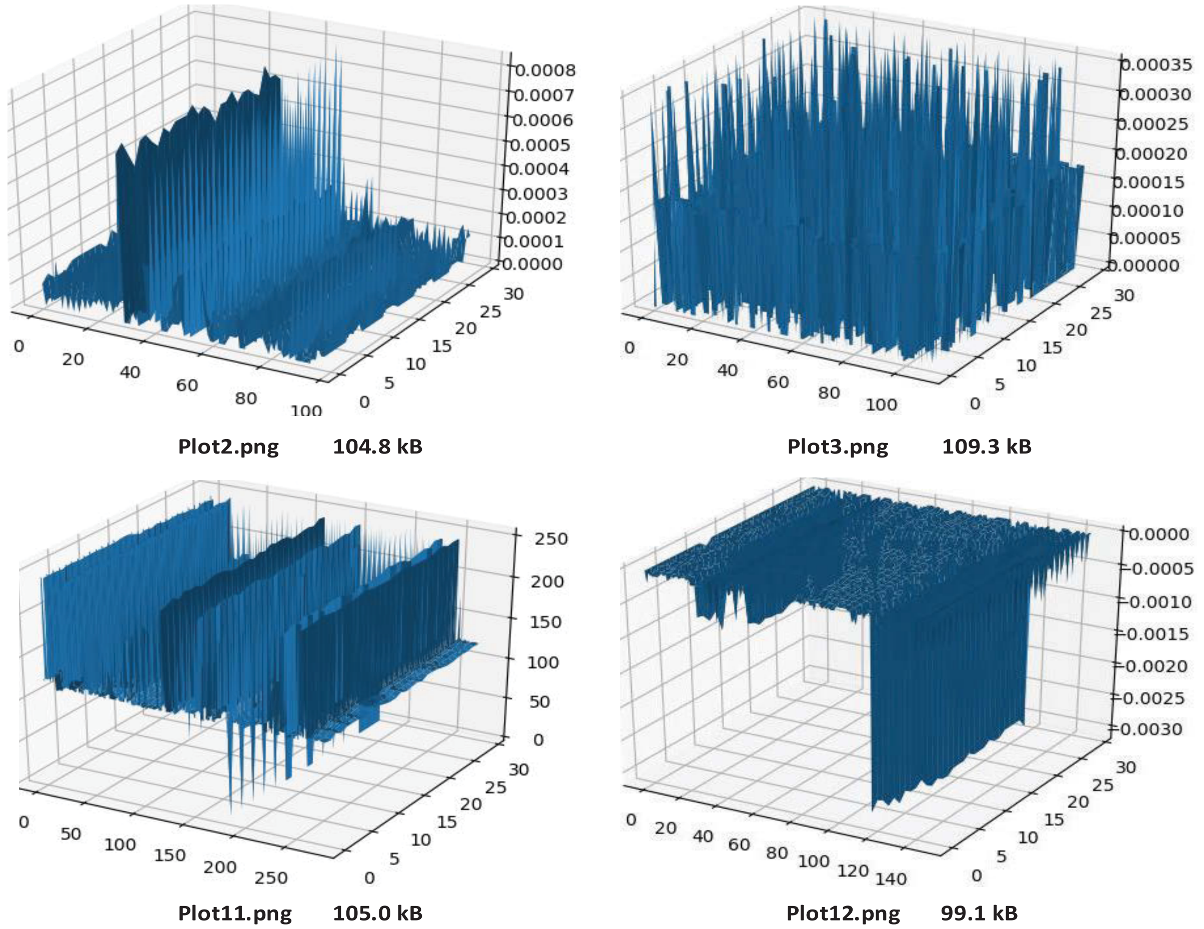


Figure 6.10 Device measurement plots for the clustered group output_75_23, MARCH_to_MAY_2013

Directories as shown. for clustered division output_75_23 and output_45_49 saved in the respective directories are shown in Figures 6.6 and 6.7. Each Clustering run behavior was analyzed and the optimum was selected with maximum divergence in member device output trends. Based on which an almost equal division showed maximum variability in cluster behavior with no similarities.

6.4.4 Tensor Processing: Tucker Decomposition

AIM: To study the space-time variant behavior of the DGDS real-time data streams, in terms of latent patterns along the modes defined in Figure 6.8.

PARAMETERS

The code applied to this stage is dependent on the tensorial framework as explained in Table 6.7

- **1** - Input data file directory;
- **2, 3** – device or time (interval to the program) or longer time window;
- **4** – How often readings are scanned and loaded into tensor;
- **5 -** How big is the 2nd dimension (in days) or paging parameter;
- The algorithm was designed with inclusive date selection.

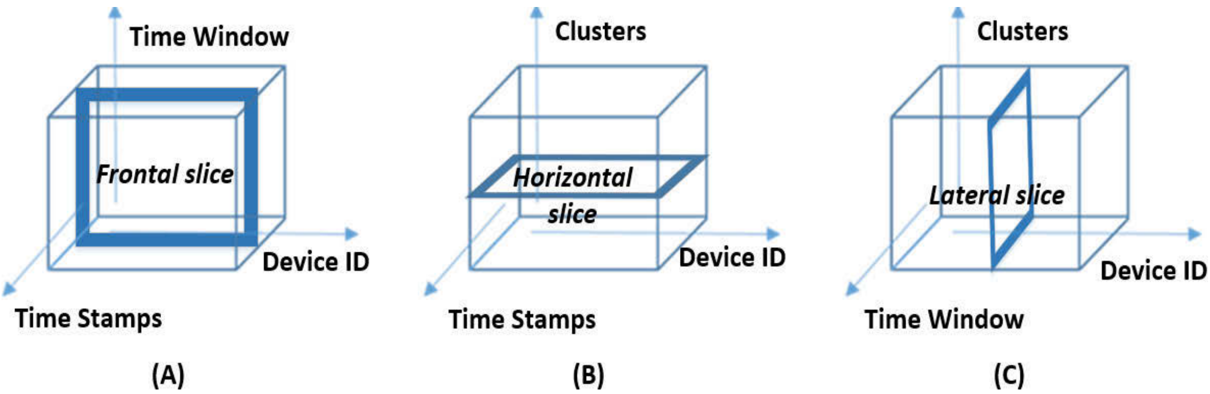


Figure 6.11 Standard tensor compositions investigated in this study, and the elemental pattern frames in the input.

The reshaped csv matrices (as a whole or partially, depending on parameters) from Pre-processing stage 2 are loaded into frontal tensor slices.

TENSORIZATION

Tensorization refers to the process of composing a raw data tensor from processed (here clustered) data, and save it as an N – way array, (memory format in JSON or CSV or any other platform compatible file format). The directory location and specific date and time range of the processed data, time window or split point, tensor pagination parameter, cluster file name (which has the minimum cluster size and number of clusters shown) and the cluster size is used to fix the tensor structure. The voids were smoothened without altering the data pattern by any fixed replacement. The method was to limit the transition across the void, by entering the average of the previous and next readings.

Table 6.6 Interpretations of the dimensions of the Data Tensors designed in the proposed method

Tensor 1	Time window = Number of days; Date range
Tensor 2	Time stamps = 15 measurements per day; [Time range / Time window + 1] * Tensor page OR ceil ((end_time - start_time) / time_window) + 1]
Tensor 3	[no. of clusters – 1 {outliers}] * cluster size = total number of devices
Core 1	min(Tensor 1 , no. of clusters - 1) = Tensor 1 OR variable
Core 2	min(Tensor 2 , no. of clusters - 1) = Tensor 2 OR variable
Core 3	min(Tensor 3 , no. of clusters - 1) = [no. of clusters – 1 {outliers}] OR variable
Factor 1	Tensor 1 x min [Tensor 1 x Core 1]
Factor 2	Tensor 2 x min [Tensor 2 x Core 2]
Factor 3	Tensor 3 x min [Tensor 3 x Core 3]

Table 6.7 Parameter inputs to the Tensorization and Tucker Decomposition scripts

No.	Parameter	Interpretation
1	*input_dir*	*output from data_extractor, e.g. outputMar13*
2	*start_date*	*e.g. 03.01.2013 09:00:00, dates are inclusive*
3	*end_date*	*e.g. 03.31.2013 16:00:00*
4	*time_window_minutes*	*how often do we take the reading, depends on the data collection model*
5	*tensor_page - number of days*	*split point of measurements along mode 2 / 3*
6	*cluster_file*	*e.g. output_75_23, which shows no. of clusters, & minimum no. of devices per cluster*
7	*cluster_size*	*no. of devices to be taken from each cluster*

TENSORIZATION PROCESS:

Tensorization refers to the process of composing a raw data tensor from processed (here clustered) data, and save it as an N – way array, (memory format in JSON or CSV or any other platform compatible file format). In our experiment, the code thus takes in directory location and specific date and time range of the processed data, time window or split point, tensor pagination parameter, cluster file name (which has the minimum cluster size and number of clusters shown) and the cluster size.

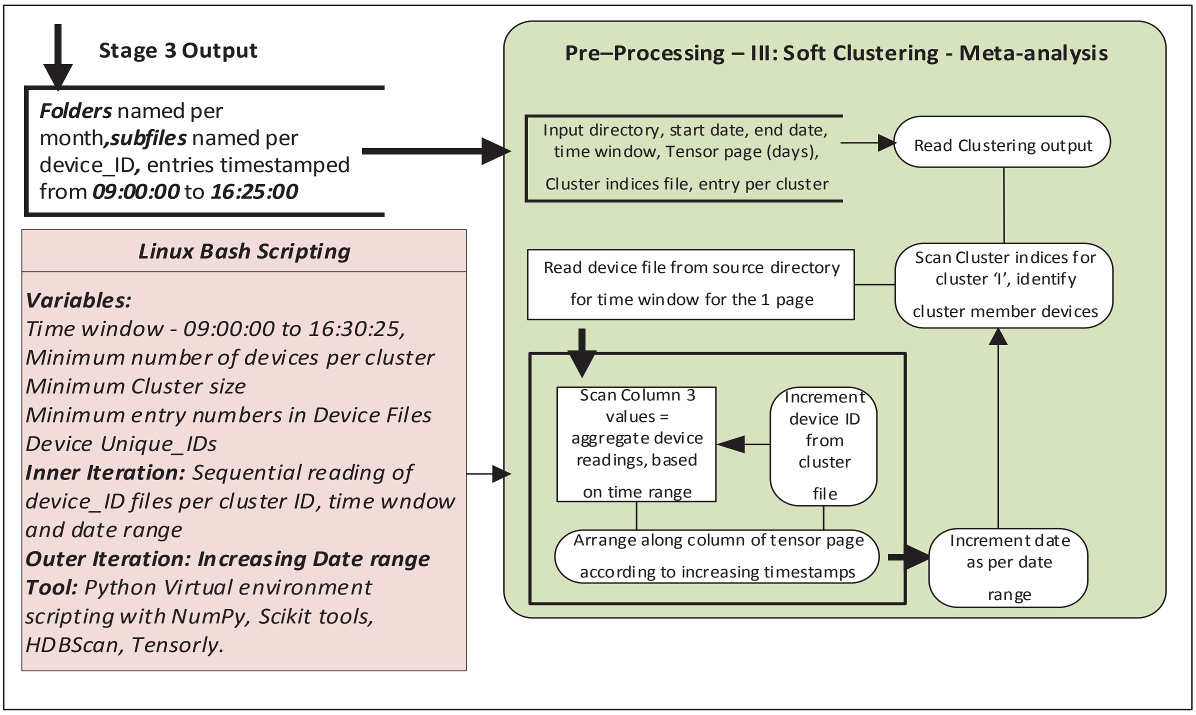


Figure 6.12 UML Representation of the Tensorization Stage

This step takes the DGDS output data for more than one month grouped in one directory, from the first data_extractor stage. This is needed to just apply clustering and then decomposition on it.

DECOMPOSITION

- **Tucker – 3 Decomposition** was chosen for Very Large-Scale Data, developed rom Tensorly library.
- **Anomaly detection model** was to observe the **Factors 0, 1 and 2** by fixing a core tensor, obtained from Tests frameworks A and B.
- **Parameters:** Date range = 5 to 60 days; time range 'n' = choice of time window of study for the day = 16:00 hours - 9:00 hours from the DGDS dataset characteristics; Tensor page = single time range period or 1 set of n * 60 minutes equivalent of readings; Time window = 30 minutes; Cluster size = < min cluster size.
- **Short time Span:** it is the daily timestamp indices; required dimension is 15 [readings across 7 hour window at 30 minute intervals], then tensor page = number of blocks of measurements from the entire date range that should fill each tensor page. Core component that is to be extracted is also a tensor of reduced dimensions, while the factors are 2 way matrices. Any core dimension can't be greater than the corresponding tensor dimension. While implementing in Python, core size is the last additional parameter for modelling tensor decomposition operation.

The various modes of the tensor data to be processed were associated against the required coding parametrization as follows, in Table 6.7

- 1st mode comprises of the "number of days" indices,
- 2nd mode is the "number of timestamps",
- 3rd mode is made up by the "total unique device IDs".

EXPERIMENT PHASES

In Figure 6.8, A, B and C are the three experimental frameworks devised. The key factor that is fixed while running the trials is the rule to set the core size. Dimensions of the core tensor depends on the sub-scale depth required for an apt representation of the original dataset information content. As per initial analysis and test runs, 3 months or more long data sets are large in quantities, and hence every step here forth needs to be carefully designed. Longer durations can be inspected, possibly 3 months in one go, but at the cost of lowering other parameters like the device number, or time window of observations.A and B were performed to study the short term pattern variations, such as less than a week. C is performed for durations 1 month or more. The tensor pages are files created as a result of the pre-processing stages. Each device ID used represents components of the various sub-

Table 6.6 Algorithm for Tucker Decomposition experiments

1	*initialize collection of devices; for readings file in readings_dir; enter time_window; clustering_file; tensor_page; cluster_device_numbr*
2	*initialize collection of devices*
3	*For cluster indices,*
4	*add to devices cluster size of best fit devices from cluster*
5	*for entry i = 1 to <cluster_IDs>;*
6	*read extracted output directory*
7	*find i = $i number of rows*
8	*initialize tensor by reshaping data from devices*
9	*exit*
10	*increment time cluster_ID*
11	*return (5)*
12	*apply chosen decomposition algorithm to tensor*
13	*for all entries j = 1 to <end_of_file>*
14	*save entries >> JSON*
15	*Hinton Plot (JSON)*
16	*exit*
17	*for entry i = 1 to <Number_of_iterations>;*
18	*calculate reconstruction error*
19	*calculation variation index*
20	*exit*
21	*exit*
22	*Print, Tensor shape, Core & Factors shape, error, variation, number of iterations $i*

systems of the Smart Grid metasystem. Timestamp labels of the different measurements are at half an hour intervals (as the measurement recording frequency in the given dataset is mostly 30mins). The values that fill in the column 3 cells of the array, are based on row 2 and column 2 labels.

1. Design A: Devices vs. Timestamps vs. Time window framework

This model studies samples of spatio-temporal-subscale behavioral patterns across the time subscale, i.e. general system behavior over longer time ranges, of the system as a whole without cluster based division (neglecting feature annotation based on kind of device). It achieves time-series data analysis, across spatio-temporal domain and identifies the subscale of pattern variations. Representative members of the devices from each cluster is aggregated in first dimension (global system behavior reflected).

In this experiment run, representative members of the device clusters aggregate along the first dimension as a continuous list to represent all possible behavioral groups. The timestamps of data measurements is in the 2nd dimension. The longer time window duration of observation is the 3rd dimension. The expected input data format which is normalized and structured is as in figure 6.10.

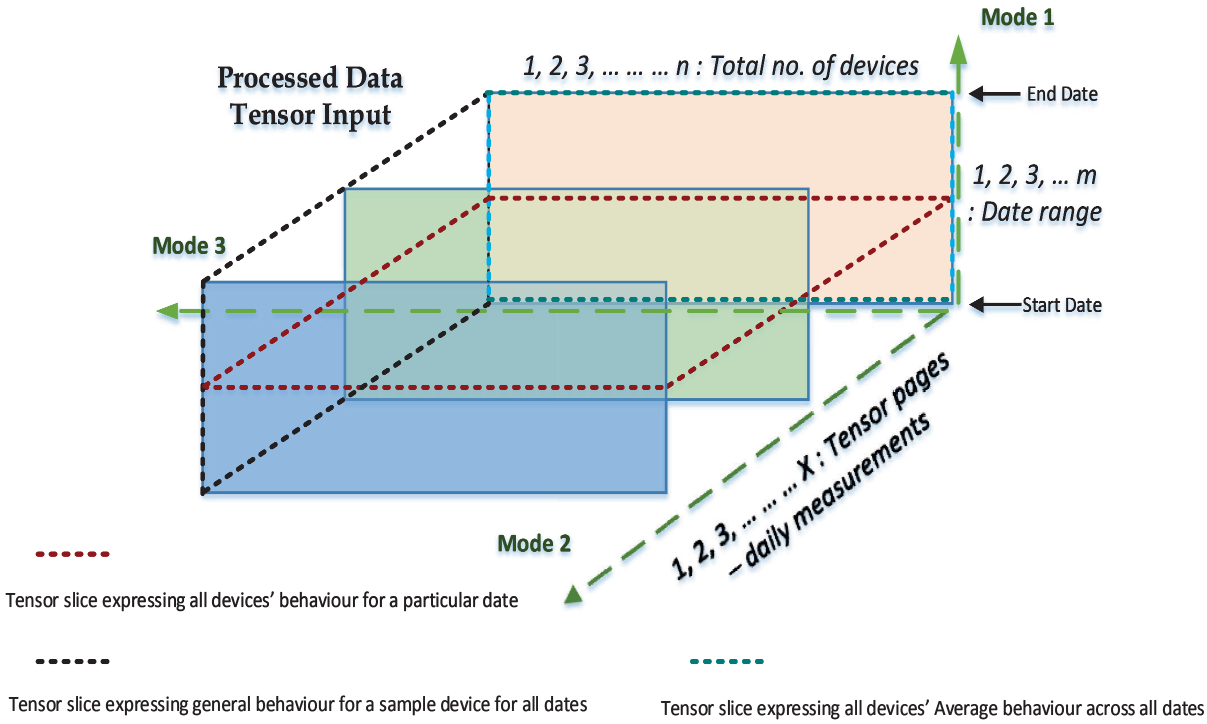


Figure 6.13 Device vs. Time window vs. Timestamp experiment - Input tensor structure and knowledge content

Training Run:

- Selected data, condensed into dates and weeks along dimension 2 and 3 was decomposed (Device vs. Timestamp vs. Days per month).
- Unfolding / matricizing it along the month dimension gave Device vs. Time behavior per month on a continuous daily time scale, equivalent of the core matrix.
- The tensor dimensions (Device vs. time vs. date] may be determined as:

((Dev x (7*2 daily) x (30days*12months*5years))) = ((Dev * 14 * 1800)) OR

((Dev x (7*2*7 weekly) x (4weeks*12months*5years))) = ((Dev * 98 * 240)) OR

((Dev x (7*2*30 monthly) x (12months*5years))) = ((Dev * 420 *60)).

Where *Dev = minimum number of devices * number of clusters*. For *cluster_id* and *min_cluster_size file* as parameters, the data frames up as tensor as shown in Figure 6.10. The split point is the depth along the time stamp mode, which is *((start time – end time) * 2) = 14 time points.* It can take any value depending on the given parameters, (weeks/5 days/10 days etc.). The representative number of devices per group was set to 10, so any behavioral changes will be reflected across 10 slices of the data tensor.

Testing phase:

Core matrices for longer time windows were concatenated to obtain Device vs. Days vs. Mon-

-ths Tensor core for Testing. Result is a ((Dev vs. (7*2*30 timestamps) vs. (N months))) structure, which can be applied to the merged monthly data, for real time processing. Repeating the decomposition for merged output, but this time with the merged core applied on the bigger data directory, extracts fragmented patterns across the pattern output, which resembles concatenated pages of the Training phase factors.

It was observed that lower core sizes improved reconstruction error marginally with drastically faster execution and lesser memory. Therefore a core size made bigger, has greater influence on memory than processing bigger tensor.

2. Design B and C: Cluster vs. Devices vs. Timestamps or Time windows framework

This model studies samples of spatio-temporal-indexed behavioral patterns i.e. system behavior over scales of observation window (both long and short), types of devices and spatial indexing. It achieves time-series data analysis, across the cyber-physical space; that means identifying every possible subscale pattern across time, different subsections and spatial arrangement. It generalizes patterns across different CPS sections and over longer time scale in detail (In short it is Device ID vs. Time stamp vs. Time window for each cluster, all at the same time). Representative or

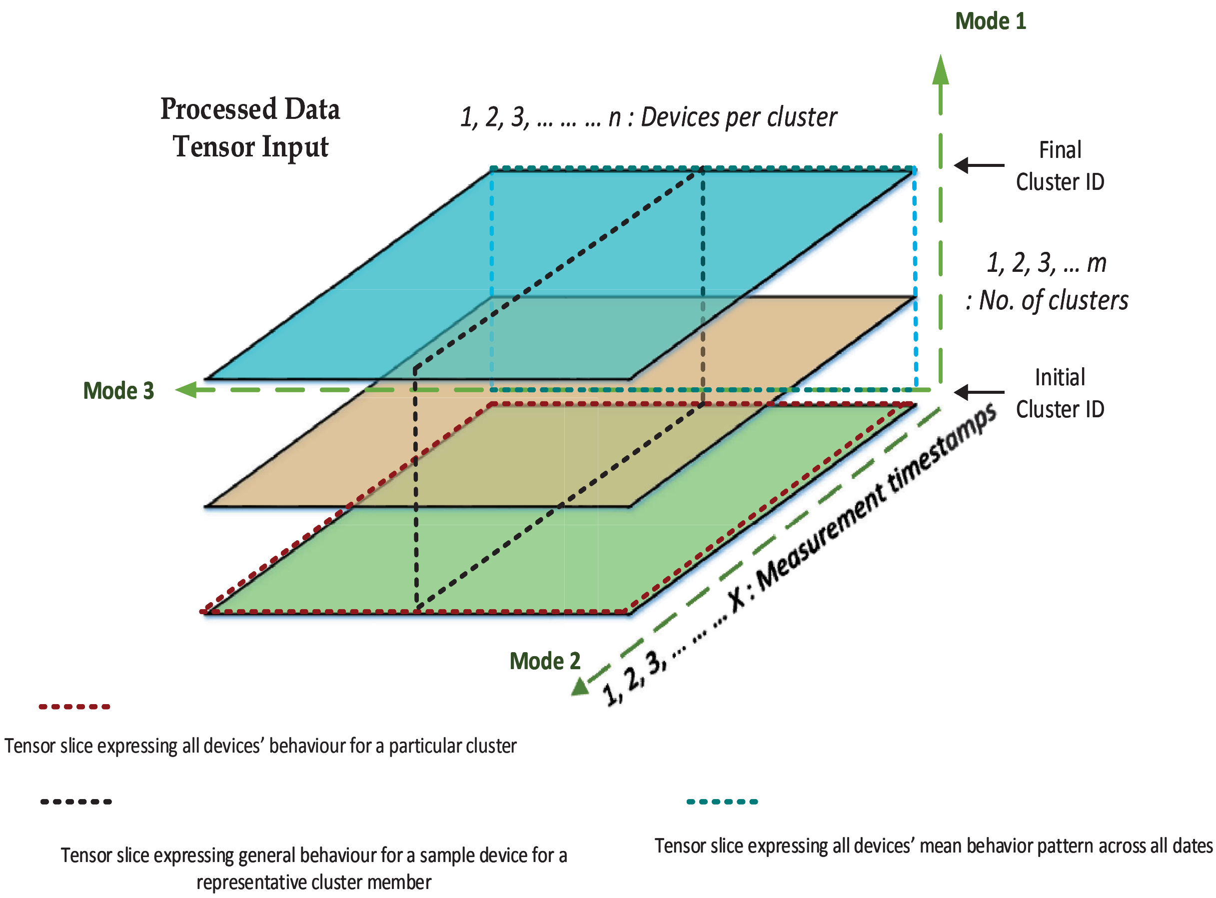


Figure 6.14 Cluster vs. Devices vs. Timestamps or Time windows framework - Input tensor structure and knowledge content

full set members of the Clusters occupy the first dimension. Members per group along the second dimension, and small-scale timestamps of data measurements or longer time window duration of observation along the third.

Training Run:

- Longer time windows gives Behavior-qualified-by-Device_Types (integrated behavioral patterns for specific groups, over longer periods).
- Measurement timestamps under observation along the 3rd dimension helps observe shorter spans or short term pattern data for the various test implementations, which is Behavior-qualified-by-Time (differentiated behavioral patterns for specific groups, over local spans)

Testing phase:

The core obtained from the clustered data processed by Framework A, and clustering indices or probabilities across the chosen specific time period is used as the decomposer reference metric.

OBSERVATION

In both Testing phases for Frameworks A and BC, the patterns obtained were compliant with the Training phase patterns unfolded, or rearranged. The reconstruction errors during the Testing phases showed very high accuracy, validating our approach.

- Experiments were performed for various core sizes, starting with 5, while incrementing other parameters, especially the 1st mode.
- Lower core sizes yielded better decodable patterns and reconstruction errors. The same held for date range more than 1 month, and cluster size balancing the minimum number of devices per cluster (45_43 set).
- If timestamps of the data to be loaded were not the same, or different, then values within the standard intervals are taken, and used for the corresponding timestamps.
- In case of multiple timestamps within the standard 30 minute interval, median reading was taken.
- For lower dimensions of tensor or core shapes, decomposition was faster, using less memory.
- The bigger core size has bigger influence on memory consumption than framing a big tensor. The final code versions were made for basic TD with missing voids smoothening, and the same implemented for Tensorization and decomposition with the option to save the factors.

A total of 14 trial experiments were made on The Tucker Decomp algo developed in Python on The Dataset March 2013 to May 2013. The following were the selected parameters with the corresponding measured output characteristics:

Initial testing culminated in the choice of the following parameters, based on the parameter variables defined:

Date range = 28 days; time range = choice of time window of study for the day = 16:00 hours - 9:00 hours OR 7hrs * 60 minutes = 420 minutes; Tensor page = single time range period or 1 set of n * 60 minutes equivalent of readings; Time window = 30 minutes; Cluster size = < min cluster size.

The subscale intermodal patterns represent semantic associations among the various modes, as modal overlaps. It qualifies as a similarity index. Hence joint pattern sets of common indexing shows a class of similarity measures, characterized by co-occurrence. However, the multiple facets introduce challenges of modelling, as a result, most prior work has focused on specific points, such as applying two-mode relations.

Throughout out the result and discussion section, the sizes of the data investigated are kept relatively small. This is because the graphs and plots generated become less visibly interpretable. However experiments were conducted over larger time windows and have corroborated the results until 1 year periods.

6.4.5 Discussion

The results are the factor matrices showing intra-modal behavior across any 2 modes of the data tensor, and the Tucker core tensor which represents the weightage of each of the factor along each slice dimension of the tensor. They are displayed as decomposed factor parameters and plots. On identifying the range of valid core and factor dimensions, the results were stored for further analysis. Segmentation fault bugs occurred during this stage, primarily due to memory saturation. They were solved by salvaging run-time to save resource consumption.

Once the required factor**s** were estimated, the code added the feature of saving the obtained factors as JSON files. After evaluating of the core and factor matrices variations and their effects, core size was turned a variable. These on the other hand, affects the tensor and factor dimensions. Appending the results was also made possible.

If any one of the core dimensions exceed the minimum cluster size, then that points out the lack of minimum number of representative points to aptly represent the full information, and hence is a violation of the Nyquist Criterion.

Also, for the 3 factors, if any 1 dimension << the others, then the factor is displayed almost like a line, or a narrow column, because the 1st dimension is small. The data is hardly visible or shows any pattern. This may be overcome by data framed in the tensor such that all 3 dimensions are balancing (magnitudes of roughly same range).

Throughout out the result and discussion section, the sizes of the data investigated are kept small. This is because the graphs and plots generated become less visibly interpretable. However experiments were conducted over larger time windows and have corroborated the results until 1 year periods.

1. Tensor representation:

The Tensor Data and core are represented with color Hinton plot. It helps to easily identify dominant relationships between components. It depicts more interpretative behavior than Iso-surface visualization and Slice-mode visualization. A Hinton Diagrams blob at position (i, j) has size == orthogonal index G_{ij}. It means that Hintons expresses the interaction of j^{th} component gj$(:)$ with the modes IJ and JK, as well as the corresponding elements. Hinton Diagrams also signify Joint Rate indices, which are measure on component interactions.

2. Output Visualization - Core and Factor:

The factors are also plotted as Hintons, as well as saved as JSONs. Hinton diagrams are used for all core and factor analysis. The size of individual block entries represent the magnitude of the factor component, while the color shows if that is a power draw from the network (positive or white), or a power injection (negative or black). The former stands for customers, battery charging or grid monitoring equipment logs showing power loss.

The latter represents non-renewable generators like Solar panels from the Prosumers (Producer-Consumers), batteries compensating during load shedding, and dynamic demand-side management. Runs of white or black entries refer to generation or pure consumption units.

Core Tensor – it represents the energy of the Data Tensor block, or the total valid information content. The 45 data points of Mode 1 shows daily subscale behavioral patterns of the various devices across the observation window. The 45 data points of Mode 2 is the representative value for every device across various long time windows (weeks), an aggregation of all daily measure. The 45 data points of Mode 3 gives signal points on characteristic observations aggregation of information for the individual component over timescale and time windows.

Factor A1 – It shows the behavioral pattern of the overall system over a week's time subscale across the entire observation window.

Factor A2 – It is the averaged pattern profile of every system component over a sample time window of 1 week, representative of the larger time window of observation.

Factor A3 – it has the mean information contribution from every data stream (every information source) over the longer observation window averaged over the time scale.

Data tensors thus had to be framed such that all 3 dimensions were balanced, where:

- Very small core sizes held, a factor of the number of consistent soft clusters.
- Other parameters may be bigger, especially Mode 3 of number of devices, gave better results.
- Time windows had to be at least 1 month.
- Cluster sizes which gave approximately balanced number cluster members & total number clusters exhibited better patterns.

All these observations were coherent with the described SGSC experiment trial descriptions.

1. Factor – 0 Analysis: Mode 31

Factor zero plots were obtained for test conditions of the core plots, as in figure 18. The trials were conducted for increasing tensor page size, which controls the data stacking. This allows determining how often the mode 2 & 3 patterns were scaled. Time windows or Date ranges were varied from 30 to 240, according to data ranges multiplied by 7 daily measurements. Tensor page was set to 1, 5 and 10 as empirical subscale frequency. It was observed that for small date ranges, cluster vs. devices vs. measurements were observed. Image no. 17 of figure 6.12 shows how the different categories of devices (clusters) varied across the month (30 day period). Images 4 -16 shows the behavior on device vs. timestamp vs. time behavior for more than a month time window (3, 4, 5 month intervals).

The 2 month interval showed aggregation of patterns, where the grouping of dependencies followed a different rule, as depicted in Images 1 – 3. This confirms with the SGC recorded observations of climate, power grid dynamics and customer involvement based trends on the data. Thus a time window of 30 days confirmed on redundancy period for the smart grid setup data. The

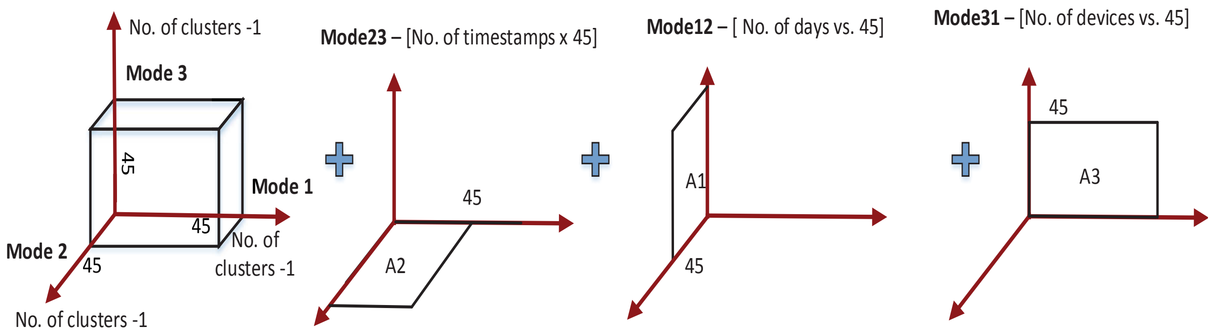


Figure 6.15 Interpretation of the obtained factors

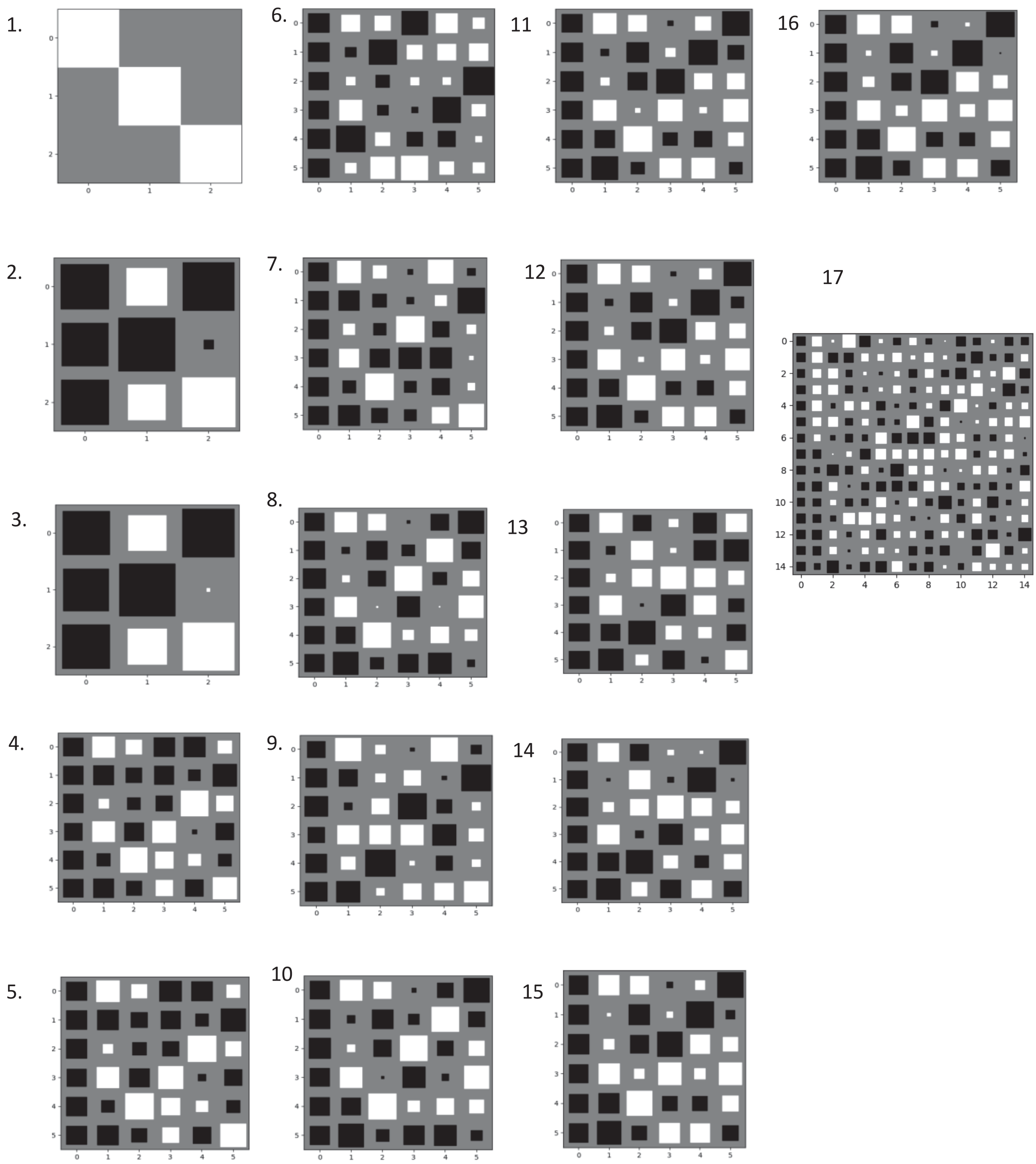


Figure 6.16 Hinton Plot Analysis of factor 0, March to May 2013

inputs compendium documents did explain how market and customer trends reflected on monthly cycles of power grid equipment performance, electrical installation dynamicity etc., over both Summer and Winter peak reduction trials.

2. Factor – 1 Analysis: Mode 23

Mode 23 represents aggregated trends of component data streams across a single time window, in Figure 19. Image 1 shows uniformity of the various clusters across devices for a small time window, 30 days. It shows that the methodology adopted was relevant to the data distribution. In 2 month time, the model now extracts within-the-cluster interrelations across various streams. It reflects on different component devices and their interactions aggregated on a daily window. Images 2 – 4 shows trends over 3 months and 8 months. It was observed that 1 month cluster vs. device vs. time format characterizes the data generating setups, while 2 months exhibit short term and long term time behavior also. For 3 month onwards interval selection reflects on overall and subscale parameters.

3. Factor – 2 Analysis: Mode 12

This factor shows dependency across the full data streams under consideration across the tensor, irrespective of pagination. The fourth image shows the complete setup behavior for a 30 day period. It proves the redundancy across various clusters by the empty spaces, and hence the need to store only representative values. Image 3 is the factor during 3 month period, which gives long term changes in data behavior irrespective of system and device characteristics. Redundancy is again observable. Since the trial data set here was from summer peak reduction experiment, the 5 month

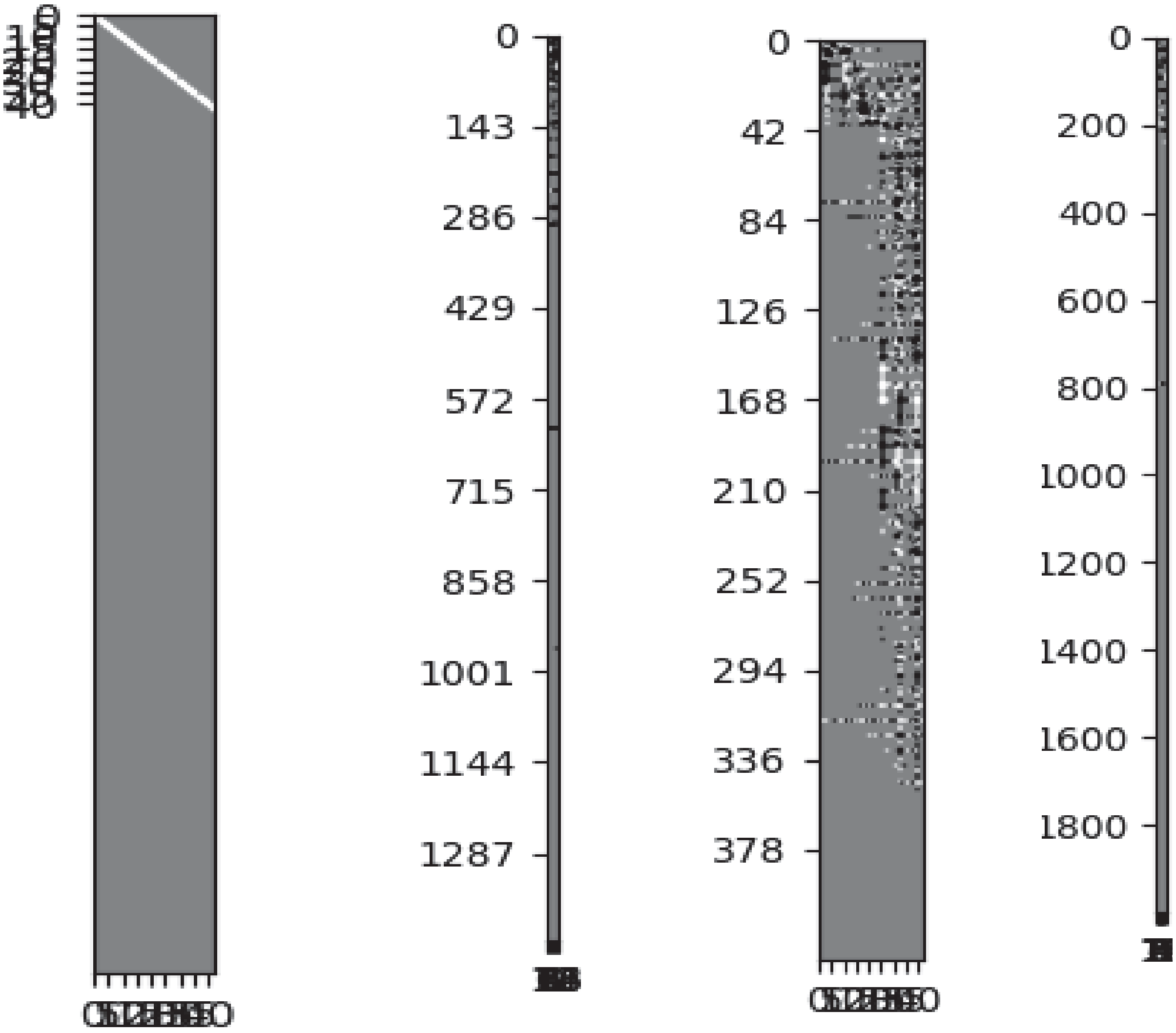


Figure 6.17 Hinton Plot Analysis of factor 2

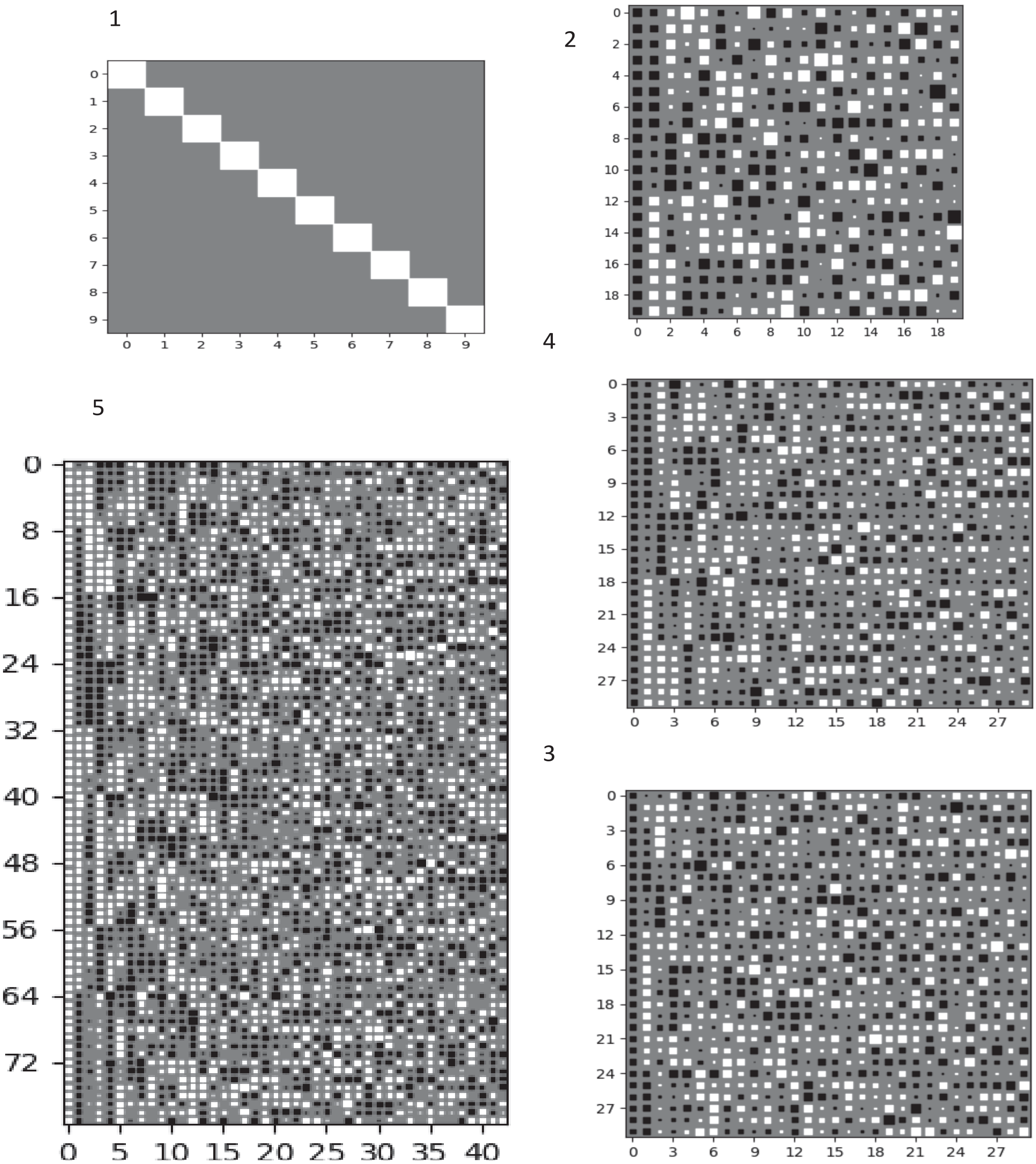


Figure 6.18 Hinton Plot Analysis of factor 1

observation period in image 2 shows periodic cycles. Image 1 was obtained in the 240 days or 8 month duration of one complete data set, aggregating net cluster performances.

4. Reconstruction and compression tests

For both cluster vs. devices vs. measurements (short time windows) and devices vs. time stamps vs. measurements (long time windows), reconstruction efficiency was greater than 99 %. Compression ratios for the March to May 2013 was obtained as below:

Raw data – 19 GB; Pre-processed files – 5 GB; and post TUCKER Decomposition: for each month period,

Table 6.7 The reconstruction erros and variations from original tensor in the testing pghase, averaged for the datasets between 2012 and 2014; the 0.0 entries were of resolution E-20 or more

Number of Devices per cluster & 'number of clusters'_'minimum cluster size'		***Reconstruction Error***	***Variation***
10	output_45_44.csv	0.000610629474052571	-2.1397065976675E-13
20	output_45_44.csv	0.004711560753764950	4.2935967281554300E-13
30	output_45_44.csv	0.000862445232131514	-1.77002943552162E-13
10	output_75_23.csv	0.000750303057149116	0.0
10	output_75_23.csv	0.0102459173293968	8.76711897523918E-14
20	output_75_23.csv	0.0353165522457474	1.91791027503995E-14
30	output_75_23.csv	0.0438020146543713	-6.90419943438769E-15
40	output_75_23.csv	0.0456620148231507	-1.70696790036117E-15
50	output_75_23.csv	0.0563942465722052	0.0
60	output_75_23.csv	0.0593242378070183	1.12271303365218E-14
75	output_75_23.csv	0.06301536905610669,	0.0.
100	output_75_23.csv	0.066861716885013	-3.15025783237388E-15
75	output_95_22.csv	0.07851639212646445,	-3.95516952522712e-15.
100	output_95_22.csv	0.0855146448419263	-4.10782519111307E-15

Table 6.8 The average Decomposition execution rate for all considered groupings across 2012 - 2014

No.	Clustering	Rate
1	output_5_483	15.56it/s
2	output_15_153	12.47it/s
3	output_25_84	25.83it/s
4	output_35_64	26.90it/s
5	output_45_44	11.32it/s

No.	Clustering	Rate
6	output_55_36	17.67it/s
7	output_65_27	18.79it/s
8	output_75_23	11.42it/s
9	output_85_22	19.03it/s
10	output_95_22	25.35it/s

Core tensor: ~ 104 Kb (Hintons) and 4.4 Kb (JSON)

Net Size acquired in decomposed data set - ~ 414.9 Kb (Hintons) and 152.91 Kb (JSON)

Factor 0: ~ 12.6 Kb (Hintons) and 3 Kb (JSON), Factor 1: ~ 8.4 Kb (Hintons) and 0.97 Kb (JSON), Factor 2: ~ 13.3 Kb (Hintons) and 42.6 Kb (JSON)

Compression ratio: 12.63e3 in terms of the preprocessed data alone.

6.5 KNOWLEDGE MINING: REFERENCE FEATURE MATCHING

The subscale intermodal patterns represent semantic associations among the various modes, as modal overlaps. It qualifies as a similarity index. Hence joint pattern sets of common indexing shows a class of similarity measures, characterized by co-occurrence. However, the multiple facets introduce challenges of modelling, as a result, most prior work has focused on specific points, such as applying two-mode relations.

Statistical-neural approach to maximize knowledge extraction from large datasets of diurnal load profile. An evaluation of the effectiveness of two cost-reflective product offerings was done based on [150]. Diurnal Load profile contains 48 samples per day at half hour rate. The following sets of observations from the DGDS trials were observed and extracted from the Tensor-Pattern Processing of section 6.4.

6.5.1 Behavioral Observations

- Most small wind turbines provided their peak power output in the early afternoon between 1:00 pm and 3:00 pm.

Table 6.9 Test Runs for Device vs. Time vs. Time framework to identify the pattern subscale, over 3 month periods from 2013 to 2014, the prime trial duration of SGSC

Tensor shape	*Core Shape*	*factor 0 shape*	*factor 1 shape*	*factor 2 shape*	*time_window _minutes*	*tensor_page - number of days*	*cluster _size*	*Iterations*	*tensor_page - number of days*
3, 30, 430	3, 30, 43	3, 3	30, 30	430, 43	240	10	10	430/430	10
3, 80, 430	3, 43, 43	3, 3	80, 43	430, 43	60	10	10	430/430	10
6, 20, 430	6, 20, 43	6, 6	20, 20	430, 43	150	5	5	430/430	5
6, 25, 430	6, 25, 43	6, 6	25, 25	430, 43	120	5	5	430/430	5
6, 20, 220	6, 20, 22	6, 6	20, 20	220, 22	150	5	5	220/220	5
6, 20, 440	6, 20, 22	6, 6	20, 20	440, 22	150	5	5	440/440	5
6, 20, 660	6, 20, 22	6, 6	20, 20	660, 22	150	5	5	660/660	5
6, 20, 880	6, 20, 22	6, 6	20, 20	880, 22	150	5	5	880/880	5
6, 20, 1100	6, 20, 22	6, 6	20, 20	1100, 22	150	5	5	1100/1100	5
6, 20, 1320	6, 20, 22	6, 6	20, 20	1320, 22	150	5	5	1320/1320	5
6, 20, 1540	6, 20, 22	6, 6	20, 20	1540, 22	150	5	5	1540/1540	5
6, 15, 1650	6, 15, 22	6, 6	15, 15	1650, 22	250	5	5	1650/1650	5
6, 20, 1650	6, 20, 22	6, 6	20, 20	1650, 20	200	5	5	1650/1650	5
6, 20, 2000	6, 20, 20	6, 6	20, 20	2000, 20	150	5	5	2000/2000	5

- Battery units provided continuous electrical output of 1.5kW with waste heat used for hot water (runs of black with this value, showed no excess generation).
- Rooftop solar PV installations normally generate the majority of their energy between 10:00 am and 4:00 pm and reduce average network load (limited cloud cover), (runs of black Hinton entries).
- Residential peak load typically occurs between 4:00 pm and 9:00 pm.
- One-fifth of the total customers has an energy storage device, assisting with peak management (observed as steady white-black-white runs).
- From 1 February 2012 to 31 January 2013 battery availability was 75.52% (white-black-white runs varying in trends occupy more volume of pattern than the pure black generators).
- A total of 267 customer sites were identified as not having exported any electricity to the grid over the period July 2010 to May 2011 (proportional white runs).
- The R510 zinc bromine flow battery store up to 10kWh of electricity, about half the daily requirements of a typical household.
- The energy consumed by the battery during the "float" period from 8:00am to 4:00pm can be significant, with an average power consumption of 300 watts needed to keep the battery charged.
- A week time frame 31 Mar 2013 to 7 Apr 2013 includes an entirely sunny day (2 April 2013), an entirely cloudy day (3 April 2013).
- A day with intermittent generation due to some cloud activity (1 April 2013).

Apart from these, certain high level trial results of the SGSC-DGDS experimentation was also identified.

- Generation profile (runs of black) from PV systems does not necessarily align with the local network profile (steady white-black-white runs matching 10am to 4pm generation trend).
- Orientation of solar PV panels towards the west offer better photovoltaic output with the late afternoon summer peak demand periods (during December to February 2012).
- Winter Peak Demand Reduction (Scheduled battery operation), 2012 to 2013 - batteries charging during the early hours of the morning and discharging during the early evening.
- Winter Peak Demand Reduction (Scheduled battery operation), 2012 to 2013 - Light zones represent times where charging is preferred and the darker zones represent the typical peaks of winter between 6:00am to 10:00 am and 5:00pm to 10:00 pm.
- Summer Peak Demand Reduction (2013-13) - the output available under "ideal performance" from the 40 devices from 4:00 pm to 7:30 pm was an average of 91 kW,
- Summer Peak Demand Reduction (2013-13) - Feeder peak occurred between 3:50pm to 4:00 pm and the batteries were in "float" mode during this time.

6.6 CHAPTER SUMMARY

The hypothesis and proposed solution represents intermodal associations for Space-Time Feature Extraction and Processing for Multi-Modal Datasets, It will also validate the use of such mined patterns for high dimensional data processing to distil intelligent management functions. Tensor decompositions and Factorizations have proven to be the most effective, detecting and combining local features in a large data tensor, across more than 1 dimension, to aggregate into global and local features. It thus represents both temporal and spatial contextual information, through representing flexible dependencies. They cover the major disadvantage of Traditional time series methods using linear or other varieties of models – being unable to model high dimensional non-linear relations. The concept will contribute towards handling magnanimous quantities of information the future will showcase.

The exploratory study of the chosen data set allowed the identification possible 'events', which were then corroborated with the factor patterns extracted. The sense of the patterns analyzed and the time-variant observation of the SSC team were congruent in meaning, validating our theory. Other performance features were also studied, such as tensor pagination, reconstruction error, variation, compression ratio, and convergence and execution rates.

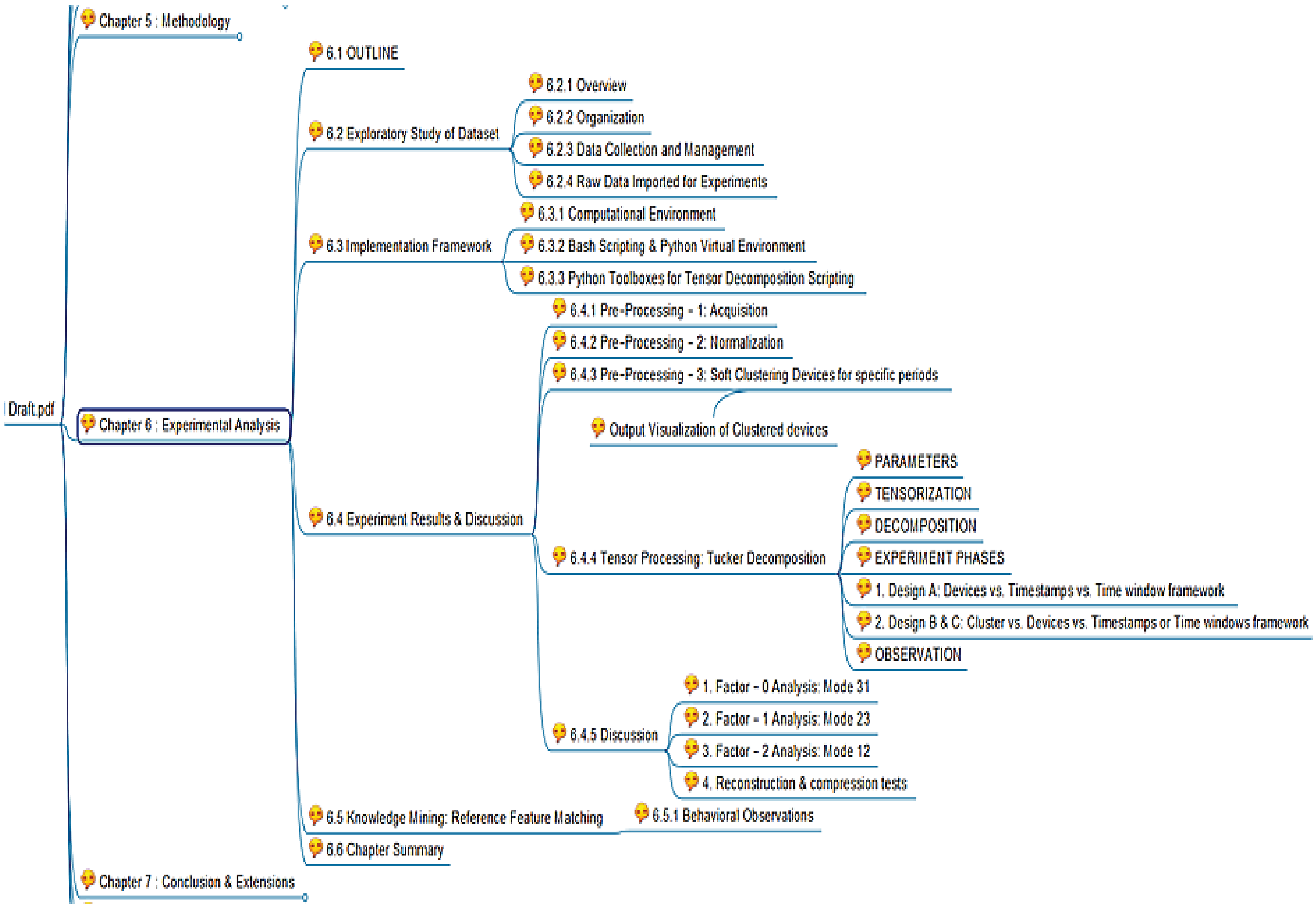


Figure 6.19 Mindmap outline of Chapter 6: Experimental Analysis

Chapter 7 : Conclusion and Extensions

7.1 OUTLINE

In this dissertation, a novel strategy is suggested to enable an entire Data Processing platform to work around the sensitivity measures of a core processing unit. The thesis defines the 'System-off-the-Loop' direction based on time series data flow processing. It ensures that the actual operation of the different CPS sub-systems (hardware or software) may be overlooked based on past behavior and responses. The methodology demonstrates how limitations from the main system response cycles can be overridden, leading to radical scale-down of various efficiency measures (processing time and latency).

Data Tensor processing based on Tucker decomposition has successfully enabled extracting high level qualitatively defined patterns. They are termed *Feature Pattern Annotations*, which will be recorded in a real-time controller unit to execute autonomic system management.

7.2 NEXT PHASE OF RESEARCH

Having verified the proposed solution to Multiway Big Data Processing, the various next stage approaches to promote the effectiveness for the solution, are summarized as follows:

1. Decompose the data tensors into Signals and Noise component subspaces. The core tensor is the signal PSNR content, and Noise subspaces will be the anomalous reading records. Further Adaptive filtering can be designed by setting dynamic convergence measures for rough clustering.

2. Anomaly Detection of multiway data can be easily modeled as a deep learning problem. They have been proven apt tools for high levels of abstractions through multiple layers of non-linear transformations. A typical operation that resembles the Multiway Anomaly detection operation in this thesis is combining a Clustering network and an Autoencoder with sigmoid activation function. K-means or SVM are suited clustering NN tools which efficiently performs unsupervised clustering. Block-wise processing of data tensors, called 'Block Tensor Analysis', will generate the parallel streams of the factors and core tensors for continuous data tensor blocks input. Aggregating each of these flows and training the corresponding Pattern detector network, will give us space-time variable traces of the patterns. The outputs maybe wired for further mechanization, like control signals to an autonomic management node of the SG-CPS.

3. A third improvement that's is applicable at this point is to upgrade the tensor processing into Dynamic Tensor Analysis (DTA). While the existing method will have a reconstruction efficiency and accuracy as high as demonstrated earlier, DTA allows a moving window to capture the tensor pages

to be processed. Thus, depending on the variability and behavioral dependencies' scale, the window sizes maybe varied. Correspondingly the pattern sizes maybe scaled within the network architecture. Window Tensor Analysis (WTA) is another approach applicable here.

The main advantage of extracting multi-modal patterns and characteristic spread components is the induction of autonomic behavior. Using such multi-way high dimensional patterns and specific characteristics, different system elements may be trained to function proactively. That is the capability to self-modulate behavior, output or signal flow, to meet a specific criteria or pre-defined operational status. Thus the new autonomic face of Cyber-Physical systems allows it them to self-regulate and adapt, rendering them intelligent.

7.3 FUTURE WORK: REAL - TIME PROTOTYPING

The future scope is this proposed system design is envisioned to establish time-updated multi-dimensional autonomic interfaces for a large-scale, heterogeneous, distributed SG infrastructure. Further Intelligent processing algorithm modules should be added to the proposed concept to handle and reduce intolerable effects such as adaptable distributes storage sub-networks and V2G and G2V connectivity and loads, through wide-sense distributed control mechanisms. As part as preparing groundwork for future directions multiple subjective, objective and practical inferences were drawn through extensive studies and investigations, within the scope of this thesis. The work-plan outline for further extension is as listed below:

- A flexible middleware platform for the CPS environments considered as the problem domain, with dedicated customization, to install the designed Tensor-based Anomaly Detection.
- Extract, analyze and optimize multiway high-dimensional data patterns (more than the 3 modes considered during experimentation).
- Develop Data and Signal patterns at the middleware, modelled as Tensorial units
- Apply Tensor Processing algorithms for detection, classification and optimization of anomalies, acting along multiple modes.
- Introduce a continuous time learning phase to process streams of data in real-time.

The objectives aimed to be achieved through the further extensions are listed as follows:

- Multi-fold reduction in response latency, execution time, computational costs, complexity, and time, power and hardware necessities.
- Autonomy concerning internal functionality in which standards are applicable and high abstraction level toward external environment for interconnectivity and interoperability.

- Elimination of complicated data pre-processing and dimensionality reduction of contemporary network architectures, and hence.
- Re-learning and optimization induced Systemic Intelligence.

7.4 CHALLENGES ENCOUNTERED

Throughout the implementation stage of this research work a significant number of difficulties were encountered. The main reason being the Big Data experimentation domain. As characterized by most Mass Data processing systems, performance of the pre-processing stage was heavily influenced by the computational environment. Also, being a novel research topic, has eliminated the chance of credible reliable data sets for trial and testing were scanty.

The main hurdles faced during the experimentation phase, which are relevant for future works and related studies, are explained below:

7.4.1 Pre-Processing

The most important issue was limited computational resources.

- The different 1 month period outputs that were generated in different folders after the Stage 2 of Data Preprocessing projected another trouble during the implementation. Some entries dint have readings or logs on all the selected days of the time experimental duration. Hence, while processing the Big data files the representation of such 'no evidence' events in the processed final high dimensional array input is a challenge. The elimination of such time slots is impossible since for studying behavior patterns across a group of devices, time frames will have to be consistent. Normally replacing void entries with zero would have sufficed. But here for Power Grid inputs, zero is a valid measurement. There for a different representation to symbolize missing values had to be framed.
- Problems due to Large-scale Data: raw data tensor and temporary variables stored in memory very large-scale = cause memory overflow error during decomposition. The solution that was devised is to leave such entries in the high dimensional array empty at first and proceed. After the data loading into the HD array was completed to form the tensor, the following action was taken - detect the lowest entry of the tensor, obtain 1/10th or 1/100th of it, and fill the null entries with this value.

7.4.2 Implementation

- Tensors, though faster, more accurate and parallel, embarrassingly are still much unpopular. The reason comes down to the libraries. Open source libraries for tensor analysis need to become more common. While it's true that tensor computations are more demanding than matrix algorithms, basic libraries if established, tensor-based algorithms will become much easier to implement than the latter.

- The scope to replace computationally voluminous and time consuming solutions and numerical computations is large. But the technological advancements of high performance hardware is not up to the mark for a corresponding software or theoretical solution, which makes realization dodgy. Yet, recent improvements in parallel and distributed computing systems have made tensor techniques feasible.

Experimentation was also limited because of lack of resources for Big Data processing (limited access of FEIT HPC cluster computing facility), and also my own issue of not having a software-related back ground. Luckily, it was my learning phase that helped me develop junior level Data analyst programmer skills.

7.5 Summary

Current technologies generate Big Data in higher order and dimensions with patterns and associations across all modes. Extracting them will isolate multi-way patterns which represent the subscale relationship of different data streams, i.e. the sources. This will help develop autonomic intelligence.

The timeliness of CP infrastructure, the main BigData sources, may be heavily impacted by unpredictability issues of real-time applications. They are mainly susceptible to performance challenges, which will be worse off with such a network level imbalance. Such issues may be effectively mapped as the patterns across different data streams from various sources.

This thesis presents a proof-of-concept for System for Space-Time Feature Extraction and Processing for Multi-Modal Datasets. Computational Intelligence in Multiway BigData streams are isolated through Tensor decomposition rules and the corresponding factor tensors and matrices.

The study was oriented around the most commercially viable Cyber-Physical Installation – Smart Energy Grids. The colorful vision of a 'revolutionized grid', in essence, is one that think, act, behave and care for itself – the apt terminology being 'smart self-management'. To solve the pressing concerns regarding the evolution of next-gen intelligent power grid ecosystems, the hypothesis uses such a futuristic deployment as the data source. The Smart Grid Smart City project Commissioned by the Department of Environment and Energy was the most successful deployment of pilot Smart Grid installations, as well as the most widespread CPS installation being trialed and prototyped. Naturally, it fit within the study domain as the data input source.

The thesis discusses some of the major milestones and hurdles with regards to the collaborative design and integration of the existing smart grid-power system components. To create an advanced self-sustained configuration, Computational Intelligence (the basic math and logic behind AI) will realize these future energy systems, allowing flow of electricity, data and control signals in all directions. We thus propose an information model which will allow future energy systems to be integrated onto the existing power transmission and distribution scenario, and allow rapid development of protocols, architecture and standards. It will lead to an integration standard for interactive coordination of different sectors of SGs.

The hypothesis and proposed solution deals with representing intermodal associations for Space-Time Feature Extraction and Processing for Multi-Modal Datasets, It validates the use of such mined patterns for high dimensional data processing to distil intelligent management functions. Tensor decompositions and Factorizations have proven to be the most effective, detecting and combining local features in a large data tensor, across more than 1 dimension, to aggregate into global and local features. It represents temporal, spatial and contextual information, through representing flexible dependencies. They cover the major disadvantage of Traditional time series methods using linear or other varieties of models – being unable to model high dimensional non-linear relations. The concept will contribute towards handling magnanimous quantities of information the future will showcase.

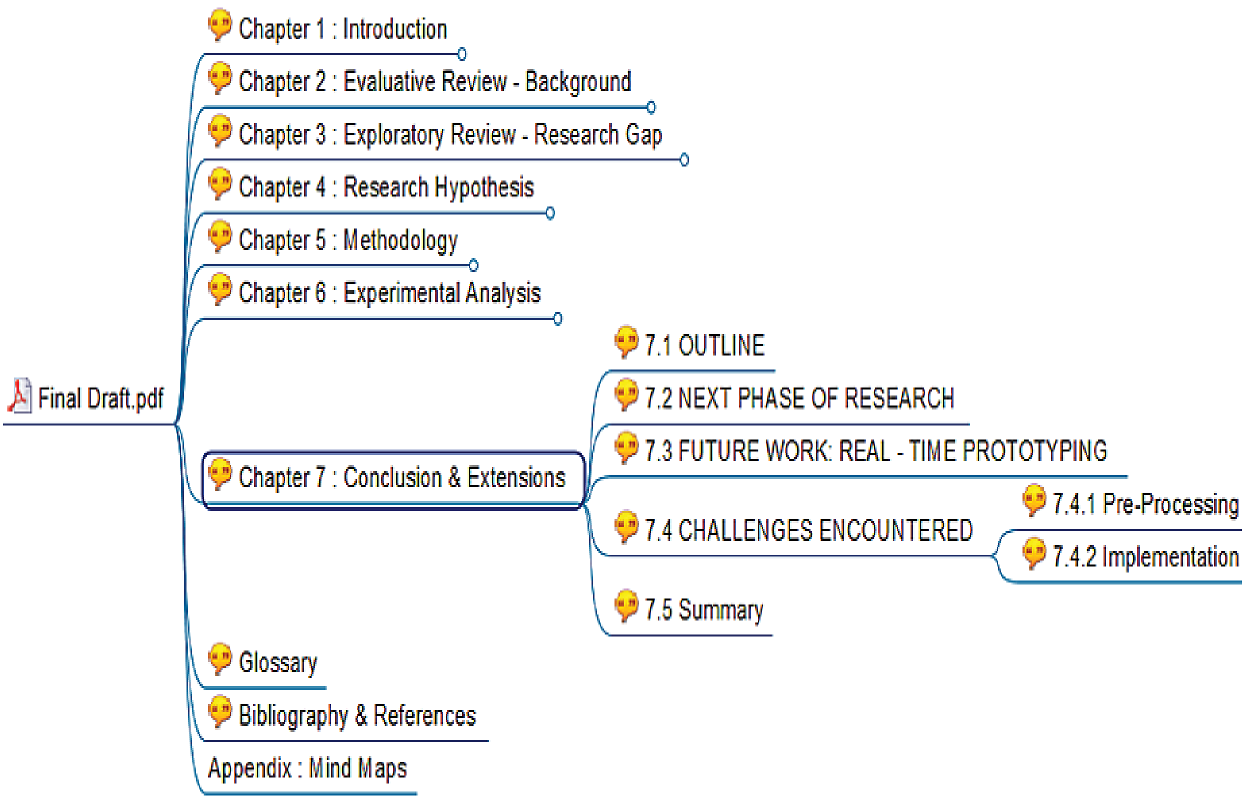


Figure 7.1 Mindmap outline of Chapter 7: Conclusions and Extensions

Glossary

1. System-of-Systems

A collection of discrete, heterogeneous, interdependent systems unified by decentralized homogeneous control, exhibiting hierarchical, heterarchical and composite functional relationships between the different information channels.

2. Cyber-Physical Systems or CPS

Most common installations of system-of-system architecture, developed along the intersection of cyber domain and physical space of data and signal flow.

3. Knowledge Representation metrics

Knowledge Representation metrics for any system or architecture encapsulates every possible intelligence or valid intel of that particular entity; they are data, communication and control signals.

4. Behavioral Patterns

Representation of information about more than one variables along different modes of a system, by applying a qualitative imaging transformation, like a tensor decomposition rule.

5. Autonomic Intelligence

It is the “free will of actions” or spontaneous-governance, providing great chances for the human system administrators and technicians to focus on higher technical issues without intervention.

6. Autonomic Management

The term refers to goal-directed reflexes, that is adaptive self-management and automated self-* characteristics.

7. Multivariate Dynamic Pattern processing

Extracting qualitatively defined patterns from data streams with each value a function of space-time and multiple-indices, visualized as a 3^{rd} order or higher structure.

8. Feature set dictionary

A set of relevant pattern records that can characterize all possible operational states of the data source system(s).

9. Dynamic Feature Patterns

Dynamically varying Feature patterns generated using a core information content tensor; successive outputs maybe compared to identify variation and anomaly patterns.

10. Data characterization

Qualitatively defining and showcasing a dataset that provides markers for processing and analyzing it to extract higher order features.

11. Higher Order Big Data

Big Data sets where, each value is a function of more than 2 variables, and can be visualized as a 3^{rd} order or higher tensor structure.

12. Heterogeneous Data Streams

Data streams that are sourced by heterogeneous components (of dissimilar or non-parallel operating characteristics) of a CPS.

13. Inter-modal dependencies

Dependencies or correlated / non-correlated behavior across continuous data streams originating along different modes of s system (e.g. across space and time).

14. Data crunching

Data Crunching in information science refers to preparation and automated processing of large volumes of information.

15. Data Tensor

Heterogeneous Data streams packed along a Tensor modes, further subjected to decomposition or factorization.

16. Annotated Latent feature set

The Pattern catalogue dynamically generated through Latent feature extraction using Tensor decompositions, to represent time-variant multimodal behavior of a system

17. Mind maps

Diagrammatic representation of information based on a central idea at the root, and associated ideas arranged around it, as derivatives.

18. Knowledge Discovery

Extracting or identifying relevant information attributes and elements from a particular stream of data, signal or control instructions, which will contribute to characterization of the overall Big Data and the source system.

19. Knowledge Representation

Systematic characterization of knowledge attainable from an entity, in terms of data or metadata, using qualitative or quantitative methods of description and depiction.

20. Multivariate Data structures

Flows of Data where each data point is a function of more than one variable, along multiple modes, normally of high dimensions.

21. Proof of concept model

A system model which demonstrates the feasibility of a concept, sometimes supported preliminary experimentation.

22. Computational Architecture

The organizational structure and flow of information or communication within a system defined in terms of computations or processes occurring.

23. Multimodal information

Information comprising of knowledge that span across various modes or indices of the system structure, like space, time, and component ID.

24. Contextual Awareness

The approach of linking changes in the computational ecosystem with the flow of information, to abstract intelligence routines within the system.

25. Internet-of-things

The complex cyber-physical setup of interconnected instruments, devices, equipment and machines and communicating autonomously via the internet and generating big data rich in all kinds of knowledge representation, thereby exhibiting smart behavior.

26. Deep learning

Deep learning is a derivative of machine learning that has the ability for unsupervised learning from unstructured or unlabeled data.

27. Metasystem

A metasystem is a system about other systems, such as describing, generalizing, modelling, or analyzing the other system(s).

28. Prosumers

Customers connected to the physical Grid installation, which are capable of both drawing power from the grid during high load conditions, and supply excess renewable power generated by themselves into the grid network to balance load management. They contribute to every smart aspect of the Smart grid ecosystem.

29. Analytics

Analytics is the discovery, interpretation, and communication of meaningful data patterns, information, relying on statistics, programming, operations research and other tools to quantify performance. They aim at describing, predicting, and improving system performance.

30. Subscale system perspective

Subscale perspective of system analysis refers the perspective into within the given scale of operation, which is usually unobservable or hidden from the macro point of view. It is the multimodal micro scale of visualization.

31. Point-of-Interconnection

The plane of homogeneous interconnection of heterogeneous subsystems of a CPS ecosystem with different characteristics and requirements.

32. Middleware

A firmware that bridges an operating system or database, congruent discrete independent component systems and applications, usually through the medium of a network.

33. Data Engineering

Data Engineering is the design and development of architectures for analyzing and processing Big Data with the required goals or result outcomes pre-defined

34. Anomaly or Anomalous event

Any kind of unusual behavior in a systemic environment may be termed Anomalous with respect to the regular behavioral trends.

35. Dependency

Any coherence or trend between 2 or more similar or dissimilar components' operation and / or information streams, which can be qualitatively or quantitatively modelled and studied.

36. Meta-analysis

The procedure that applies statistical analysis on a large volume of quantitative findings, to integrate the observations and enhance understanding.

Bibliography and References

[1] Y. Mordecai and D. Dori, "Minding the Cyber-Physical Gap: Model-Based Analysis and Mitigation of Systemic Perception-Induced Failure," *Sensors,* vol. 17, no. 7, p. 1644, 2017.

[2] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley and Sons, 2009.

[3] S. Ying and J. Sztipanovits, "Foundations for innovation in cyber-physical systems," in *Workshop Report, Energetics Incorporated, Columbia, Maryland, US*, 2013.

[4] Z. Jianchao, S. Boon-Chong, L. Tek-Tjing, and F. Chuan Heng, "Opportunities for Software-Defined Networking in Smart Grid," in *2013 9th International Conference on Information, Communications and Signal Processing*, 2013, pp. 1-5.

[5] B. Zhou *et al.*, "Smart home energy management systems: Concept, configurations, and scheduling strategies," *Renewable and Sustainable Energy Reviews,* vol. 61, pp. 30-40, 8// 2016.

[6] S. D. J. McArthur *et al.*, "The Autonomic Power System - Network operation and control beyond smart grids," in *Innovative Smart Grid Technologies (ISGT Europe), 2012 3rd IEEE PES International Conference and Exhibition on*, 2012, pp. 1-7.

[7] N. V. R. Reddy and P. V. Krishna, "A Novel Context Aware Power Management Model for Smart Grid Communication Networks," *Contemporary Engineering Sciences, Vol. 9, 2016, no. 11, 531 - 538,* 2016. HIKARI Ltd,

[8] C. Alcaraz, L. Cazorla, and G. Fernandez, "Context-Awareness Using Anomaly-Based Detectors for Smart Grid Domains," *Springer International Publishing Switzerland 2015,* vol. J. Lopez et al. (Eds.): CRiSIS 2014, LNCS 8924, pp. 17–34, 2015., 2015.

[9] M. Donohoe, B. Jennings, and S. Balasubramaniam, "Context-awareness and the smart grid: Requirements and challenges," *Computer Networks,* vol. 79, pp. 263-282, 3/14/ 2015.

[10] Y. E. Song, Y. Liu, S. Fang, and S. Zhang, "Research on Applications of the Internet of Things in the Smart Grid," in *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2015 7th International Conference on*, 2015, vol. 2, pp. 178-181.

[11] M. Moness and A. M. Moustafa, "A Survey of Cyber-Physical Advances and Challenges of Wind Energy Conversion Systems: Prospects for Internet of Energy," *IEEE Internet of Things Journal,* vol. 3, no. 2, pp. 134-145, 2016.

[12] S. Agnew and P. Dargusch, "Effect of residential solar and storage on centralized electricity supply systems," *Nature Clim. Change,* Perspective vol. 5, no. 4, pp. 315-318, 04//print 2015.

[13] Z. Jinquan, Z. Zefeng, Y. Yiping, L. Changnian, and W. Wenhui, "Optimal charging/discharging scheme of battery storage systems in active distribution network," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016, pp. 1-5.

[14] A. M. A. Haidar, K. Muttaqi, and D. Sutanto, "Smart Grid and its future perspectives in Australia," *Renewable and Sustainable Energy Reviews,* vol. 51, pp. 1375-1389, 11// 2015.

[15] G. M. Shafiullah, A. M. T. Oo, A. B. M. S. Ali, and P. Wolfs, "Smart Grid for a Sustainable Future," (in English), *Smart Grid and Renewable Energy,* vol. 4, no. 1, p. 23, Feb 2013, 2013-06-14 2013.

[16] (2012, August 2012). *GREEN ECONOMY IN ACTION - UNDP Report*.

[17] P. Effendi and J. Courvisanos, "Political aspects of innovation: Examining renewable energy in Australia," *Renewable Energy,* vol. 38, no. 1, pp. 245-252, 2// 2012.

[18] International Energy Agency, "Tracking Clean Energy Progress 2016," in "Energy Technology Perspective 2016 " International Energy Agency2016, Available: http://www.iea.org/publications/freepublications/publication/TrackingCleanEnergyProgress2016.pdf.

[19] M. Erol-Kantarci and H. T. Mouftah, "Energy-Efficient Information and Communication Infrastructures in the Smart Grid: A Survey on Interactions and Open Issues," *Communications Surveys and Tutorials, IEEE,* vol. 17, no. 1, pp. 179-197, 2015.

[20] A. Usman and S. H. Shami, "Evolution of Communication Technologies for Smart Grid applications," *Renewable and Sustainable Energy Reviews,* vol. 19, pp. 191-199, 3// 2013.

[21] K. S. Reddy, M. Kumar, T. K. Mallick, H. Sharon, and S. Lokeswaran, "A review of Integration, Control, Communication and Metering (ICCM) of renewable energy based smart grid," *Renewable and Sustainable Energy Reviews,* vol. 38, pp. 180-192, 10// 2014.

[22] H. Farhangi, "The path of the smart grid," *Power and Energy Magazine, IEEE,* vol. 8, no. 1, pp. 18-28, 2010.

[23] A. M. Grilo, J. Chen, M. Díaz, D. Garrido, and A. Casaca, "An Integrated WSAN and SCADA System for Monitoring a Critical Infrastructure," *IEEE Transactions on Industrial Informatics,* vol. 10, no. 3, pp. 1755-1764, 2014.

[24] S. Bahrami and A. Sheikhi, "From Demand Response in Smart Grid Toward Integrated Demand Response in Smart Energy Hub," *Smart Grid, IEEE Transactions on,* vol. 7, no. 2, pp. 650-658, 2016.

[25] R. C. Qiu *et al.*, "Cognitive Radio Network for the Smart Grid: Experimental System Architecture, Control Algorithms, Security, and Microgrid Testbed," *IEEE Transactions on Smart Grid,* vol. 2, no. 4, pp. 724-740, 2011.

[26] W. Kohn, Z. B. Zabinsky, and A. Nerode, "A Micro-Grid Distributed Intelligent Control and Management System," *Smart Grid, IEEE Transactions on,* vol. 6, no. 6, pp. 2964-2974, 2015.

[27] G. K. Venayagamoorthy, "Innovative smart grid control technologies," in *2011 IEEE Power and Energy Society General Meeting*, 2011, pp. 1-5.

[28] M. Beaudin and H. Zareipour, "Home energy management systems: A review of modelling and complexity," *Renewable and Sustainable Energy Reviews,* vol. 45, pp. 318-335, 5// 2015.

[29] G. Venayagamoorthy, "Dynamic, Stochastic, Computational, and Scalable Technologies for Smart Grids," *IEEE Computational Intelligence Magazine,* vol. 6, no. 3, pp. 22-35, 2011.

[30] J. Lazar and M. McKenzie, "Australian Standards for Smart Grids – Standards Roadmap," in "Standards Australia 2012," Department of Resources, Energy and TourismJune 2012 2012, vol. Version 1.0 Available: http://www.standards.org.au/Documents/120904%20Smart%20Grids%20Standards%20Road%20Map%20Report.pdf.

[31] J. Kim, F. Filali, and Y.-B. Ko, "Trends and Potentials of the Smart Grid Infrastructure: From ICT Sub-System to SDN-Enabled Smart Grid Architecture," *Applied Sciences,* vol. 5, no. 4, p. 706, 2015.

[32] D. Niyato, L. Xiao, and P. Wang, "Machine-to-machine communications for home energy management system in smart grid," *IEEE Communications Magazine,* vol. 49, no. 4, pp. 53-59, 2011.

[33] N. Goel and M. Agarwal, "Smart grid networks: A state of the art review," in *2015 International Conference on Signal Processing and Communication (ICSC)*, 2015, pp. 122-126.

[34] S. Luthra, S. Kumar, R. Kharb, M. F. Ansari, and S. L. Shimmi, "Adoption of smart grid technologies: An analysis of interactions among barriers," *Renewable and Sustainable Energy Reviews,* vol. 33, pp. 554-565, 5// 2014.

[35] A. R. S. Vidal, L. A. A. Jacobs, and L. S. Batista, "An evolutionary approach for the demand side management optimization in smart grid," in *Computational Intelligence Applications in Smart Grid (CIASG), 2014 IEEE Symposium on*, 2014, pp. 1-7.

[36] S. Mišák, J. Stuchlý, J. Platoš, and P. Krömer, "A heuristic approach to Active Demand Side Management in Off-Grid systems operated in a Smart-Grid environment," *Energy and Buildings,* vol. 96, pp. 272-284, 6/1/ 2015.

[37] M. Albano, L. L. Ferreira, and L. M. Pinho, "Convergence of Smart Grid ICT Architectures for the Last Mile," *IEEE Transactions on Industrial Informatics,* vol. 11, no. 1, pp. 187-197, 2015.

[38] J. Giraldo, E. Mojica-Nava, and N. Quijano, "Synchronization of isolated microgrids with a communication infrastructure using energy storage systems," *International Journal of Electrical Power and Energy Systems,* vol. 63, pp. 71-82, 12// 2014.

[39] A. R. Khan, A. Mahmood, A. Safdar, Z. A. Khan, and N. A. Khan, "Load forecasting, dynamic pricing and DSM in smart grid: A review," *Renewable and Sustainable Energy Reviews,* vol. 54, pp. 1311-1322, 2// 2016.

[40] S. Aman, Y. Simmhan, and V. K. Prasanna, "Energy management systems: state of the art and emerging trends," *IEEE Communications Magazine,* vol. 51, no. 1, pp. 114-119, 2013.

[41] J. Bremer and M. Sonnenschein, "Parallel tempering for constrained many criteria optimization in dynamic virtual power plants," in *Computational Intelligence Applications in Smart Grid (CIASG), 2014 IEEE Symposium on*, 2014, pp. 1-8.

[42] F. Mwasilu, J. J. Justo, E.-K. Kim, T. D. Do, and J.-W. Jung, "Electric vehicles and smart grid interaction: A review on vehicle to grid and renewable energy sources integration," *Renewable and Sustainable Energy Reviews,* vol. 34, pp. 501-516, 6// 2014.

[43] G. K. Venayagamoorthy, "Intelligent sense-making for smart grid stability," in *Power and Energy Society General Meeting, 2011 IEEE*, 2011, pp. 1-3.

[44] M. Honarmand, A. Zakariazadeh, and S. Jadid, "Integrated scheduling of renewable generation and electric vehicles parking lot in a smart microgrid," *Energy Conversion and Management,* vol. 86, pp. 745-755, 10// 2014.

[45] S. Grijalva and M. U. Tariq, "Prosumer-based smart grid architecture enables a flat, sustainable electricity industry," in *Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES*, 2011, pp. 1-6.

[46] N. Phuangpornpitak and S. Tia, "Opportunities and Challenges of Integrating Renewable Energy in Smart Grid System," *Energy Procedia,* vol. 34, pp. 282-290, 2013/01/01 2013.

[47] O. Vermesan, L.-C. Blystad, P. Hank, R. Bahr, R. John, and A. Moscatelli, "Smart, connected and mobile: Architecting future electric mobility ecosystems," in *Design, Automation and Test in Europe Conference and Exhibition (DATE), 2013*, 2013, pp. 1740-1744.

[48] P. Valerio, "Is the IoT a Tech Bubble for Cities?: With more cities joining the smart city revolution and investing in sensors and other IoT devices, the risk of a new tech bubble is rising," *IEEE Consumer Electronics Magazine,* vol. 5, no. 1, pp. 61-62, 2016.

[49] M. Zetlin, "How the Internet of Things will transform your workplace? ," Accessed on: 06 Oct 2015, Available: https://enterprisersproject.com/article/2015/10/how-internet-things-will-transform-your-workplace

[50] B. Gao *et al.*, "A New Smart Grid Control and Operation Concept - Autonomic Power System," *MATEC Web of Conferences,* vol. 40, p. 07011, 2016.

[51] M. Randles, D. Lamb, and A. Taleb-Bendiab, "Engineering Autonomic Systems Self-Organisation," in *Fifth IEEE Workshop on Engineering of Autonomic and Autonomous Systems (ease 2008)*, 2008, pp. 107-118.

[52] A. Cichocki, R. Zdunek, A. H. Phan, and S.-I. Amari, "Introduction – Problem Statements and Models," in *Nonnegative Matrix and Tensor Factorizations*: John Wiley and Sons, Ltd, 2009, pp. 1-80.

[53] L. Suganthi, S. Iniyan, and A. A. Samuel, "Applications of fuzzy logic in renewable energy systems – A review," *Renewable and Sustainable Energy Reviews,* vol. 48, pp. 585-607, 8// 2015.

[54] P. Siano, "Demand response and smart grids—A survey," *Renewable and Sustainable Energy Reviews,* vol. 30, pp. 461-478, 2// 2014.

[55] A. Windham and S. Treado, "A review of multi-agent systems concepts and research related to building HVAC control," *Science and Technology for the Built Environment,* vol. 22, no. 1, pp. 50-66, 2016/01/02 2016.

[56] S. M. Ali *et al.*, "Wide area smart grid architectural model and control: A survey," *Renewable and Sustainable Energy Reviews,* vol. 64, pp. 311-328, 10// 2016.

[57] V. C. Gungor *et al.*, "Smart Grid Technologies: Communication Technologies and Standards," *IEEE Transactions on Industrial Informatics,* vol. 7, no. 4, pp. 529-539, 2011.

[58] J. Bhatt, V. Shah, and O. Jani, "An instrumentation engineer's review on smart grid: Critical applications and parameters," *Renewable and Sustainable Energy Reviews,* vol. 40, pp. 1217-1239, 12// 2014.

[59] T. Strasser *et al.*, "A Review of Architectures and Concepts for Intelligence in Future Electric Energy Systems," *IEEE Transactions on Industrial Electronics,* vol. 62, no. 4, pp. 2424-2438, 2015.

[60] R. N. Anderson, A. Boulanger, W. B. Powell, and W. Scott, "Adaptive Stochastic Control for the Smart Grid," *Proceedings of the IEEE,* vol. 99, no. 6, pp. 1098-1115, 2011.

[61] X. Zhang, A. J. Flueck, and C. P. Nguyen, "Agent-Based Distributed Volt/Var Control With Distributed Power Flow Solver in Smart Grid," *Smart Grid, IEEE Transactions on,* vol. 7, no. 2, pp. 600-607, 2016.

[62] L. Yu, H. Li, X. Feng, and J. Duan, "Nonintrusive appliance load monitoring for smart homes: recent advances and future issues," *IEEE Instrumentation and Measurement Magazine,* vol. 19, no. 3, pp. 56-62, 2016.

[63] H. Xiao, A. Luo, Z. Shuai, G. Jin, and Y. Huang, "An Improved Control Method for Multiple Bidirectional Power Converters in Hybrid AC/DC Microgrid," *IEEE Transactions on Smart Grid,* vol. 7, no. 1, pp. 340-347, 2016.

[64] Z. Wang, B. Chen, J. Wang, and C. Chen, "Networked Microgrids for Self-Healing Power Systems," *IEEE Transactions on Smart Grid,* vol. 7, no. 1, pp. 310-319, 2016.

[65] Y. Wang, X. Lin, and M. Pedram, "A Near-Optimal Model-Based Control Algorithm for Households Equipped With Residential Photovoltaic Power Generation and Energy Storage Systems," *IEEE Transactions on Sustainable Energy,* vol. 7, no. 1, pp. 77-86, 2016.

[66] G. K. Venayagamoorthy, R. K. Sharma, P. K. Gautam, and A. Ahmadi, "Dynamic Energy Management System for a Smart Microgrid," *IEEE Transactions on Neural Networks and Learning Systems,* vol. PP, no. 99, pp. 1-14, 2016.

[67] D. Tang and P. Wang, "Probabilistic Modeling of Nodal Charging Demand Based on Spatial-Temporal Dynamics of Moving Electric Vehicles," *Smart Grid, IEEE Transactions on,* vol. 7, no. 2, pp. 627-636, 2016.

[68] W. Sheng, K.-y. Liu, S. Cheng, X. Meng, and W. Dai, "A Trust Region SQP Method for Coordinated Voltage Control in Smart Distribution Grid," *IEEE Transactions on Smart Grid,* vol. 7, no. 1, pp. 381-391, 2016.

[69] D. Ranamuka, A. P. Agalgaonkar, and K. M. Muttaqi, "Online Coordinated Voltage Control in Distribution Systems Subjected to Structural Changes and DG Availability," *Smart Grid, IEEE Transactions on,* vol. 7, no. 2, pp. 580-591, 2016.

[70] I. P. Panapakidis, "Application of hybrid computational intelligence models in short-term bus load forecasting," *Expert Systems with Applications,* vol. 54, pp. 105-120, 7/15/ 2016.

[71] M. H. Moradi, M. Abedini, and S. M. Hosseinian, "A Combination of Evolutionary Algorithm and Game Theory for Optimal Location and Operation of DG from DG Owner Standpoints," *Smart Grid, IEEE Transactions on,* vol. 7, no. 2, pp. 608-616, 2016.

[72] A. Botta, W. de Donato, V. Persico, and A. Pescapé, "Integration of Cloud computing and Internet of Things: A survey," *Future Generation Computer Systems*.

[73] "Seven astounding technology trends for 2016," (in English), *SecurityInfoWatch.com,* 2015 Dec 22, 2015-12-23 2015.

[74] M. Fazio, A. Celesti, A. Puliafito, and M. Villari, "Big Data Storage in the Cloud for Smart Environment Monitoring," *Procedia Computer Science,* vol. 52, pp. 500-506, // 2015.

[75] F. Al-Doghman, Z. Chaczko, A. R. Ajayan, and R. Klempous, "A review on Fog Computing technology," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 001525-001530.

[76] T. W. Liao, "Clustering of time series data—a survey," *Pattern recognition,* vol. 38, no. 11, pp. 1857-1874, 2005.

[77] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: applications to image and text data," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 2001, pp. 245-250: ACM.

[78] Y. Wenlu, A. D. Silva, and M. L. Picard, "Computing data quality indicators on Big Data streams using a CEP," in *Computational Intelligence for Multimedia Understanding (IWCIM), 2015 International Workshop on*, 2015, pp. 1-5.

[79] S. Yun, J. H. Park, and W. T. Kim, "Data-centric middleware based digital twin platform for dependable cyber-physical systems," in *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*, 2017, pp. 922-926.

[80] D. Cartes, J. H. Chow, D. McCaugherty, S. Widergren, and G. K. Venayagamoorthy, "The IEEE Computer Society Smart Grid Vision Project opens opportunites for computational intelligence," in *Evolving and Adaptive Intelligent Systems (EAIS), 2013 IEEE Conference on*, 2013, pp. 144-150.

[81] C. Alcaraz, L. Cazorla, and G. Fernandez, "Context-Awareness Using Anomaly-Based Detectors for Smart Grid Domains," in *Risks and Security of Internet and Systems: 9th International Conference, CRiSIS 2014, Trento, Italy, August 27-29, 2014, Revised Selected Papers*, J. Lopez, I. Ray, and B. Crispo, Eds. Cham: Springer International Publishing, 2015, pp. 17-34.

[82] CEN, CENELEC, and ETSI, "Smart Grid Reference Architecture," Smart Grid Coordination Group (SGCG), November 2012, 2012, Available: https://ec.europa.eu/energy/sites/ener/files/documents/xpert_group1_reference_architecture.pdf.

[83] B. Meehan, "Enterprise GIS and the Smart Electric Grid." doi: G33457 ESRI10/08dh Available: http://www.esri.com/library/articles/enterprise-seg.pdf

[84] F. Shahnia, "Analysis and correction of voltage profile in low voltage distribution networks containing photovoltaic cells and electric vehicles," PhD, 2011.

[85] K. Sharma and L. Mohan Saini, "Performance analysis of smart metering for smart grid: An overview," *Renewable and Sustainable Energy Reviews,* vol. 49, pp. 720-735, 9// 2015.

[86] V. Namboodiri, V. Aravinthan, and W. Jewell, "Communication Needs and Integration Options for AMI in the Smart Grid," in "“The Future Grid to Enable Sustainable Energy Systems”," May 2012 2012.

[87] "Communications network solutions for smart grids," Siemens AG2011, Available: https://w5.siemens.com/belux/web/fr/infrastructure-and-cities/smart-grid/smart-

communications/Documents/Communication%20Network%20Solutions%20for%20Smart%20Grids_110638_WS_CommSolutions_Dach_E.pdf.

[88] B. Speer and M. Miller, "The Role of Smart Grids in Integrating Renewable Energy," in "ISGAN Synthesis Report," National Renewable Energy Laboratory,NREL/TP-6A20-63919, May 2015 2015, Available: http://www.nrel.gov/docs/fy15osti/63919.pdf.

[89] R. Kempener, P. Komor, and A. Hoke, "Smart Grids and Renewables: A Guide for Effective Deployment," International Renewable Energy Agency (IRENA), www.irena.orgNovember 2013 2013.

[90] J. Lin, R. Ravichandiran, H. Bannazadeh, and A. Leon-Garcia, "Monitoring and measurement in software-defined infrastructure," in *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 2015, pp. 742-745.

[91] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognition Letters,* vol. 42, pp. 11-24, 2014.

[92] I. Jayawardene and G. K. Venayagamoorthy, "Comparison of echo state network and extreme learning machine for PV power prediction," in *Computational Intelligence Applications in Smart Grid (CIASG), 2014 IEEE Symposium on*, 2014, pp. 1-8.

[93] S. A. Mehdi, J. Khalid, and S. A. Khayam, "Revisiting Traffic Anomaly Detection Using Software Defined Networking," in *Recent Advances in Intrusion Detection: 14th International Symposium, RAID 2011, Menlo Park, CA, USA, September 20-21, 2011. Proceedings*, R. Sommer, D. Balzarotti, and G. Maier, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 161-180.

[94] S. Wang and X. Huang, "Aggregation points planning for software-defined network based smart grid communications," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, 2016, pp. 1-9.

[95] D. Alahakoon and X. Yu, "Smart Electricity Meter Data Intelligence for Future Energy Systems: A Survey," *IEEE Transactions on Industrial Informatics,* vol. 12, no. 1, pp. 425-436, 2016.

[96] M. Smith. (2015). *THE TRUTH ABOUT INTELLIGENT AUTOMATION WITH HFS RESEARCH*. Available: http://digitally.cognizant.com/digital-technologies/talking-raw-truth-intelligent-automation-hfs-research/

[97] F. Shrouf and G. Miragliotta, "Energy management based on Internet of Things: practices and framework for adoption in production management," *Journal of Cleaner Production,* vol. 100, pp. 235-246, 8/1/2015 2015.

[98] D. K. Mohanta, C. Murthy, and D. Sinha Roy, "A Brief Review of Phasor Measurement Units as Sensors for Smart Grid," *Electric Power Components and Systems,* vol. 44, no. 4, pp. 411-425, 2016/02/25 2016.

[99] E. Acar, A. Anandkumar, L. Mullin, S. Rusitschka, and V. Tresp, "Tensor Computing for Internet of Things," in "Report from Dagstuhl Perspectives Workshop 16152," Dagstuhl Reports: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany, GermanyApril 10–13, 2016 2016, vol. Vol. 6, Issue 4, pp. 57–79 Available: http://www.dagstuhl.de/16152.

[100] Quim Llimona Torras, "Multimodal Data: Acquisition, Processing, Storage and Exploration," in *lemonzi.files.wordpress.com*, ed. Wordpress.com, , 2013.

[101] N. Vervliet, O. Debals, L. Sorber, and L. D. Lathauwer, "Breaking the Curse of Dimensionality Using Decompositions of Incomplete Tensors: Tensor-based scientific computing in big data analysis," *IEEE Signal Processing Magazine,* vol. 31, no. 5, pp. 71-79, 2014.

[102] N. Zheng, Q. Li, S. Liao, and L. Zhang, "Flickr group recommendation based on tensor decomposition," in *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, 2010, pp. 737-738: ACM.

[103] P. R. Patil and M. Bhamare, "Multidimensional Data Mining for Anomaly Extraction," in *2013 Third International Conference on Advances in Computing and Communications*, 2013, pp. 5-8.

[104] A. Cichocki, "Era of Big Data Processing: A New Approach via Tensor Networks and Tensor Decompositions," in *International Workshop on Smart Info-Media Systems in Asia,*, 2014.

[105] J. Liang, G. Yu, B. Chen, and M. Zhao, "Decentralized Dimensionality Reduction for Distributed Tensor Data Across Sensor Networks," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 27, no. 11, pp. 2174-2186, 2016.

[106] F. Angiulli and F. Fassetti, "Distance-based outlier queries in data streams: the novel task and algorithms," *Data Mining and Knowledge Discovery,* vol. 20, no. 2, pp. 290-324, 2010/03/01 2010.

[107] D. C. Alexander, J. C. Gee, and R. Bajcsy, "Similarity Measures for Matching Diffusion Tensor Images," in *BMVC*, 1999, pp. 1-10.

[108] M. Muja and D. G. Lowe, "Fast Matching of Binary Features," in *2012 Ninth Conference on Computer and Robot Vision*, 2012, pp. 404-410.

[109] IDRE. (accessed August 22, 2016). *What Is The Difference Between Categorical, Ordinal And Interval Variables?* Available: https://stats.idre.ucla.edu/other/mult-pkg/whatstat/what-is-the-difference-between-categorical-ordinal-and-interval-variables/

[110] A. Sanfeliu and M. L. Cortés, "Progress in pattern recognition, image analysis and applications," in *9th Iberoamerican Congress on Pattern Recognition, CIARP*, 2004: Springer.

[111] T. Jeong, J. G. Park, and C. Lee, "Feature Extraction Based on Difference Vectors," in *2007 2nd International Workshop on Soft Computing Applications*, 2007, pp. 183-186.

[112] X. Zhang, G. Wen, and W. Dai, "A tensor decomposition-based anomaly detection algorithm for hyperspectral image," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 54, no. 10, pp. 5801-5820, 2016.

[113] T. Penkova, "Principal component analysis and cluster analysis for evaluating the natural and anthropogenic territory safety," *Procedia Computer Science,* vol. 112, pp. 99-108, 2017.

[114] B. Boashash, *Time-frequency signal analysis and processing: a comprehensive reference*. Academic Press, 2015.

[115] J. Murakami, N. Yamamoto, and Y. Tadokoro, "High-speed computation of 3D tensor product expansion by the power method," *Electronics and Communications in Japan (Part III: Fundamental Electronic Science),* vol. 85, no. 7, pp. 63-72, 2002.

[116] G. Chen, F. Wang, and C. Zhang, "Collaborative filtering using orthogonal nonnegative matrix tri-factorization," *Information Processing and Management,* vol. 45, no. 3, pp. 368-379, 2009.

[117] L. Kuang, F. Hao, L. T. Yang, M. Lin, C. Luo, and G. Min, "A Tensor-Based Approach for Big Data Representation and Dimensionality Reduction," *IEEE Transactions on Emerging Topics in Computing,* vol. 2, no. 3, pp. 280-291, 2014.

[118] T. G. Kolda, "Tensor Analysis for Evolving Networks," in "Workshop on Time-varying Complex Network Analysis," Office of Advanced Scientific Computing Research, U.S. Department of Energy, , Cambridge, UK, , Workshop repot, September 19, 2012 2012, , Available: https://www.cl.cam.ac.uk/research/srg/netos/events/temporalnets-workshop12/Kolda.pdf.

[119] A. Cichocki, R. Zdunek, A. H. Phan, and S.-I. Amari, "Multi-Way Array (Tensor) Factorizations and Decompositions," in *Nonnegative Matrix and Tensor Factorizations*: John Wiley and Sons, Ltd, 2009, pp. 337-432.

[120] I. Perros, R. Chen, R. Vuduc, and J. Sun, "Sparse Hierarchical Tucker Factorization and Its Application to Healthcare," in *2015 IEEE International Conference on Data Mining*, 2015, pp. 943-948.

[121] S. Sizov, S. Staab, and T. Franz, "Analysis of Social Networks by Tensor Decomposition," in *Handbook of Social Network Technologies and Applications*, B. Furht, Ed. Boston, MA: Springer US, 2010, pp. 45-58.

[122] L. Kuang, L. T. Yang, and K. Qiu, "Tensor-based software-defined internet of things," *IEEE Wireless Communications,* vol. 23, no. 5, pp. 84-89, 2016.

[123] W. Hu, J. Gao, J. Xing, C. Zhang, and S. Maybank, "Semi-Supervised Tensor-Based Graph Embedding Learning and Its Application to Visual Discriminant Tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, no. 1, pp. 172-188, 2017.

[124] G. R. Pinheiro, G. S. Soares, A. L. Costa, R. A. Lotufo, and L. Rittner, "Divergence map from diffusion tensor imaging: Concepts and application to corpus callosum," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 1120-1123.

[125] W. Luan, G. Liu, and C. Jiang, "Collaborative tensor factorization and its application in POI recommendation," in *2016 IEEE 13th International Conference on Networking, Sensing, and Control (ICNSC)*, 2016, pp. 1-6.

[126] X. Y. Liu, S. Aeron, V. Aggarwal, X. Wang, and M. Y. Wu, "Tensor completion via adaptive sampling of tensor fibers: Application to efficient indoor RF fingerprinting," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 2529-2533.

[127] K. M. Lee, M. Li, and C. Y. Lin, "Magnetic Tensor Sensor and Way-Finding Method Based on Geomagnetic Field Effects With Applications for Visually Impaired Users," *IEEE/ASME Transactions on Mechatronics,* vol. 21, no. 6, pp. 2694-2704, 2016.

[128] P. Koniusz and A. Cherian, "Sparse Coding for Third-Order Super-Symmetric Tensor Descriptors with Application to Texture Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5395-5403.

[129] G. Cui, L. Gui, Q. Zhao, A. Cichocki, and J. Cao, "Bayesian CP factorization of incomplete tensor for EEG signal application," in *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2016, pp. 2170-2173.

[130] X. Cheng, Y. Qiu, W. Hou, and W. K. Ching, "A semi-tensor product approach for Probabilistic Boolean Networks," in *2014 8th International Conference on Systems Biology (ISB)*, 2014, pp. 85-90.

[131] H. Kasai, W. Kellerer, and M. Kleinsteuber, "Network Volume Anomaly Detection and Identification in Large-Scale Networks Based on Online Time-Structured Traffic Tensor Tracking," *IEEE Transactions on Network and Service Management,* vol. 13, no. 3, pp. 636-650, 2016.

[132] H. Jingfei, S. Guiling, Z. Ying, and G. Tianyu, "Data recovery in heterogeneous wireless sensor networks based on low-rank tensors," in *2016 IEEE Symposium on Computers and Communication (ISCC)*, 2016, pp. 616-620.

[133] L. Kuang, L. T. Yang, X. Wang, P. Wang, and Y. Zhao, "A tensor-based big data model for QoS improvement in software defined networks," *IEEE Network,* vol. 30, no. 1, pp. 30-35, 2016.

[134] A. Cichocki, R. Zdunek, A. H. Phan, and S.-I. Amari, "Selected Applications," in *Nonnegative Matrix and Tensor Factorizations*: John Wiley and Sons, Ltd, 2009, pp. 433-472.

[135] F. Huang, "Discovery of latent factors in high-dimensional data using tensor methods," *arXiv preprint arXiv:1606.03212,* 2016.

[136] J. C. Ho *et al.*, "Limestone: High-throughput candidate phenotype generation via tensor factorization," *Journal of Biomedical Informatics,* vol. 52, pp. 199-211, 12// 2014.

[137] J. Dauwels, K. Srinivasan, R. M. Ramasubba, and A. Cichocki, "Multi-channel EEG compression based on matrix and tensor decompositions," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 629-632.

[138] J. Byun, I. Hong, B. Kang, and S. Park, "A smart energy distribution and management system for renewable energy distribution and context-aware services based on user patterns and load forecasting," *IEEE Transactions on Consumer Electronics,* vol. 57, no. 2, pp. 436-444, 2011.

[139] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.,* vol. 41, no. 3, pp. 1-58, 2009.

[140] K. Mets, J. A. Ojea, and C. Develder, "Combining Power and Communication Network Simulation for Cost-Effective Smart Grid Analysis," *IEEE Communications Surveys and Tutorials,* vol. 16, no. 3, pp. 1771-1796, 2014.

[141] E. Tsampasis, L. Sarakis, H. Leligou, T. Zahariadis, and J. Garofalakis, "Novel Simulation Approaches for Smart Grids," *Journal of Sensor and Actuator Networks,* vol. 5, no. 3, p. 11, 2016.

[142] J. McDuling, "Tesla gears up for Australia's electricity market," November 4, 2015 - 10:49AM. Available: http://www.smh.com.au/business/energy/tesla-gears-up-for-australias-electricity-market-20151103-gkq4he.html

[143] (2014). *ARUP SGSC Executive Report:*.

[144] (2014). *SGSC Technical Compendium: Distributed Generation and Distributed Storage*. Available: https://www.google.com.au/search?q=Network+models+for+the+Newcastle%2C+Upper+Gundy+and+Newington+areas+%E2%80%93+to+understand+the+interconnectivity+of+feeders+and+measurement+devices%2Candoq=Network+models+for+the+Newcastle%2C+Upper+Gundy+and+Newington+areas+%E2%80%93+to+understand+the+interconnectivity+of+feeders+and+measurement+devices%2Candaqs=chrome..69i57.4709j0j7andsourceid=chromeandie=UTF-8

[145] Ausgrid, "DGDS3052 – Customer Grade Distributed Energy Power Quality Investigation – A Case Study in a Suburban Environment," 2013, .

[146] Ausgrid, "DGDS3053 – Customer Grade Distributed Energy Power Quality Investigation – A Case Study in an Urban Environment," 2013, .

[147] Ausgrid, "DGDS3051 – Customer Grade Distributed Energy Power Quality Investigation – A Case Study in a Rural Environment," 2013, .

[148] J. Kossaifi. (2016 - 2017). *tensorly.decomposition.tucker*. Available: https://tensorly.github.io/stable/modules/generated/tensorly.decomposition.tucker.html#tensorly.decomposition.tucker

[149] I. Oseledets. *oseledets/ttpy - Python implementation of the TT-Toolbox*. Available: https://github.com/oseledets/ttpy

[150] O. Motlagh, G. Foliente, and G. Grozev, "Knowledge-mining the Australian smart grid smart city data: A statistical-neural approach to demand-response analysis," in *Planning Support Systems and Smart Cities*: Springer, 2015, pp. 189-207.