CITY UNIVERSITY LONDON

Smart Grid Framework Analysis and Artificial Neural Network in Load Forecast

By

Fang Yuan Xu

This thesis is submitted for the Degree of

Doctor of Philosophy

At

City University London

School of Engineering and Mathematical Sciences

September 2011

Acknowledgement

It is my honour to express my respect and gratitude to my supervisor Professor Loi Lei Lai for his professional support and supervision to my PhD study with patient. His novel researching style directs me on forming my own research skills. With the bi-directional communication, he introduced the world-leading technologies and latest development to me. Every time I met difficulties, his encouragement is my incentive towards the challenges. I am highly grateful to his help and kindness.

I would like to express my thankfulness to Dr Ma Yingnan, Dr Zhou Long, Zhang Hao Tian and many other colleagues in Energy System Group, City University London. Researching together for years, communications with them not only give me inspiration but also make my research a pleasure.

I would like to convey my appreciation and gratefulness to my parents, Jian Ping Xu and Jia Ning Dai for their self-forgetful loves and support to me. I miss them every day abroad.

It is my honour to express my thanks to Lai Chi Ken, Zhu Feng, Li Zhi, Hao Zhou and Si Jia Lu. They offer their helps on providing materials and resources when I need.

I appreciate the solicitude from my girl-friend Li Pei Peng. Her patient and love is my promotion to future.

Finally I would like to thank all guides from Kildare Walk 21: Zhi Hao Zhong, Li Yan Zhang, Zhi Hong Zhong, Cheng Cheng Cui, Sheng Li Jiang and Tracy Zhang. They bring love and joy to my study life in the UK.

Bow to ALL ones supporting me

Abstract

Power system is the one of the most critical parts of the whole energy utilization around the world. Recently people pay more attention to the energy utilization, new types of generations, storages and power utilization need to increase energy efficiency and reduce carbon emission. Due to the power grid currently is still mainly under the old-designed approach, it is increasingly exposed limitation on efficiency enhancement, security and reliability improvement, new technologies compatibility and meeting larger power capacity requirements.

Thus, Smart Grid is 'born' to improve power grid for these requirements. It is an overlapping area between power system and digital technology, intelligent technology, communication technology and so on. Smart Grid can provide updates for nearly all sections of traditional power grid. It is a systematic framework that new technologies integration, system development strategy and planning, customers' awareness improvements and supports from all relevant areas. The areas must be operated in coordination and parallel.

Firstly, this thesis introduces Smart Grid and Smart Metering on its definition, characteristics and deployment.

Secondly, this thesis describes a load forecasting system for macro-grid. Artificial Neural Network (ANN) was introduced to achieve this work for its excellent mapping approximation ability.

In the third section, thesis focuses on load forecasting for micro-grid. Back-Propagation method is used to train the Multi-layer Perceptron (MLP) ANN and its results were compared to that from Radial Basis Function (RBF) ANN. Analysis was focused not only on the two networks but also ANN generalization problems and differences between micro-grid load and macro-grid load prediction.

List of Publications

[1].Fang yuan Xu, Loi Lei Lai, '*A Study on Design and Functionalities of Smart Grid*', Power and Energy Society General Meeting, 2011 IEEE, pp. 1-5, Oct 2011, Print ISBN: 978-1-4577-1000-1.

[2]. Fang yuan Xu, Long Zhou, Yi Lin Wu, Yingnan Ma, '*Standards, Policies and Case studies in smart metering*', Power and Energy Society General Meeting, 2010 IEEE, pp. 1-5, Sep 2011, Print ISBN: 978-1-4244-6549-1

[3]. Fang yuan Xu, Long Zhou, Loi Lei Lai, '*Application of Artificial Neural Network in electrical analysis of micro-grid load*', Power and Energy Society General Meeting, 2010 IEEE, pp. 1-5, Sep 2010, Print ISBN: 978-1-4244-8357-0

[4]. Fang yuan Xu, Loi Lei Lai, '*Simple intelligent 3-phase power quality detecting framework for microgrid*', 2011 International Conference on Machine Learning and Cybernetics, Volume 3, pp. 1319 – 1323, Sep 2011, Print ISBN: 978-1-4577-0305-8.

[5]. Fang yuan Xu, Leung, M.C., Long Zhou, '*A RBF network for short – term load forecast on microgrid*', 2010 International Conference on Machine Learning and Cybernetics, Volume 6, pp. 3195 – 3199, Sep 2011, Print ISBN: 978-1-4244-6526-2

[6]. Hao-Tian Zhang, Fang yuan Xu, Long Zhou, '*Artificial neural network for load forecasting in smart grid*', 2010 International Conference on Machine Learning and Cybernetics, Volume 6, pp. 3200 – 3205, Sep 2010, Print ISBN: 978-1-4244-6526-2.

[7]. Long Zhou, Fang yuan Xu, Ying nan Ma, '*Impact of smart metering on energy efficiency*', 2010 International Conference on Machine Learning and Cybernetics, Volume 6, pp. 3213 – 3218, Sep 2010, Print ISBN: 978-1-4244-6526-2.

[8]. Fang yuan Xu, Loi Lei Lai, '*Multi-Agent Demand Side Management Framework Design for Micro - Grid*', IEEE Transactions on Smart Grid. (to be submitted)

[9]. Fang yuan Xu, Loi Lei Lai, '*RBF Logical Input Element Robustness Analysis*', IEEE Transactions on Neural Networks. (to be submitted)

Table of Content

List of Figures

List of Tables

Copyright Declaration

Fang Yuan Xu

London, September 2011

Chapter 1

Introduction

Power system is the one of the most critical parts of the whole energy utilization around the world. Recently people pay more attention to the energy utilization, new types of generations, storages and power utilization need to increase energy efficiency and reduce carbon emission. Due to the power grid currently is still mainly under the old-designed approach, it is increasingly exposed limitation on efficiency enhancement, security and reliability improvement, new technologies compatibility and meeting larger power capacity requirements.

Thus, Smart Grid is 'born' to improve power grid for these requirements. It is an overlapping area between power system and digital technology, intelligent technology, communication technology and so on. Smart Grid can provide updates for nearly all sections of traditional power grid, including renewable energy generation and new storage integration, Demand Response (DR) and Demand Side Management (DSM), Transmission & Distribution Automation, Electric Vehicle (EV) integration, Advanced Metering Infrastructure (AMI) and so on. It is a systematic framework that new technologies integration, system development strategy and planning, customers' awareness improvements and supports from all relevant areas. The areas must be operated in coordination and parallel [1] – [6].

In current status, Smart Grid developments still stay at the initial points. Various works are placed at Smart Grid definition, characteristics summarization, standardization and Smart Grid test bedding. But seldom people have organized the above work in a reasonable development procedure. This thesis aims to establish one systematic procedure framework that formalizes the design of Smart Grid scope. Works will be constituted with plain sequences in this procedure framework.

Power load, as a main requirement for power system, affects the power flow in every electricity cable. A prediction of power load influences not only planning for all stakeholders but also the reliability and security. Smart Grid introduce new feature to power system like micro-grid, which lead to new requirement to microgrid load forecast for Distributed Generation and other micro-grid management. This thesis aims to achieve macro-gird load forecast and micro-grid load forecast by Artificial

Neural Network. Moreover, through analysis and compare, this thesis figures out the feature differences between macro-grid load forecast and micro-grid load forecast, Smart Meter plays an important role in demand response. It is more than a measurer but also a platform for demand response and dynamic pricing. In demand response, customers receive the latest load information from Smart Meter before their consumption. The utilities will provide a predicted price for customer to manage their consumptions, which the predicted price is based on the predicted load. Due to the huge contribution from Smart Meter to load forecast applications, this thesis also concerns about the Smart Metering development.

1.1 Thesis Organization

This thesis mainly focuses on Smart Grid and Power System load forecasting. It consists of 6 Chapters.

Chapter 1 is the main Introduction and describes the layout the whole thesis.

Chapter 2 introduces Smart Grid on its definition, characteristics and constructions. Furthermore, the competitive Smart Grid standardizations is also revealed. Smart Grid demonstration projects worldwide are included to summarize the countries' behaviour toward this new concept.

As the earliest application, Smart Metering system is introduced in Chapter 3. This Chapter provides analysis on policy and standards of Smart Metering worldwide. Case study for each country of their Smart Metering application is included.

Chapter 4 describes a load forecasting system for Macro-grid. Artificial Neural Network (ANN) is introduced to achieve this work for its excellent mapping approximation ability. Back-Propagation training and its improvements are introduced with analysis.

Chapter 5 focuses on load forecasting for Micro-grid. Back-Propagation trained MLP ANN and Radial Basis Function (RBF) ANN are applied for comparison. Analysis is placed not only on the two networks but also ANN generalization problems and differences between Micro-grid load and Macro-grid load prediction.

Chapter 6 summarizes the work done in the study. Based on the current work, direction on future study is pointed out, such as load forecasting system for a mixed load Micro-grid, which integrated with the framework of Demand Side Management (DSM).

1.2 Original Contribution

1. Comparison, analysis and summary of Smart Metering Standards and Policies. This work summarizes the advantages and week points of various Smart Metering projects with their development procedure. It provides a good reference for future Smart Grid development. (Chapter 3)
2. Analysis on Smart Grid design, functionalities and standards. This work organizes Smart Grid's characteristics, functionalities, necessary technologies into a scope design procedure. Comparison is also applied to Smart Grid standards worldwide, which provide a good tutorial for areas aiming to develop Smart Grid system. (Chapter 2)
3. Smart Grid load forecasting system framework design for Macro-grid in Ontario, Canada. This work introduces a Smart Grid load forecast design procedure with considering general influencing factors and Ontario local factors. (Chapter 4)
4. An Artificial Neural Network based load forecasting system design for Ontario, Canada. This work compares results from different ANN training algorithms and provides a novel explanation for the differences. (Chapter 4)
5. Micro - Grid load forecasting system framework design for City University of Hong Kong. This work compares the differences between macro-grid and micro-grid load forecast problems so as to figure out the traditional method has limitation in Micro-grid load forecast (Chapter 5)

6. MLP network based and RBF network based load forecasting system design. This work introduces a real-time hourly load forecasting for City University of Hong Kong by comparing two different ANNs. (Chapter 5)

Chapter 2

Smart Grid

2.1 Introduction

The Smart Grid vision presents a new power system with more automatics, more intelligence, more decentralization, more options and consumer participation, and better resilience and management. It is an upgrade from the traditional power grid in all levels, including not only the technologies and management but also the value and the characteristics.

This Chapter proposed a construction of Smart Grid in multi-level with an entire scope framework design procedure. An orbicular Smart Grid description including characteristics, metrics, standards and technologies will be unfurled.

2.2 Smart Grid Definition

Smart Grid is a large and complicated concept which is still holding debate on its definition because of the expected emphasis addressed by each participant.
Various definition of Smart Grid is raised, such as:

- "The infrastructure to transmit renewably generated electricity from a variety of small and large generation sites scattered over wide areas with the ability to manage both fluctuating supply and loads" by European Climate Forum (ECF) [7].

- "A smart grid is an electricity network that uses digital and other advanced technologies to monitor and manage the transport of electricity from all generation sources to meet the varying electricity demands of end-users" by International Energy Agency [8].

- "The "smart grid" has come to describe a next-generation electrical power system that is typified by the increased use of communications and information technology in the generation, delivery and consumption of electrical energy." by IEEE [9].

Other Smart Grid definitions are shown in Table 2.1:

Definition provider	Smart Grid Definition
EU Functionalities of Smart Grids and smart meters	A Smart Grid is an electricity network that can cost efficiently integrate the behaviour and actions of all users connected to it – generators, consumers and those that do both – in order to ensure economically efficient, sustainable power system with low losses and high levels of quality and security of supply and safety [10].
DOE Smart Grid book: The smart grid: An Introduction	A smarter grid provides chances to make the transformation from a centralized, producer-controlled network to one that is less centralized and more consumer-interactive, by bringing the philosophies, concepts and technologies that enabled the internet to the utility and the electric grid. More importantly, it enables the industry's best ideas for grid modernization to achieve their full potential [6].
European Regulators' Group for Electricity and Gas	A Smart Grid is an electricity network that can cost efficiently integrate the behaviour and actions of all users connected to it – generators, consumers and those that do both – in order to ensure economically efficient, sustainable power system with low losses and high levels of quality and security of supply and safety is [11].

Table 2.1: Smart Grid Definition

But whatever definitions are used, Smart Grid will always be focused when facing the following problems:

- High proportion of high-carbon centralized generation.
- Monotonous market.
- Few options to power quality, services and consumption types.
- Low efficiency in asset management.

- Vulnerable to attack and disasters

Due to face similar problems, most of the Smart Grid definitions will include the following sections:

- To use new technical methods to improve the efficiency, security and power reliability of each part and the whole electricity grid.
- To provide new services, new customer options and to enable the grid compatibility for new products and new services.
- To set up an entire communicating system and the associated assets that improve the interpretability among related devices for better effect.

2.3 Drive of Smart Grid

New thing is splendid as born with promotion and requirement. Smart Grid, a new set of new value, new characteristics and new technologies, is raised as facing several requirements other than from whimsy.

In 21st century, electric systems in several major economic entities are going to suffer serious bottlenecks, which are mainly placed in persisting supplying clean, reliable and affordable energy services. According to the requirement, the general drivers for Smart Grid are [6]:

- **Reliability**: Blackouts and brownouts are happen frequently as lacking of automated analytics, slow response times of mechanical switches and lacking of situational awareness on the part of grid operators.
- **Efficiency**: Based on the large scale of power system, only a small improvement of efficiency representing not only a large consumption reduction but also a significant carbon emitting decrease.
- **National Economy**: In 2005, extreme weather causing extensive damage of overhead lines in Southern Sweden, burning 400 million € with 70 million m^3 wood damaged [12]. The traditional grid not only fails to meet the requirement of economic development but also fails in providing enough protection to current economic status.
- **Affordability**: As the requirement of economy grow and new electricity consumption, traditional grid can no longer afford the development.

- **Security**: The traditional grid's centralized structure leaves all the society into the risk of attack. A too much dependency on grid could bring national banking, communications, and traffic and security systems among others to a complete standstill when attack occurs.
- **Environment/climate change**: Over half of the electricity worldwide is burning by coal, producing pollution and green house gas.
- **Global competitiveness**: Facing the development of Smart Grid, countries and organization worldwide are rapidly raising their solutions to the above problems, trying to capture a leading chance for development and business [6].

The above drives plus several local drives specified for local countries or local areas promote the research and deployment of the Smart Grid.

2.4 Smart Grid Scope Design Procedure.

Different Smart Grid scopes or landscapes are published by various organizations. E.g. [17] introduces scope from Department of Energy (DOE) in US while [18] introduces a scheme for UK Smart Grid development. In the scope reports it is easy to find out description of multiple types of technologies and standards. But it is difficult to find out a systematic reflection from technologies to the national development aims. For other nations or organizations that prefer to form up a scope of their own, it is better to organize all the researches together into a procedure revealing the way to sketch the scope other than just show up what the scope is. This section selects Smart Grid scope from DOE as example to introduce a scope procedure.

Fig 2.1 reveals the procedures. One nation should use their national development object to guide its Smart Grid scope. From the national developments, summarize characteristics on what power system should look like. Then based on these characteristics, finding out what technologies and standards could satisfy them. Finally the scope is formed. National developments relate to power system are list below:

- **Aim 1:** less carbon emit and pollution to protect global environment.
- **Aim 2:** Spend less for every energy unit generation.
- **Aim 3:** Increase energy consumption security and reliability.
- **Aim 4:** Promote economic developments

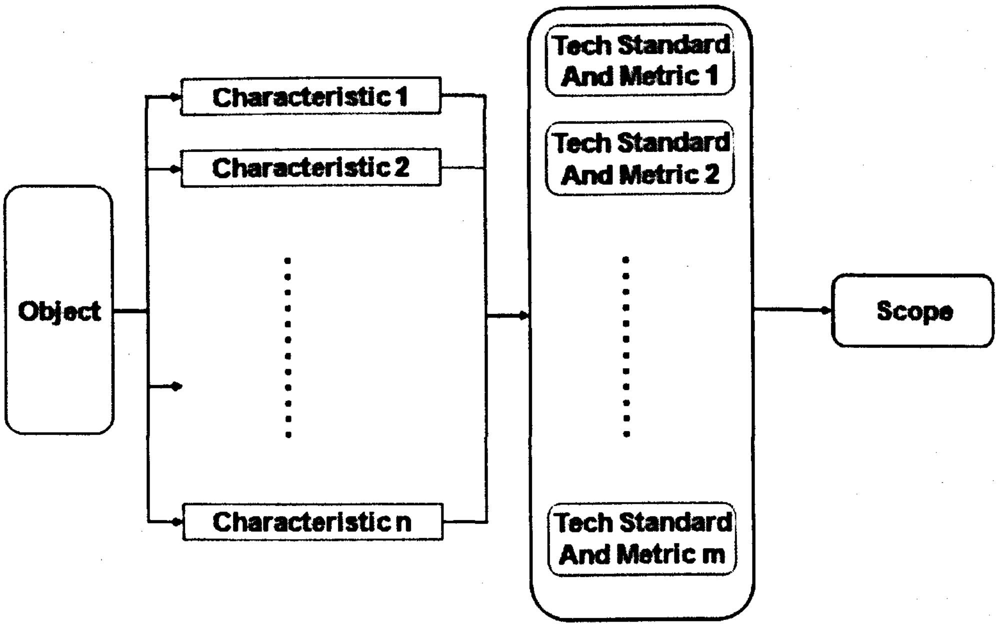


Fig 2.1: Smart Grid Scope Design Procedure [14]

2.4.1 Smart Grid Characteristics

To better understand the objects and classification for new technologies and services in Smart Grid, a classification of characteristics of Smart Grid is a fundamental support. One should summarize out their own Smart Grid Characteristics basing on their development object and the current situation. These characteristics are further description of smart grid definition and they provide guidance for the smart grid new technologies.

Set US as an example. In US, Department of Energy publishes a report in reference [15], describing smart grid in 6 characteristics. The Fig 2.2 reveals these Smart Grid characteristics.

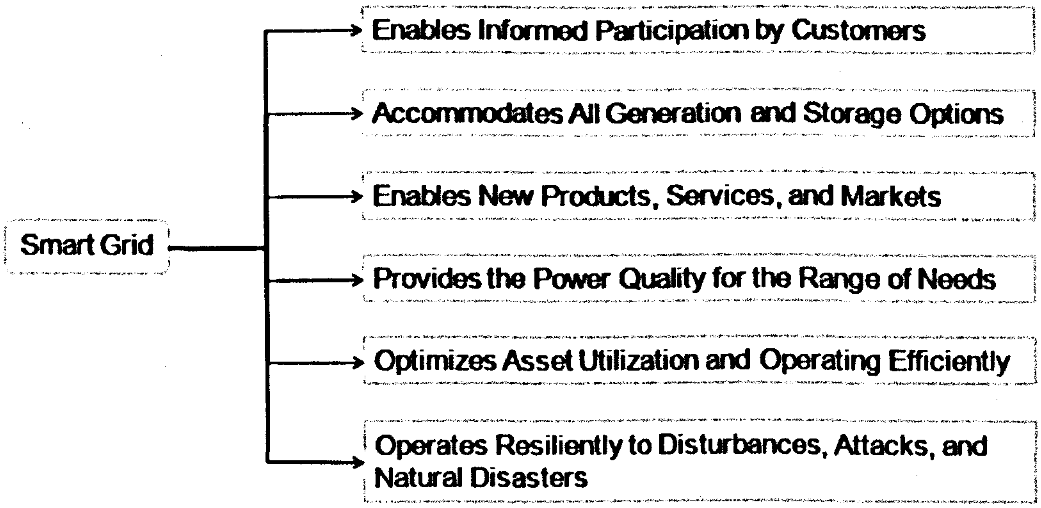


Fig 2.2: Smart Grid Characteristics by NELT [14]

From Fig.2.2, each characteristic are based on their national development aims:

- **Enable Informed Participation by Customers (C 1):** This characteristic introduces bi-directional information flow and energy flow between utilities and consumer. This two-way flow not only saves cost of utilities through appliances like Metering Automation, but also suggests a greener consumption style to customers by dynamic pricing. Therefore, this characteristic reflects incentive from Aim 1 and Aim 2.
- **Accommodate All Generation and Storage Options (C2):** This characteristic introduces new types of generations, like renewable bulk generation and distributed generation, which may reduce carbon emit and the traditional generation cost. Moreover, requirement of new types of generations may create new market and provide new motivation for economic development. So this characteristic reflects incentive from Aim 1, Aim 2 and Aim 4.
- **Enable New Products, Service and Markets (C3):** This characteristic covers new consumption like EV, new service like dynamic pricing. All these new products and service brings new markets and change traditional market feature. It does not only promote greener consumption but also produce new incentives for economic grow. So this characteristic reflects incentive from Aim 1, Aim 4.
- **Provide Power Quality for Range of Needs (C4):** This characteristic covers solutions to Power Quality disturbance which may increase the cost from reliability. So this characteristic reflects incentive from Aim 2 and Aim 3.

- **Optimize Asset Utilization and Operation Efficiency (C5):** This characteristic stands for the efficiency increasing and asset optimal utilization by the real-time information communication in all section of power system. It directly leads to lower cost and higher reliability. So this characteristic reflects incentive from Aim 2 and Aim 3.
- **Operates Resiliently to Disturbances, Attacks, and Natural Disasters (C6):** This characteristic covers the Smart Grid solutions that reduce the harms from disturbances, attacks and natural disasters. In other words, enhance the security and reliability so as to save cost from the harms. This characteristic reflects incentive from Aim 2 and Aim 3.

Though nations may have similar development, but their Smart Grid Scope should be based on their own situations. Comparing to DOE's Smart Grid characteristics, Electric Network Strategy Group (ENSG) has summarized their own characteristics for UK Smart Grid development in Fig 2.3.

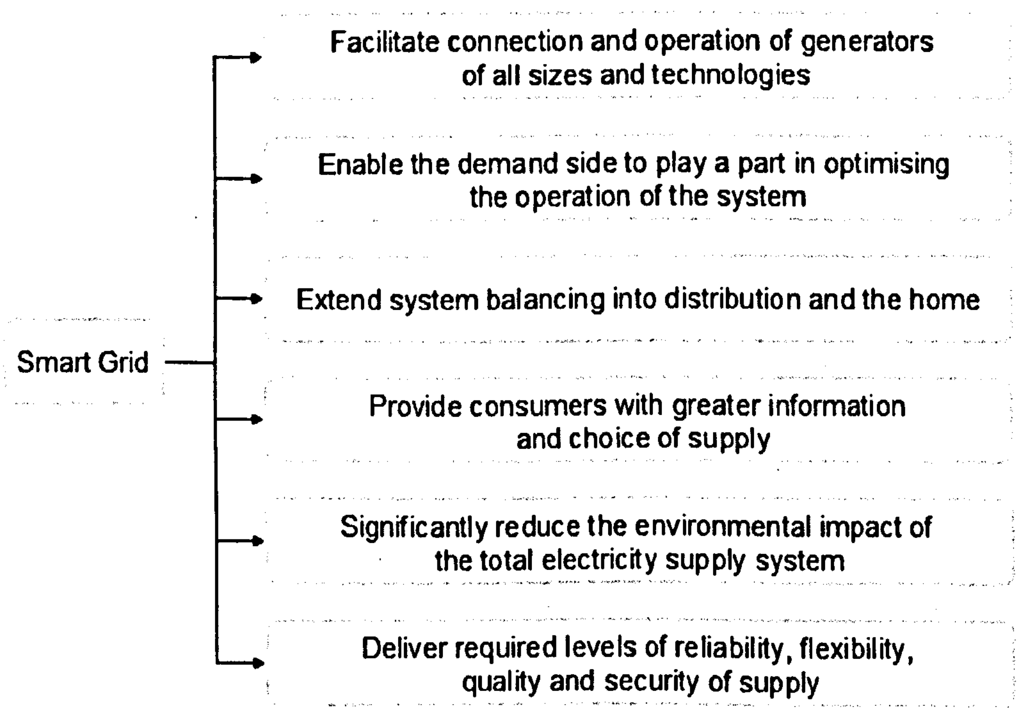


Fig 2.3: Smart Grid Characteristics by ENSG [18]

The six characteristics in Fig 2.3 reflect the 4 national development aims as well. While the characteristics from DOE mentioned the reliability and security in 3 items, the one from UK only mentioned this in one item. This is because US suffer much more serious reliability and security problem than UK. Various materials reveal that in the largest blackout worldwide which affecting more than 30 millions people, there are 2 from US but non from UK. In these blackouts which affect at least 1 million person*hour in the passed 5 years, US suffered more than 25 but UK only takes 10. The different situation suffered by US and UK influence their characteristics making of Smart Grid.

2.4.2 Smart Grid Technologies Metrics

With the characteristics, the next step in procedure is to find out technologies and standards that satisfy all the characteristics. So technologies in Smart Grid should reflect the Smart Grid characteristics.

Set US as example. U.S. Department of Energy (DOE) established a workshop for identifying metrics of measuring progress toward implementation of smart-grid technologies, practices, and services with 140 experts on June 20, 2008. At last over 50 metrics was hand in for smart-grid progress, in which 20 are for smart grid deployment [15]. Table 2.2 shows these 20 metrics.

	Metric Title	Characteristic Reflection
	Area, Regional, and National Coordination Regime	
1	**Dynamic Pricing**	C 1, C3
2	**Real-time System Operations Data Sharing**	C 5, C 6
3	**Distributed-Resource Interconnection Policy**	C 2, C 5
4	**Policy/Regulatory Progress**	C 1 to C 6
	Distributed-Energy-Resource Technology	
5	**Load Participation Based on Grid Conditions:**	C 1,
6	**Load Served by Microgrid**	C 1, C 5
7	**Grid-Connected Distributed Generation (renewable and non-renewable) and Storage**	C 2
8	**EVs and PHEVs**	C 3
9	**Grid-Responsive Non-Generating Demand-Side Equipment**	C 5
	Delivery (T&D) Infrastructure	

10	**T&D System Reliability**	C 4, C 5, C 6
11	**T&D Automation**	C 4, C 5, C 6
12	**Advanced Meters**	C 1, C 5
13	**Advanced System Measurement**	C 5, C 6
14	**Capacity Factors**	C 5
15	**Generation and T&D Efficiencies**	C 5
16	**Dynamic Line Ratings**	C 4, C 5
17	**Power Quality**	C 4
Information Networks and Finance		
18	**Cyber Security**	C 6
19	**Open Architecture/Standards**	C 1 to C 6
20	**Venture Capital**	C 1 to C 6

Table 2.2: Smart Grid Metrics

Each metric is corresponding to one or more characteristics in Table 2.2. Column 'Characteristics Reflection' introduce the characteristics relate to the metrics.

2.5 Smart Grid Scope

With the analysis on characteristics and metrics, following the scope design procedure, the Smart Grid Scope could be revealed. Various new technologies and standards satisfying the characteristics make the Smart Grid advanced and obviously different from traditional power grid. The following section introduces one scope sample from NIST of DOE, revealing the new component of Smart Grid.

The Conceptual Smart Grid framework model from NIST is mainly focused in this thesis for its wide compatibility and integration of various possible new technologies, new power consumptions and necessary element of traditional power grid.

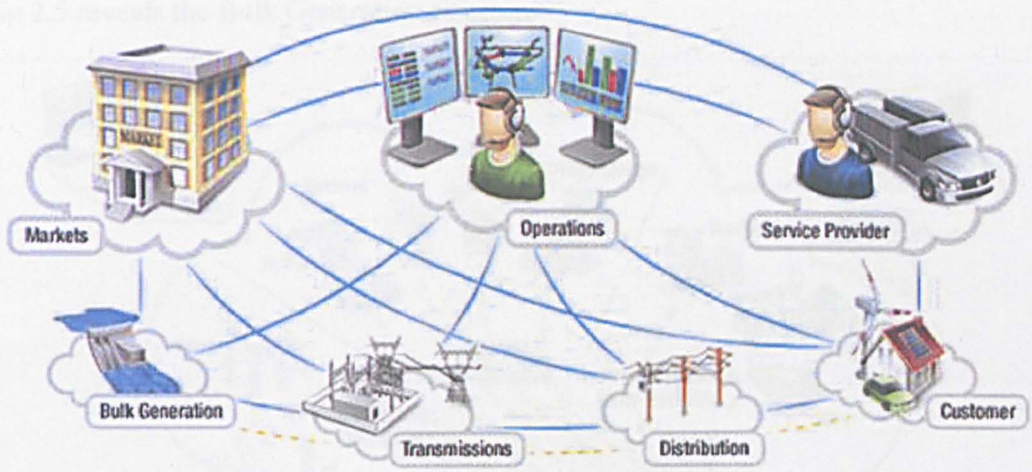


Fig 2.4: NIST Smart Grid Conceptual Framework Model [17]

"The National Institute of Standards and Technology (NIST) Smart Grid Conceptual Model provides a high-level framework for the smart grid that defines seven important domains: **Bulk Generation, Transmission, Distribution, Customers, Operations, Markets** and **Service Providers** [16]. All the 7 domains construct a sub-system of their own, though not completely separated. A communication and monitoring network are established covering all the above domains, aiming to provide a bi-directional information flow between related domains.

2.5.1 Bulk Generation

Smart Grid accommodates all generations and storage options. As centralized generations still plays a critical role, this domain mainly integrates all kinds of centralized power generation, storage types and specified monitoring and management for each generation and the whole generation systems. Traditional generations, e.g. Coal and large hydro, are definitely included. Renewable energy, including intermittent renewable energy, like solar and wind, and un-intermittent renewable energy, like wave energy and biomass, are also covered in this domain for they are the optimal choice in taking place of high polluted and carbon-emitted generations.

Comparing to tradition power system, the advanced areas are placed at:

- New types of generation seize larger percentage to traditional generation.
- New generations, Storage and their associate device create new market chances.
- Communication between components improves the asset utilization and operational efficiency.

Fig 2.5 reveals the Bulk Generation structure:

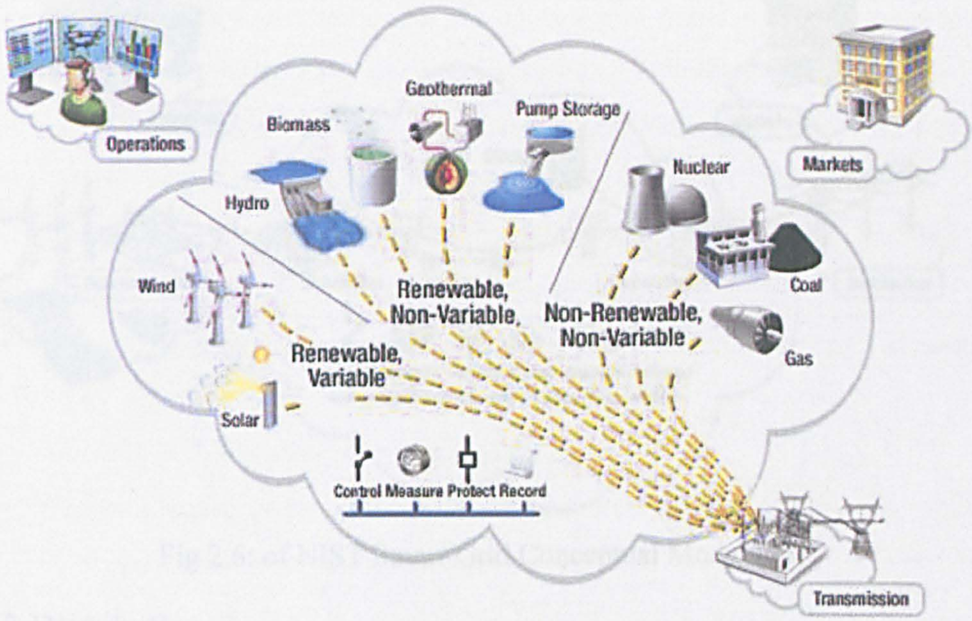


Fig 2.5: Bulk Generation of NIST Smart Grid Conceptual Model [17]

2.5.2 Transmission

To transport the energy from Bulk Generation to load centre, power transmission grid is still the necessary consideration of Smart Grid. In traditional case, large transmission grid swallows more than 10% energy of generation. Smart Grid aims to apply new technologies and management, e.g. HVDC, FACTS, Transmission Dispatch Automation, Communication Network, to reduce the consumption and enhance the efficiency and stability in this domain.

Comparing to tradition power system, the advanced areas are placed at:

- New types power delivery technologies reduce the line lost.
- Wide Area Communication platform encourage data transmission and sections communication that helps in better power dispatch and problems diagnosis.
- Substation Automation improves the operational efficiency and asset utilization.

Fig 2.6 reveals the Transmission of NIST Smart Grid Conceptual Model.

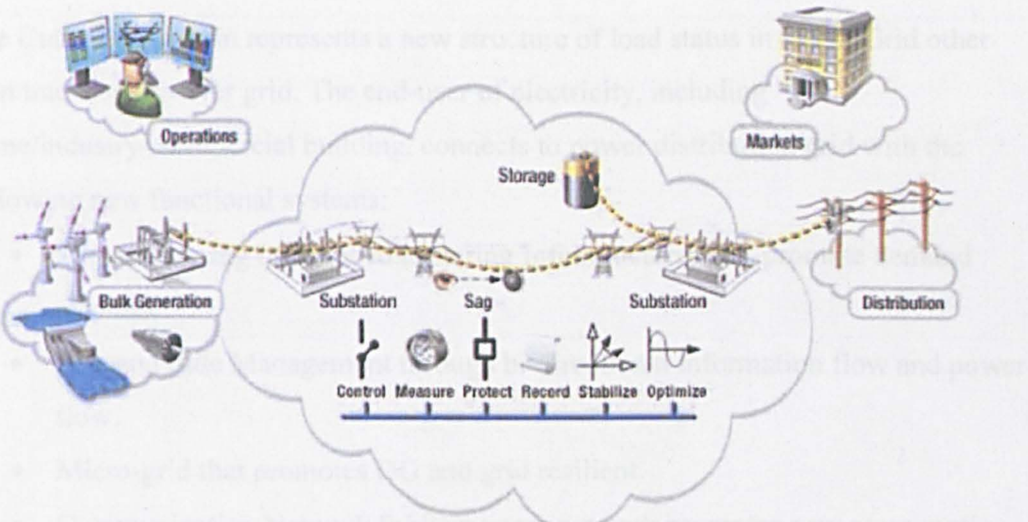


Fig 2.6: of NIST Smart Grid Conceptual Model [17]

2.5.3 Distribution

Smart Grid Distribution not only achieves the ability in traditional power grid but also integrate several new technologies including:

- Distributed Generation: Another generation type other than Bulk Generation.
- Real time monitoring, data analysis and management.
- Optimization and automation on power dispatch and grid protection.
- Various Power Quality selections.

Fig 2.7 reveals the Distribution domain of NIST Smart Grid Conceptual Model.

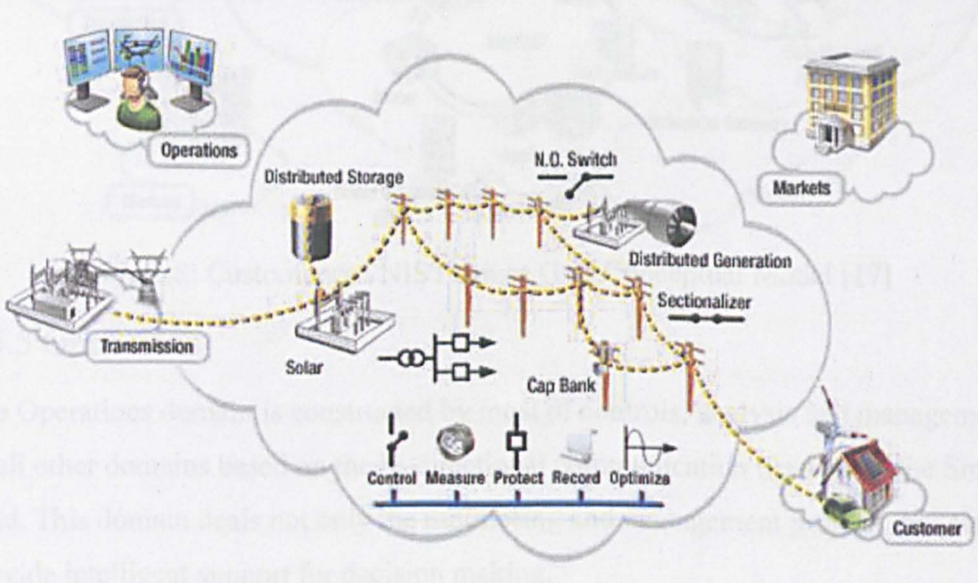


Fig 2.7: of NIST Smart Grid Conceptual Model [17]

2.5.4 Customer

The Customer domain represents a new structure of load status in Smart Grid other than traditional power grid. The end-user of electricity, including home/industry/commercial building, connects to power distribution grid with the following new functional systems:

- Smart Metering (Advanced Metering Infrastructure) that promote demand response.
- Demand Side Management through bi-directional information flow and power flow.
- Micro-grid that promotes DG and grid resilient.
- Communication Network for home area network promotes new consumption style and creates new service & markets.

Fig 2.8 reveals the Customers domain of NIST Smart Grid Conceptual Model

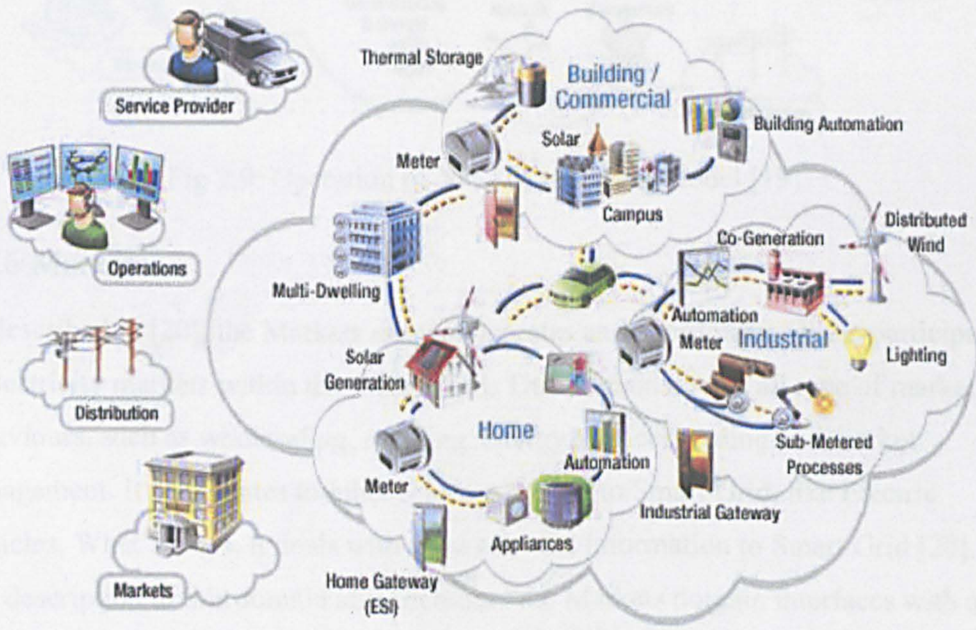


Fig 2.8: Customers of NIST Smart Grid Conceptual Model [17]

2.5.5 Operations

The Operations domain is constructed by most of controls, analysis and management of all other domains based on the bi-directional communication network in the Smart Grid. This domain deals not only the monitoring and management problem but also provide intelligent support for decision making.

Comparing to traditional power system:

- New types of equipment and services, like renewable generation, distributed generation, HVDC and dynamic pricing, require specified operations.
- Communication platform establishment has brought larger information sharing, so as have change the operation towards traditional sections in power system.

Fig 2.9 reveals Operation domains of NIST Conceptual Model.

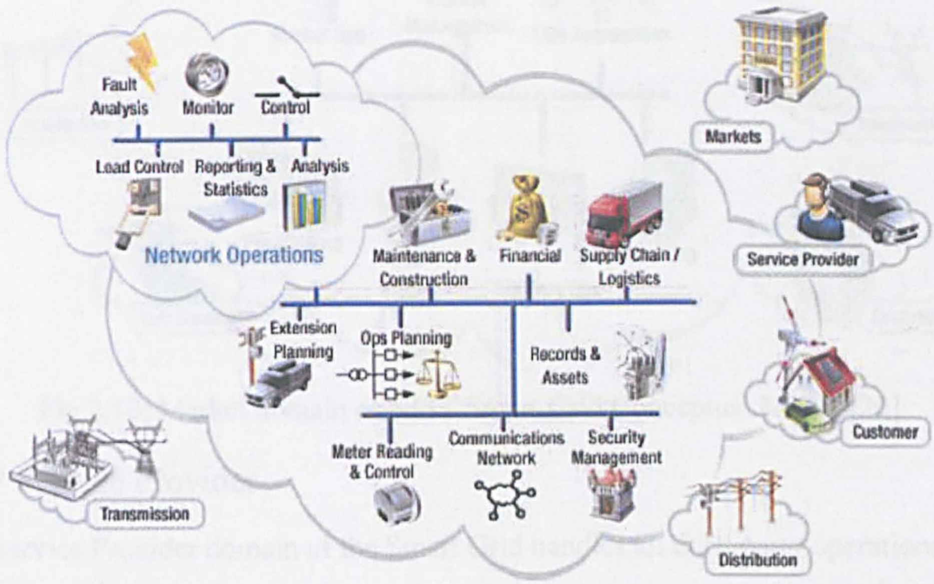


Fig 2.9: Operation of NIST Conceptual Model [19]

2.5.6 Markets

As described in [20], the Markets domain operates and coordinates all the participants in electricity markets within the Smart Grid. This domain covers all type of market behaviours, such as wholesaling, retailing, energy services trading and market management. It also relates to other relevant market to Smart Grid, like Electric Vehicles. What's more, it deals with most relevant information to Smart Grid [20]. The description to this domain also includes "the Markets domain interfaces with all other domains and make sure they are coordinated in a competitive market environment". This will be the ability of a power grid with permitting competition. Some monopolized or market-unopened power grid should establish their own rules to maintain a health market operations other than this Conceptual Model introduced. Comparing traditional power market, new types of consumption like EV and DG require new market behaviours to maximize their contribution to Smart Grid. Fig 2.10 reveals the Market domain of NIST Smart Grid Conceptual Model.

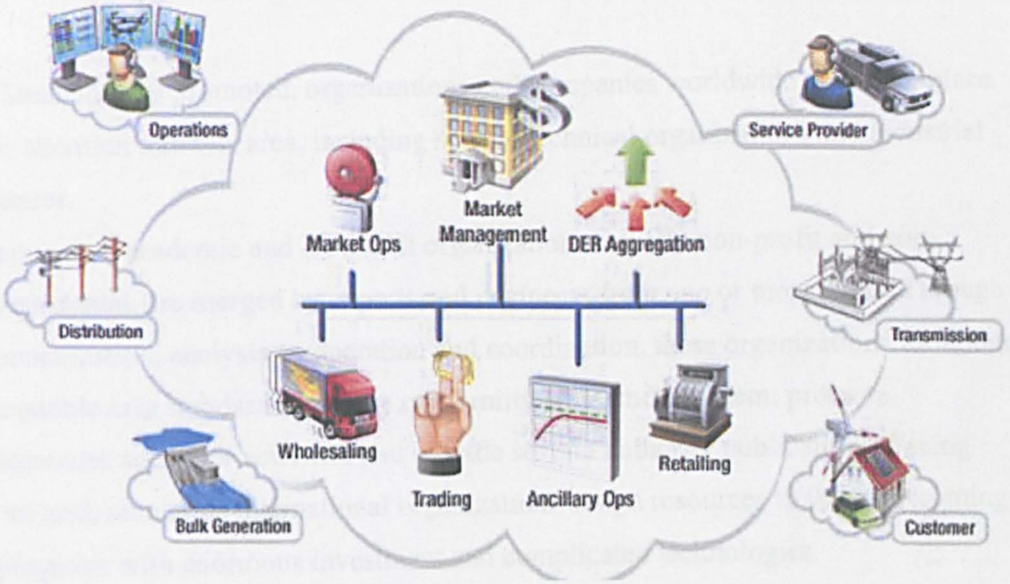


Fig 2.10: Market domain of NIST Smart Grid Conceptual Model [20]

2.5.7 Service Provider

The Service Provider domain of the Smart Grid handles all third-party operations among the domains. This section covers all requirements for establishment and maintenance for other sections. Fig 2.11 reveals the structure of Service Provider of NIST Smart Grid Conceptual Model [21].

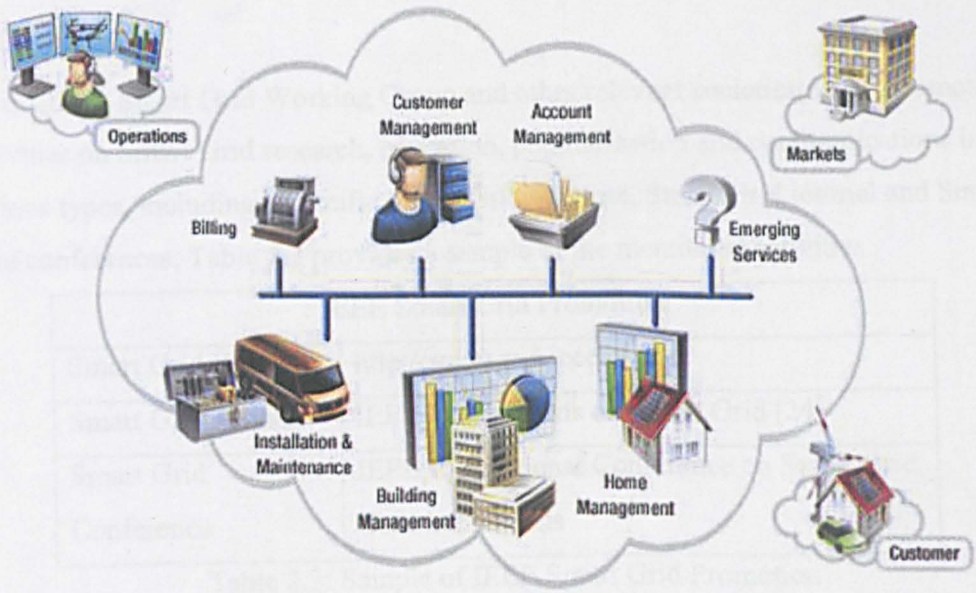


Fig 2.11: Service Provider of NIST Smart Grid Conceptual Model [21]

2.6 Smart Grid Related Organization and Standards

As Smart Grid is promoted, organizations and companies worldwide gradually place their attention into this area, including several technical organizations and industrial alliances.

International academic and industrial organizations, usually non-profit and non-governmental, are merged by experts and engineers from one or more areas. Through communication, analysis, cooperation and coordination, these organizations formulate compatible area standards; manage conformity assessment system; promote conferences and other activities and provide several authority publications. Facing smart grid, numerous international organizations assign resources to this new coming deployment with enormous investment and complicated technologies.

Smart Grid development is driven by various requirements from market and technology fields. For a few requirements, more than one solution could be discovered. To avoid the standards chaos from casual selection, companies and other relevant organizations start to unite together for supporting one possible optimal choice, promoting the nativity of alliance. The standards and policies supported by a large alliance are usually standing for widely accepted and influencing.

2.6.1 Institute of Electrical and Electronics Engineers (IEEE) and Its Smart Grid

Through the Smart Grid Working Group and other relevant societies, IEEE promotes activities on Smart Grid research, education, popularization and communications in various types, including a specific Smart Grid Website, Smart Grid journal and Smart Grid conferences. Table 2.3 provides a sample of the mentioned activities.

IEEE Smart Grid Promotion	
Smart Grid Website	http://smartgrid.ieee.org/
Smart Grid Journal	IEEE Transactions on Smart Grid [24]
Smart Grid Conference	IEEE International Conference on Smart Grid Communications

Table 2.3: Sample of IEEE Smart Grid Promotion

IEEE is a technical organization with various standards publications. In the standards approved and approving, there are nearly 100 standards and standards relevant to smart grid, including the over 20 IEEE standards named in NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 1.0. Facing the interoperability brought by Smart Grid, IEEE raise a standards series, 2030 Smart Grid Interoperability Series of Standards, as compensation for interoperability support. The approving standards from 2030 share the common goal of interoperability supported by interrelated and complementary technologies [26]

Samples of IEEE Smart Grid Standards are list in Table 2.4 [25].

Series No.	Working Group Approving	Title	Status
1547.3-2007	SCC21	Guide For Monitoring, Information Exchange, and Control of Distributed Resources Interconnected With Electric Power Systems.	Approved
802.11-2007	IEEE 802	IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications	Approved
P2030	SCC21	Guide for Smart Grid Interoperability of Energy Technology and Information Technology Operation with the Electric Power System (EPS), and End-Use Applications and Loads	Approving
P802.11	IEEE 802	IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications	Approving

Table 2.4: Samples of IEEE Smart Grid Standards

2.6.2 International Electrotechnical Commission (IEC) and Its Smart Grid

In June 2010, IEC SG3 published a document, "IEC Smart Grid Standardization Roadmap" edition 1.0, for describing the present situation of standards and regulating the future standardization planning. The whole Smart Grid standardization is divided into several clauses for specific areas. For each clause, the document describes systematically in the following structure:

- Area Description.
- Requirements.
- Existing Standards.
- Gaps between the existing standards and the requirements.
- Recommendations.

Basing on the describing structure mentioned above, IEC specifies a catalogue with about 20 items including 3 general items highly related to others. Table 2.5 selects some of the catalogues in IEC Smart Grid Roadmap [28].

	Title	Content
1	Smart transmission systems, Transmission Level Application	Including description of FACTS, HVDC, Cable Transmission and Long-distance transmission.
2	Distributed Energy Resources	Including description of Energy Management System (EMS), forecasting system.
3	Advanced Metering for Billing and Network Management	Including Advanced Meter Infrastructure and the bidirectional communication network between the smart grid and metering devices and business systems.

Table 2.5: Samples of catalogues in IEC Smart Grid Roadmap

Based on the Smart Grid Roadmap, IEC has established more than 100 standards relevant to Smart Grid. By classification, they can be put into 13 categories as Table 2.6 reveals [29]. Unlike IEEE Smart Grid standardization, IEC does not satisfy with only a standard compensation. The standards in roadmap form up a well organized standard framework that specified for Smart Grid.

Communication	Distribution Automation (DA)	Distributed Energy Resources (DER)
Distributed Management System (DMS)	Demand Response (DR)	Energy Management System (tech.) (EMS)
Electric Vehicle (EV)	Flexible Alternating Current Transmission	High Voltage Direct Current (HVDC)

	System (FACTS)	
Substation Automation (SA)	Storage	Smart home

Table 2.6: IEC established standards classification

Under these categories, examples of IEC Smart Grid standards are list in Table 2.7.

Reference	Topic	Title
IEC 61970-2	Common Information Model	Energy management system application program interface (EMS-API) - Part 2: Glossary.
ISO/IEC 14543-3-3	Information Technology – HES	Information technology - Home electronic system (HES) architecture - Part 3-3: User process for network based control of HES Class 1.
IEC 60633	HVDC - High Voltage Direct Current	Terminology for high-voltage direct current (HVDC) transmission.
IEC 61400-24	Wind Turbines	Wind turbines - Part 24: Lightning protection.

Table 2.7: Examples of IEC Smart Grid standards

2.6.3 ZigBee Alliance and Its ZigBee Communication Tech for Smart Grid

ZigBee is a standard-based wireless technology for the requirement of low-cost, low-power wireless sensor and control networks, which is interested by Smart Grid communication platform establishment. The communication platform from Smart Grid requires a communication technology that covers large network size, long battery life but do not necessary a too high data rate. ZigBee will be one of the best choice compare to Bluetooth and Wi-Fi as shown in Table 2.8.

Basing on IEEE 802.15.4 for Wireless Personal Area Networks (WPAN), ZigBee establishes an easy-used mesh network which mainly works around 2.4 GHz radio frequencies. The possible application areas are listed below [30]:

- Commercial building management.
- Energy management.
- Health care and fitness.
- Telecommunications.

- Residential management.
- Retail management.

ZigBee delivers unprocessed metadata at the rate of 250Kbs at 2.4 GHz (16 channels) for global utilizations, 40Kbs at 915 MHz (10 channels) for Americas and 20Kbs at 868 MHz (10 Channel) for Europe. Its low-power solution ranging from 10 to 1600 meters are used for transmission with dependence on the environmental conditions and the power output. A low-power feature could be seen in Table 2.8, a technical compare between ZigBee and other wireless communication techniques [31].

Market Name Standard	ZigBee 802.15.4	GSM/GPRS CDMA/1*RTT	Wi-Fi TM 802.11b	Bluetooth TM 802.15.1
Application Focus	Monitoring & Control	Wide Area Voice & Data	Web, Email, Video	Cable Replacement
System Resource	4KB – 32KB	16MB+	1MB+	250KB+
Battery Life (days)	100 – 1000+	1 – 7	0.5 – 5	1 – 7
Network Size	Approximate to Unlimited (64K)	1	32	7
Maximum Data Rate	20 - 250	64 – 128+	11,000+	720
Transmission Range (meters)	1 – 100+	1000+	1 – 100	1 – 10+
Success Metrics	Reliability, Power, Cost	Reach, Quality	Speed, Flexibility	Cost, Convenience

Table 2.8: Technical compare between ZigBee and other techniques

2.6.4 HomePlug Powerline Alliance and Its HomePlug Powerline Tech for Smart Grid

The HomePlug Powerline technology enables the power lines to transmit signals for communications as well as electrical power. This double-duty role for power line has been attempted for several decades and is achieved with a complex amalgam of signal processing technologies and new modulation techniques.

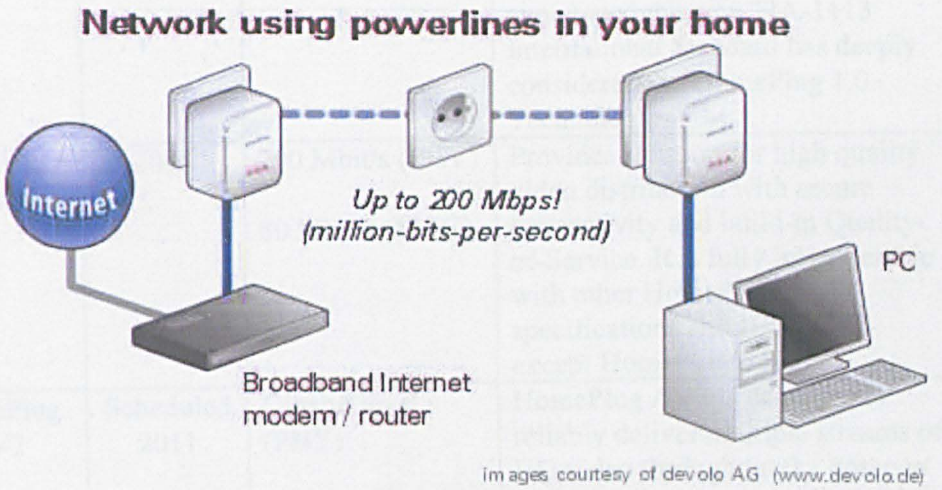


Fig 2.12: Powerline Network Diagram [34]

The popularization of technology also comes from the ubiquitous power outlets. The maturing area power wiring system provides a ready-made platform for powerline communication which leads to convenient installation. HomePlug powerline networks can also be used to extend wireless coverage by plugging access points into the powerline network at optimum points. A large number of applications by this technology are reveals in Table 2.9 [33].

	Areas
1	Whole-home broadband internet
2	HDTV Networking
3	Gaming
4	Smart Grid / Smart Energy
5	Whole-Home Audio

Table 2.9: HomePlug Powerline Technology Application [33]

After founded in 2000, the HomePlug Powerline Alliance keeps providing specification for Powerline networking standardization. Table 2.10 has shown a compare between different specifications from HomePlug Powerline Alliance.

Title	Time Published	Peak Speed	Description
HomePlug 1.0	2001	14 Mbit/s	In 2008 Telecommunication Industry Association (TIA) announced the new TIA-1113 International Standard has deeply considered the HomePlug 1.0 Technology
HomePlug AV	2005	200 Mbit/s (PHY) 80 Mbit/s (MAC)	Provides solution for high quality video distribution with secure connectivity and build-in Quality-of-Service. It is fully interoperable with other HomePlug specifications and IEEE 1901 except HomePlug 1.0.
HomePlug AV2	Scheduled 2011	Gigabit level (PHY) +600 Mbit/s (MAC)	HomePlug AV2 is designed to reliably deliver multiple streams of HD video throughout the home as well as next generation low latency content such as 3D and 4K HD video.
HomePlug Green PHY	2010	3.8 Mbit/s (PHY) 1 Mbit/s (MAC)	Developed as Smart Grid Communications protocol for connecting home appliances like HVAC and Smart Meters.

Table 2.10: HomePlug Specifications compare [33] [35]

Other than a splendid communication solution for entertainment distribution, HomePlug Powerline network will also allow Home Area Network (HAN) to communicate with smart meters and provide energy management for consumers. By working with ZigBee Alliance, HomePlug Powerline Alliance is helping to set up HAN ecosystem that enables intelligent energy management and efficiency in local area [36].

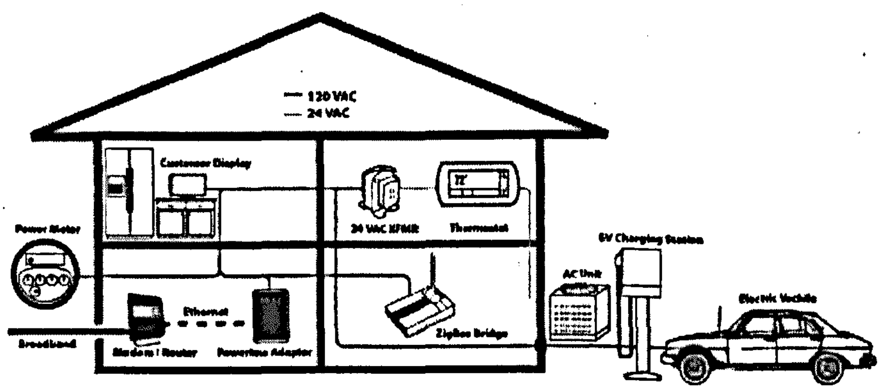


Fig 2.13: Smart Grid Home Area Network by Power Line Communications [36]

Smart Grid considers Home Plug Power Line as one of its communication technologies candidates as it utilize the existing network instead of invest to build a new one. This feature will reduce the initial investment but the network access is limited to the existing network sockets, which has less flexibility than the choice of wireless network.

2.7 Smart Grid Benefit

Power grid smartening and modernization, like any investment, does not come for free. But the benefits go far beyond the costs. Almost all the new technologies are born for new requirements with compatibility of new developments. So Smart Grid development covers the following benefits [37]:

- **Improvements in Reliability.**
  - Major Reduction in Outage Duration and Frequency.
  - Far Fewer Power Quality Disturbances.
  - Virtual Elimination of Regional Blackouts
- **Improvements in Security and Safety.**
  - Significantly Reduced Vulnerability to Terrorist Attack And Natural Disasters.
  - Improved Public and Worker Safety.
- **Improved Economics.**
  - Reduction or Mitigation of Prices.
  - New Options for Market Participants.
- **Improved Efficiency.**
  - More Efficient Operation and Improved Asset Management at Substantially Lower Costs.
- **More environmentally friendly.**
  - Much Wider Deployment of Environmentally Friendly Resources.
  - Electrical Losses Reduced.

If targeting on stakeholder angle, Smart Grid contributes its advantages as follow items reveals [38]:

- **Residential and Small Commercial Customers**: Smart Grid provides assessment and tool for customers to manage their energy consumptions with dynamic pricing.

Moreover, the enhanced reliability will help to reduce the risk and price by limit the risk of outage for specially needed people.

- **Large Customers**: Smart Grid provides more informatics and reliable power supply with multi Power Quality (PQ) Options.
- **Local Governments**: Including the consumer benefits, local governments could also gain advantages on the reduction of accidents or disaster. The more informatics ability from Smart Grid will help to provide a faster and more accurate decision making and action enforced.
- **Utility/Grid Operators**: A more efficient communication platform and data analysis system could increase the automation and the operation efficiency of the utility, so as to reduce the cost.
- **State and Local Economies**: The Smart Grid affordability ensures the power supply for economic development. Moreover, Smart Grid exploits the traditional market and introduces new markets, providing business chances and jobs. In power utilization angle, Smart Grid integrates much more renewable energy, supporting a sustainable economic development.

2.8 Smart Grid Demonstration Projects

Though Smart Grid appears to contain significant advantages, in specific environment, how excellent each Smart Grid advantage will be become an urgent question to answer before deployment. There is still requirement on proves and evidences for Smart Grid abilities. So demonstration projects or test-beds are applied for examining the Smart Grid abilities and working status.

2.8.1 Pacific Northwest Smart Grid Demonstration Project

The Pacific Northwest Smart Grid Demonstration Project is a regional project funded through a competitive process by the DOE under the American Recovery and reinvestment Act of 2009 (ARRA) across five Pacific Northwest states: Idaho, Montana, Oregon, Washington, and Wyoming, which involve more than 60,000 metered customers, and will engage, using smart grid technologies, system electricity assets exceeding 112 megawatts. The intent of the project is to verify the viability of smart grid technology and quantify smart grid costs and benefits which can be used to validate new smart grid business models at a scale that can readily be adapted and replicated nationally [39].

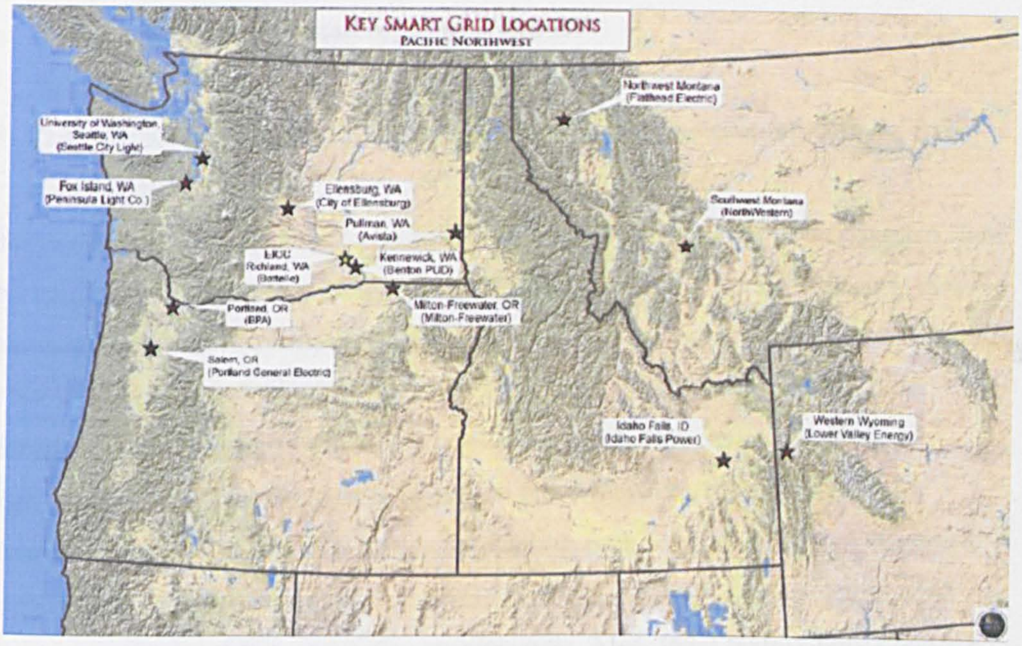


Fig 2.14: Key Smart Grid Locations of Pacific Northwest Demonstration Project [39]

The primary objectives of this $178M cost project are listed as follows [39]:

- Develop and validate an interoperable distributed communication and control infrastructure using incentive signals.
- Measure and validate smart grid costs and benefits for end-users, utilities, regulators, and others thereby laying the foundations of business cases for future smart grid investments.
- Contribute to the development of standards and control methodologies.

Apply smart grid capabilities to support the integration of renewable resources. This five-year project formally started on February 1, 2010. With its influence, up to 1,500 jobs created or retained at the peak of the project. New technology helps to update the aging electricity delivery system infrastructure and enhance reliability. The cost-benefit analysis will guide utility decisions on their future investments. The distributed energy source and the interoperability of the network will help to optimise the system efficiency and reduce the greenhouse gas emission. Customers will have more information about their own energy consumption so that they can become more aware have more choices [39] [40].

This project is one of the largest Smart Grid projects focusing on demand side issues. It provides a comparatively framework test to:

- General Distributed Generation, like wind and solar.
- New appliances, like smart appliances and EV.
- Reliability
- Demand Side Management Issues, including metering, communications, data analysis and control.

This project is mainly related to customer side and distribution in power system. Test bed for Smart Grid appearance when deploying general DG, Smart Metering, new consumption like EV and bi-directional information flow are the main concern of this project. Though it is difficult to evaluate this project with details as it is still in process, it is still doubtable to deploy a large project before several small experimental test as the impact is huge once fail. Moreover, the reliability and security of Smart Grid is inter-influencing among each sections of power system. It is in doubt that if a major work in customer side could reflect the entire Smart Grid ability on solution to Power Quality and disasters. Also, local feature may promote special DG type. Focusing on general DG type may miss the optimal generation plan.

2.8.2 EPRI Smart Grid Demonstration Initiative

"The EPRI Smart Grid Demonstration Initiative is a multi-year international collaborative initiative demonstrating the integration of Distributed Energy Resources (DER) in large scale demonstration projects. DERs integrated include demand response, storage, distributed generation, and renewable generation to advance widespread, efficient, and cost-effective deployment of utility and customer-side technologies in the distribution system and to enhance overall power system operations [41].

This Demonstration Initiative contains several Host-Site Demonstration Projects, with new structures of DER and several control & management methods. This Host-Site Demonstration Projects include: [41]

- American Electric Power (AEP) Smart Grid Demonstration Project.
- Consolidated Edison Smart Grid Demonstration Project.
- Duke Energy Smart Grid Demonstration Project
- Electricité de France (EDF) Smart Grid Demonstration Project.

- ESB Networks Smart Grid Demonstration Project
- Exelon Smart Grid Demonstration Project.
- The FirstEnergy Smart Grid Demonstration Project.
- KCP&L Smart Grid Demonstration Project
- PNM Resources Smart Grid Demonstration Project
- Southern California Edison (SCE) Smart Grid Demonstration Project
- Southern Company smart Grid Demonstration Project.

The EPRI Smart Grid Demonstration Initiative is a large Smart Grid project promoting plan:

- **Technologies Promotion**: Projects from EPRI Smart Grid Demonstration Initiative cover most areas in DER, including DG, EV, DR, Pricing, and Communications and so on. They provide inspections on DER framework integration and each project focuses on one core task of Smart Grid.
- **Market Promotion**: Most of projects are supported by one or more Smart Grid relevant utilities. The cooperation with industry not only relieves the pressure of government but also increases the industrial participation. Prepare of Smart Grid deployment is on the way.
- **International Cooperation Promotion**: Some of the projects do not locate in the U.S. only. Like '*Electricité de France (EDF) Smart Grid Demonstration Project*' is a French project by EDF.

All the projects in EPRI Demonstration Initiative have placed their target or a part of target into customer side. So Smart Grid development in consumers section is a general point of all the projects. E.g. all projects contain Demand Response Technology. But due to the aim difference, different projects may contain distinct task. For example, project with EDF is targeting to optimize the integration of distributed generation, storage and energy efficiency measure for providing load relief and reduce carbon emission. So this project includes Distributed Generation as it is a main content but do not include AMI. Comparatively, project ESB prefer to research on maximize the existing grid ability, connection with large wind farms and the customer response with real-time demand and consumption management. So AMI is certain content for it is the base for real-time demand and consumption management. And DG will not be included as unlike EDF, this project only interest in bulk generation like wind farm.

2.9 Conclusion

Smart Grid is a framework of multi-types of technologies and various stakeholders. This chapter is written aiming to introduce a generally Smart Grid scope design procedure which reveals the way of organizing various research work into scope establishment. It also attached Smart Grid details information on Standardization and demonstration worldwide. The procedure is introduced as follows.

1) Summarize the aims of national development to form an object for Smart Grid.
2) Based on the object, find out the new contributions to these aims from new power system, then the contribution forms up the characteristics. So each characteristic could reflect the aims.
3) With the characteristics, find out what technologies and standards could help to achieve these characteristics.
4) With the above three steps, a scope design finished.

Considering the necessity of compatibility among segments in Smart Grid, Section 2.6 introduces the current Smart Grid standardization situation worldwide. A variety of demonstration projects for Smart Grid are revealed as well in Section 2.8.
More Smart Grid reports and papers may refer to [43] – [52].

Chapter 3

Smart Metering Infrastructure

3.1 Smart Metering Introduction

AS high speed development of technologies today, more and more problems occur in power utilizations, such as:

- High cost energy consumption.
- Lack of equitable collocation of multi-energy-resources.
- Low fault detecting speed.
- Pollution and green house gas emit.

Facing these problems, smart grid, a topic recently attracted much attention, is one of the best choice to improve the situation. And Smart Metering, which is an important part of Smart Grid, is the closest way for energy user to be affected by the advantage of smart grid technology.

Considering the local requirement, functionalities classification and some other factors, the definition of Smart Metering Infrastructure, which short as Smart Metering, appear to be tiny different between countries. In United States, the Smart Metering Infrastructure, which named as Advanced Metering Infrastructure (AMI), is given an definition by United States Federal Regulatory Commission (FERC):

"Advanced Metering is a system that records customer consumption (and possibly other parameters) hourly or more frequently and that provides for daily or more frequent transmittal of measurements over a central collection point." [53].

European Smart Metering Alliance provides a structure to describe the Smart Metering definition, which shown in Fig 3.1:

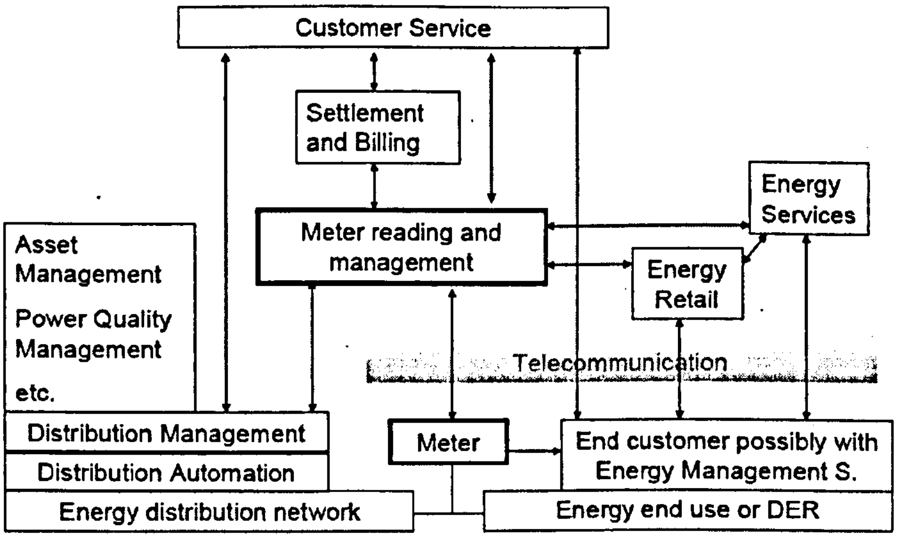


Fig 3.1: Structure of Smart Metering Infrastructure from ESMA [54]

But whatever definitions, Smart Metering Infrastructure should cover the following sections:

- Metering Automation.
- Data Analysis.
- Communication Platform.
- Monitoring and Management

The above 4 sections highlight the advantages between Smart Meter and traditional meter: bi-directional information flow and platform for new services.

3.2 Smart Metering Infrastructure Benefits

As a kernel segment of Smart Grid for demand side, Smart Metering Infrastructure was born for urgent requirement from multi-factors. By classification, drives are classified by stakeholders:

- Consumers.
- Utilities.
- Social and Environment Benefits.
- Other Benefits

3.2.1 Consumer Benefits

Smart Metering Infrastructure provide chances for more options, more informed services and better optimized control meeting the requirement of higher reliability, better power quality and more accurate billing [55]. Examples of consumer benefits are list below [54] [56] [57]:

- More accurate and timely billing.
- Improved access to the electricity market via accurate consumption history and possibilities to benefit from demand flexibility.
- Customer access various informed services.
- Improved safety of humans and equipment through better power quality and fault management.
- Ability to manage consumption by allowing customers devices remote control and communicate with meters.

3.2.2 Utilities Benefits

The Smart Metering Infrastructure helps utilities mainly in billing and operations improvement. [55]

- For billing, Smart Metering Infrastructure provide automatic meter reading for decreasing the cost of manual meter reading. Moreover, the bi-directional communication with consumers in new billing system helps the utilities in better decision making.
- For operations, Smart Metering Infrastructure provides a real-time optimized grid and assets monitoring and management for improvement of reliability, power quality and energy efficiency.

Examples of benefits for utilities are listed below: [55]

- Assess equipment health.
- Maximize asset utilization and life.
- Optimize maintenance, capital and O&M spending.
- Pinpoint grid problems.
- Improve grid planning
- Locate/identify power quality issues.
- Detect/reduce energy theft.

3.2.3 Social and Environment Benefits

The world is facing problems in shortage of energy supply, pollution and green house gas over emit. As a critical section of Smart Grid, Smart Metering Infrastructure improve energy consumption efficiency through demand response and other applications, which not only reduce the direct cost and cost from fault but also enhance customers awareness of low carbon lifestyle. Clean energy production is also encouraged for promotion of Renewable Energy from Smart Metering Infrastructure.

3.2.4 Other Benefits

The traditional meters deployed are continually over their life-period. New smart meters' deployment has just fit to this status by replacing new meters [57].
When recognizing Smart Metering Infrastructure as a platform of Smart Grid, communication infrastructure and other applications from Smart Metering Infrastructure could benefit other segments of Smart Grid, such as distribution, transmission and assets management. Also, Smart Metering Infrastructure promotes relevant development of technologies and industries such as the following areas [57]:

- Integrated Communications.
- Sensing and Measurement.
- Advanced Control Methods.
- Advanced Grid Components.
- Improved Interfaces & Decision Support.

3.3 Smart Metering Technologies

Once the requirement and definition of Smart Metering is clear, its necessary related technologies are obvious. The new metering system not only suffer the data reading ability from traditional system but also achieves communication, analysis and providing management of relevant segment of power grid.
European Smart Metering Alliance summarizes the Smart Metering technology options in Table 3.1 [58];

Technologies Classifications	Relevant Consideration	Description
Meter Design Options	Measured quantities	Measure data of electricity, gas, heat/cooling, and water and so on.
	Time interval	Multi-time-scale; short from millisecond to long as monthly or yearly.
	Disaggregated	Data classification for further analysis

	Data	
	Switch/valve	For requirement of limits, payment service and remote meter management.
	Multiple and dynamic tariffs	Multi-pricing services
Wide area data communication	Considerations for wide area networks (WAN)	Universality, reliability availability and transfer time of WAN, security, relevant Support, data accuracy and consistency, bandwidth, speed of response, public/private communications networks, interoperability, multi-utility
Software systems and data stores	Data collection, processing and storage.	
Customer feed back and local area communication	Customer feedback route options	Several feedback data assessment such as mobile text, internet and so on.
	LAN and final customer feedback design considerations.	Considering: Security, multi-utility and smart homes, demand response and embedded generation, data transfer rate, installation cost, meter battery life and disposal, meter and display energy usage.
	Feedback and display to final customer consideration	Display services and technologies selection

Table 3.1: Samples of Smart Metering technology options from ESMA

Other than the above European description, the United States National Energy Technology Laboratory (NETL) summarizes another Smart Metering technologies classification in NETL Modern Grid Strategy. It collects the relevant technologies into the following five classes [55]:

- Smart Meters
- Wide-area communications infrastructure.

- Home (local) area networks (HANs).
- Meter Data Management Systems (MDMS)
- Operational Gateways

Compare to the Smart Metering from ESMA, NETL describes Smart Metering in an application framework other than in technical side.

3.4 Smart Metering Standards

Smart Metering Standardization is essential for industrial and market promotion while preventing redundancies and chaos. Several standardizations for Smart Metering are being processed or born already such as:

- The United States National Institute of Standards and Technology (NIST) have promoted a Smart Grid Interoperability Standards, covering Smart Metering interoperability [59].
- The Commission to the European Standardization Organization (ESOs) has promoted a specific mandate (M/441), which is "To create European Standards that will enable interoperability of utility meters (water, gas, electricity, heat) which can then improve the means by which customers' awareness of actual consumption can be raised in order to allow timely adaptation in their demands" [10].

As the Chapter 2 describe, IEC has published its "IEC Smart Grid Standardization Roadmap" for the whole Smart Grid planning and deployment. Smart Metering, as the core section of Smart Grid, is described systematically in this roadmap. Examples of IEC Smart Metering Standards are revealed in Table 3.2:

IEC Standards Types	IEC Standards	Description
Product standards	IEC 62054 parts 11 and 21	Electricity metering (a.c.) – Tariff and load control specify type test requirements and test methods for tariff and load control equipments.
Payment systems standard	IEC 62055 series Electricity Metering	Payment systems specify a framework for standardization.

Reliability standards	IEC 62059 series Electricity metering equipment	Dependability specifies reliability prediction and assessment methods.
Standards for data exchange	IEC 62056 series	Data exchange for meter reading, tariff and load control specifies meter data exchange.

Table 3.2: Samples of IEC Smart Metering Standards [29]

3.5 Smart Metering Worldwide

Smart Metering Infrastructure is focused around the whole world as Smart Grid development in all countries. Different countries perform different procedures. Fig 3.2 reveals parts of the Smart Metering progress worldwide:

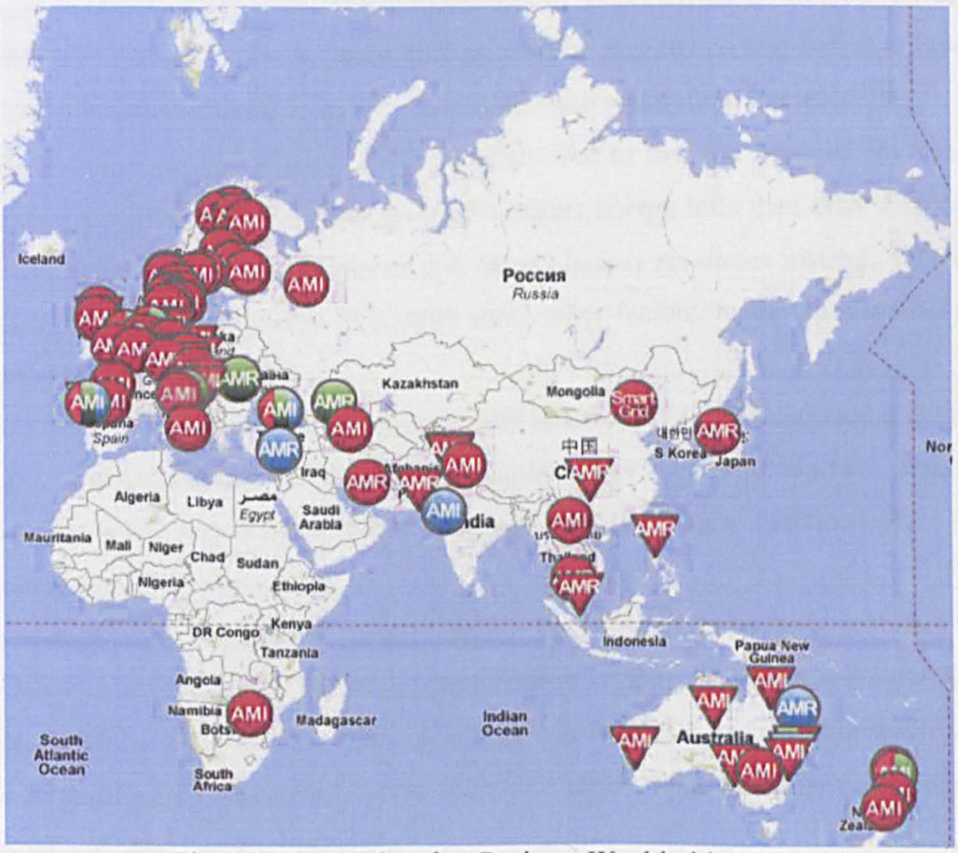


Fig 3.2: Smart Metering Projects Worldwide [60]

3.5.1 Smart Metering in UK

In 2009, following other EU countries' application, UK government had announced that the smart metering program was paved and First Utility was chosen to be the first provider of smart meters.

The upgrade of smart meters cost each household 340 GBP, which is believed to be saved back from the high bills in the future. Compare to the bills of old style meters, £800 for gas and £445 for electricity annually, new smart meters could help people save more than £28 every year, which is 2% to 3% on estimation.

This smart metering program is supposed to be finished by 2020. UK government place strong wishes on this meter upgrading, hoping it could change the power consumption habit of people. Also, on estimated, after the finishing of this program UK could reduce CO2 emissions by 2.6 tones, which contributes to improvement of climate situations [61].

3.5.2 Smart Metering in Italy

When talking about smart metering industrial program, Italy is an unavoidable topic. With the largest and earliest development of smart metering, Italy, cooperated with ENEL, becomes a star in smart grid application worldwide and Italian achieved to be the first group of people to share to benefit from smart metering technology.

Before meters upgraded from 2001, probably due to that the material for energy generation is hydrocarbons, Italian suffered a higher energy bills than other European countries, which using nuclear power and other cheaper resources instead. The high demand of reduction in energy bills, with some other factors, made the midwifery of smart metering application.

Until 2001, the third-largest energy provider in EU, ENEL, had depicted a plan for 5 years, covering 40 million homes and business. New meters in used are based on power line technology from Echelon Corporation. All the new functions other than old meter were own designed.

As the result of early deploy of smart metering, Italy is the first country, probably the only country, enjoying the rich benefit from energy savings. The harvest is up to $750 million annually. Italy has provided a good experiment and example to the world. After its deployment, countries all over the world started their own smart metering program one after another [13] [62].

3.5.3 Smart Metering in United States

Holding the largest economy and nearly the highest electricity consumption in the world, the United State requires a better effective in power system as well as China.

California is a typical example. As a result of its climate, a summer peak demand for near 50-100 hours every year appears to this state. Hot weather leads to more usage of air-conditioner, then influencing the electricity usage deeply. In this situation, the state's three largest investor-owned utilities, Pacific Gas & Electric Co., Southern California Edison and San Diego Gas & Electric, started their projects on deploying new meters of two-way communicating to their customers.

Set PE&G as an example. On July 20, 2006, PE&G received a project of meters updating from California's energy agency on near 9 million household customers in Northern California. New meters have the ability on recording and reporting the consumed gas and electricity and the price hourly. Users could shift their energy usage to a cheaper time, e.g. off-peak time, to save their bills [13] [62].

According to the California Public Utilities Commission, the savings from new meters will cover about 70% of smart meter investment.

Other states in US are following. Los Angeles Department of Water and Power (LADWP), which is the largest municipal utility in the USA, decided to apply the AMI (Automatic Metering Infrastructure) to their own customers. Austin Energy also begun to deploy a two-way RF mesh network and around 260,000 residential smart meters from 2008, and more than 165,000 smart meters have been installed by 2009[13].

An interesting phenomenon in USA is that IT giants like Google, Microsoft and Cisco have also entered this area providing their products. Google and Microsoft have Web-based software, which offer chances for them to popularize in partnerships with utilities and smart meter makers, and Cisco has plans to make home energy management hardware as part of a broad-ranging set of smart grid efforts [13].

3.5.4 Smart Metering in Oceania

In Australia, states have already taken actions on smart metering deployment, like Victoria State. The so-call Advanced Metering Infrastructure (AMI) program is started, trying to reduce their energy consumption and the carbon emission. The influence will spread up to 2.2 million homes and 300,000 businesses. The new meters include most of new abilities of smart meters, such as accurate electricity pricing system for every 30 minutes and a two-way communication between power companies and customers. A study made by National Cost Benefit Analysis reveals near $700 million benefit coming from the coming two decades.

In another country of Oceania, New Zealand started its smart metering program at

2006. But the effects of the new technologies are not as good as expected. The New Zealand Parliament was presented a report in June 2009; bring the installed smart meters to account. This report reveals that new meters have problems in real-time monitoring functions. What's more, the new meters cannot communicate with other devices for lacking a device at initial installation. Problems also occur in lacking of other basic functions and compatibility as the result of no an agreed standard for corporations. These make the deployment of smart metering in New Zealand an example of opposite side [13] [62].

3.6 Conclusion

As a core section of Smart Grid, Smart Metering becomes a base for a lot of technologies integration. As Section 3.1 introduces, Smart Metering is not only a better automatic digital measurer but also helps in data analysis and management. It has a wide beneficial range (Section 3.2) and needs a verity of technologies support (Section 3.3). It also appears to be one of the Smart Grid segments which achieve deepest development. Various standards and deployment procedures are born (Section 3.4 and 3.5). More Smart Metering reports and papers may refer to [63] to [67].

Chapter 4

Artificial Neural Network in Load Forecasting of Smart Grid

4.1 Introduction to load forecasting

The load being forecasted in Smart Grid is the power utilized by a specific group of customers at a time point of a period of time. It is a multi-factor-related nonlinear problem including the time factors, the weather factors and other influencing factors. It is widely applied in all power system segments facing different specific requirements from commercial, industrial and residential situations.

Separated from the range of time, power system load forecasting could be classified into three categories: short-term load forecasting which focus from minutes to one week; medium forecasting which focus from one week to a year and the long-term load forecasting which focus from one year to more than decades. The nature of load variation between different time horizons is different. Generally speaking the larger the time range is, the more influencing factors will appear and the more complexity the problem will be. For example, if talking about short term load of a city in a day, it could be influenced by weather, time, and customer behaviours. But when the time range expanded to decades, the economic variation should be considered as it may manipulate the power consumption. Due to this reason, short-term load forecasting could achieve accuracy as less than 8% for next day load but the long-term load forecasting could not. Load forecasting for longer time range usually focuses on more general power load, like the next year peak load forecasting. Figure 4.1 reveals an example of load forecasting.

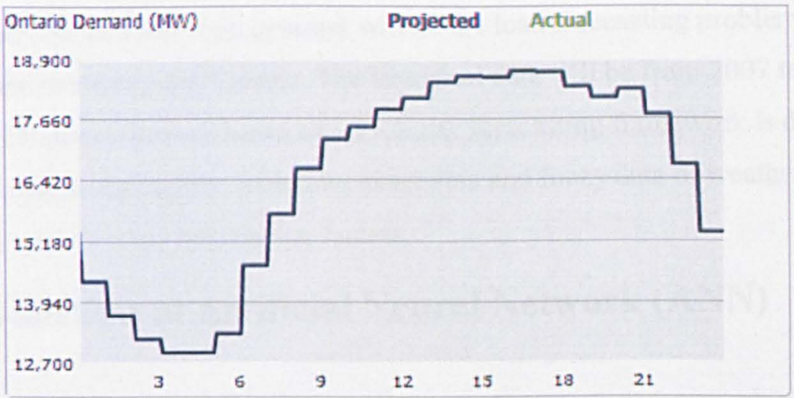


Fig 4.1: Load forecasting on August 30, 2011 for Canadian Ontario Demand by ieso [79]

Load forecasting helps decision making for electric utilities in electric power generation, load switching and infrastructure development. Base on weather information, utilities could approximate load flow situation and prevent overloading by short-term load forecasting data. It also helps in decreasing the failure and blackouts. In Smart Grid demand response, customers will receive the predicted price from utilities at smart meters that help users to manage their consumption. The predicted price is based on load forecast with the latest load information. So load forecast is a basic element for demand response in Smart Grid.

Various short term load forecast methods have already developed. [103] introduce an exponential smoothing algorithm for weekdays' load forecast. It provides a smooth enough result but only recognize the load curve shape instead. Large error may occur if the model does not focus on the influencing pattern but only the curve shape. [104] suggested a Similar Day method based on searching historical days. This method mainly based on the similarity of the historical data and will achieve a good result if the load repeating regularly. However, it is not sufficient enough to capture complex pattern. [105] provides a wavelet based neural networks for short term load forecasting. It successfully achieves a pattern recognized forecaster with smooth output. But in actual situation it is difficult to say if some of harmonics are belongs to real demand. So the prediction of wavelet neural network will have better generalization ability but to a degree will ignore some practical situation. [106] has introduce load forecast by Artificial Neural Network with weather data, but the weather influence is not all exact but contains fuzzy concept. This factor may lead to large error in some special case. More materials for load forecasting are introduced in [68] to [78].

In this chapter the load forecasting target will be the load forecasting problem in Ontario power consumption, Canada. The historical data will be from 2007 to 2009. An Artificial Neural Network based exact – fuzzy forecasting framework is designed with multi-influencing factors, including exact data and fuzzy data of weather condition as well as other influencing factors.

4.2 Introduction to Artificial Neural Network (ANN)

4.2.1 Artificial Neural Network Model

Human's biological brain can be recognized as a biology-based information processing system. It receives signal from sensors, like eyes, skin, ears, to collect information outside, forming an inside recognizable model and guide the decision making, including pattern recognition and forecasting.

Neurons are the chief components of the brain. The connections between neurons shape various highly coupled serial or parallel networks. Generally the human brain palliums contain over 10 billion neurons [80]. There are multi-types of specific neurons in our brain. A neuron from each type could connect by signal to other hundreds or thousands neurons [81]. Though a single neurons take time to process a single event in milliseconds level, which far longer than 10^{-9} second in silicon chips. But the human brain's each action per second only takes energy at level 10^{-16} J, comparing to the best computer today at 10^{-6} J [80]. Moreover, the information collected and stored in human brain is quite larger than the best technology today.

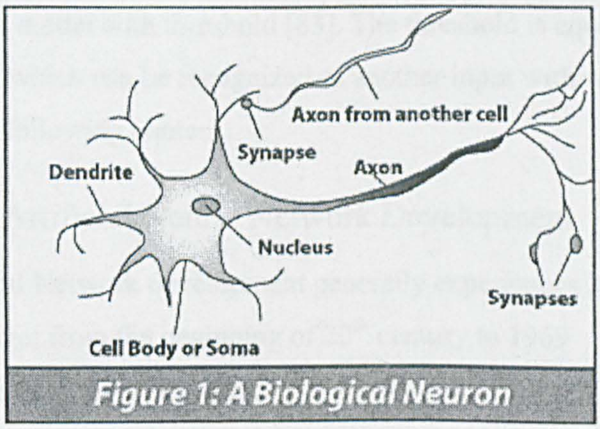


Fig 4.2: Biological Neuron [82]

Fig 4.2 reveals the structure of a biological neuron that the brain works with. There 3 core segments:

- Dendrite: short branch stretching out from the neuron cell body. It collects signals from outside or other neurons with weights.
- Axon: a long branch stretching out from the neuron cell body. It passes the signals from the neuron cell body.
- Neuron cell body: It collects the signal from dendrites and process the sum of the signal. The result is passed to Axon.

So by extracting the signal processing procedure from a single neuron's operation status, the result will be as shown in Fig 4.3.

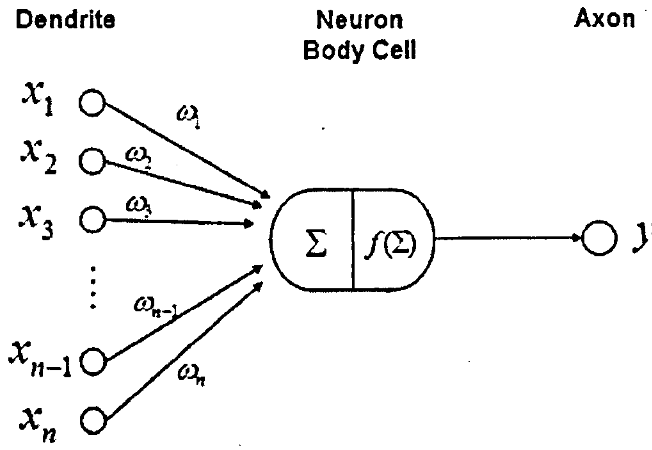


Fig 4.3: Neuron working procedure

From Fig.4.3, signals x are passed from Dendrite with weights w to Neuron Cell Body. After a summarization and a processing function, the neuron gives out an output y. This is the basic model of Artificial Neural Network. In 1940s McCulloch-Pitts mentioned a neuron model with threshold [83]. The threshold is equal to add a bias to the summarization, which can be recognized as another input with weight 1. This will be discussed in the following content.

4.2.2 History of Artificial Neural Network Development

The Artificial Neural Network development generally experiences 3 stages.

- Starting Stage: from the beginning of 20^{th} century to 1969
- Silent Stage: from 1969 to 1982.
- Prosperous Stage: from 1982 till now.

The development of math, biology and computing technology are the three main promotions to ANN development, influencing the developing limit as well. Table 4.1 reveals some important events of ANN development history.

Time	Events
1943	McCulloch and Pitts summarized basic neuron model, MP model, which is said to be the creation of Artificial Intelligence. [83]
1949	D.O. Hebb promote Hebb Learning Rules in his book "The Organization of Behaviour" for weight learning in ANN
1957	Rosenblatt promoted a new concept: Perceptron, which suggesting a new supervising learning method for pattern recognition.
1960	Widrow and Hoff introduce Least Mean-Square, LMS, for Adaptive Linear Element.
1980	Grossberg established Adaptive Resonance Theory for self-organizing theory.
1984	Hopfield design his ANN model electric circuits.
1986	Rumelhart, Hinton and Williams develop the Back-Propagation algorithm in their book "Parallel Distributed Processing: Explorations in the Microstructures of Cognition"

Table 4.1: Important Events of ANN Development

The ability of non-linear and parallel learning and memory attracts attention in various areas as modelling, time series analysis, pattern recognition, signal processing and control, especially facing problems that lack of physical understanding or non-linear varying data.

4.3 Theory of Perceptron

4.3.1 Perceptron Basic

Perceptron was first introduced by Rosenblatt in 1957 as a typical model of supervise learning. It is a basic feed forward network that achieves approximation to specific mapping. Fig 4.4 introduces the model of a single layer Perceptron.

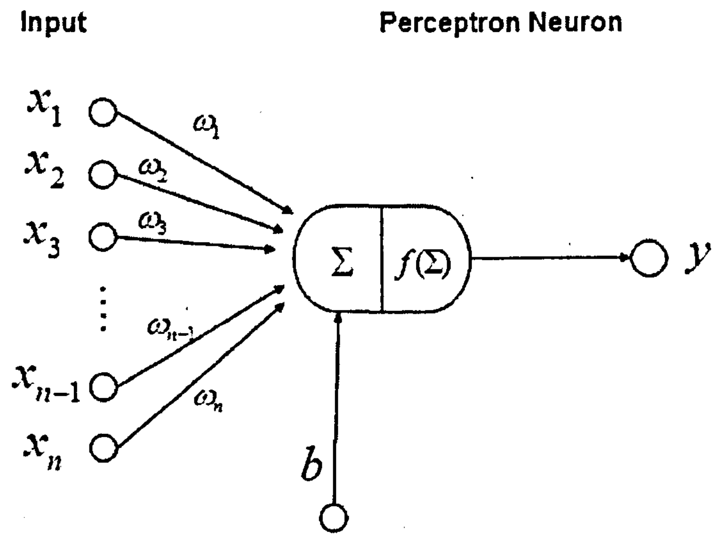


Fig 4.4: Perceptron neuron model

The model accepts n inputs with corresponding n weights. Neuron sums up all the input with weights then passes to a function f, which named activation function. b in Fig 4.4 is a bias, which standing for the threshold value mentioned in 4.2.1. Equation 4.1 expresses the relation in Perceptron model.

$$y = f(\sum_{i=1}^{n} \omega_i x_i + b) \tag{4.1}$$

The activation function plays a core segment in ANN information processing. It is a function with the following features:

- Non-linear: Achieves non-linear mapping between function input and output, making it possible for non-linear mapping between Perceptron input and output.
- Continuous and differentiate: The differentiable and continuous feature support the weight variation in Back-Propagation training, which will be mentioned in 4.5.
- Between a range limited by an upper value and bottom value: Limits the output value.

Normally activation function has three types: Step function and sigmoid functions.

Step function

The Step function is firstly introduced when single layer perceptron was created. It simulates the working mode of biological neuron. Table 4.2 introduces the step function.

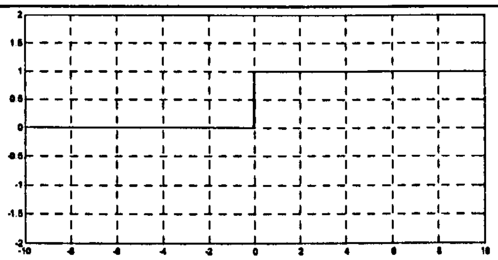
Step Function
$y=\begin{cases}+1, x \geq a \\ -1, x < a\end{cases}$ α is the value that function jump


Table 4.2: Step function for activation function

Sigmoid function

Sigmoid function stands for all the function whose shape similar to letter S. It is the most widely used function, including series of logarithm functions and series of hyperbolic tangent functions. Sample is selected in Table 4.3.

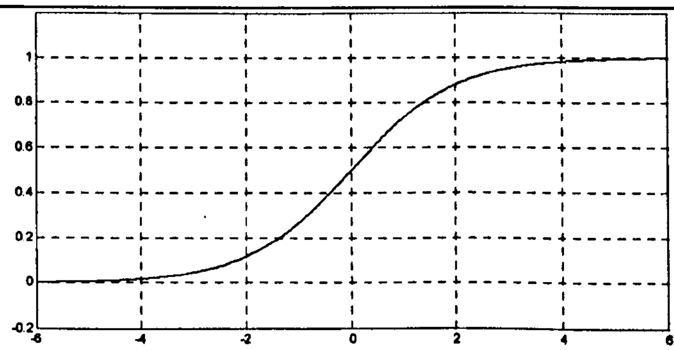
Step Function
$y=\dfrac{1}{1+e^{-x}}$


Table 4.3: Sigmoid function for activation function

4.3.2 Multi-layer Perceptron (MLP)

Multi-layer Perceptron is an expansion of Perceptron neuron model. As Fig 4.5 reveals, there are three types of layers in the MLP's architecture, which are input layer, hidden layers and output layer. Input layer accepts network input from outside while hidden layers and output layer contains the perceptron neurons. Number of neurons in input layer is the dimension of input vector while dimension of output vector determines the number of neurons in output layer. The number of hidden layers and the number of neurons in each hidden layer are controllable by users. It deeply influences the complexity and the performance of MLP. Between each two layers, the output from one neuron is passed to neurons in the next layer with weights. These weights are the core members for neural network training. In other words, Artificial Neural Network is processed by architecture selection and weights selection to achieve a preferred mapping between input and output.

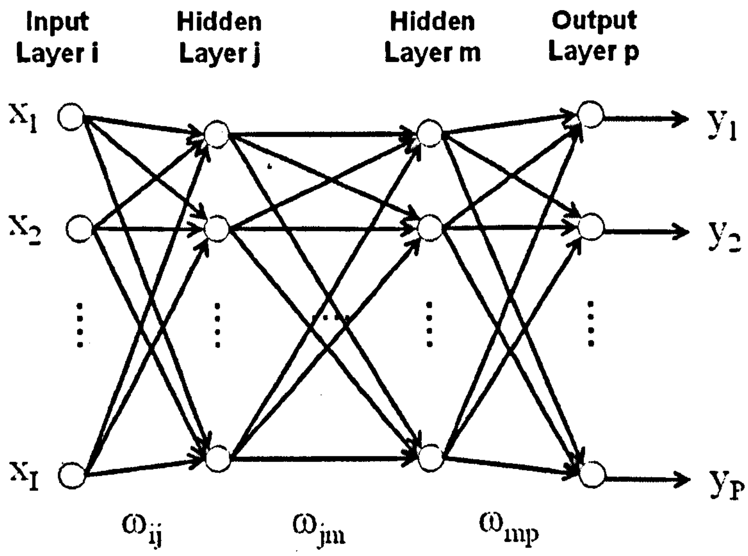


Fig 4.5: Simplified Model of Multi-Layer Perceptron

In 1989, G.Cybenkot provide a demonstration in Reference [84], revealing that any continuous function can be uniformly approximated by a continuous neural network having only one internal hidden layer and with an arbitrary continuous sigmoidal nonlinearity in the unit hypercube. In the same year KEN-ICHI FUNAHASHI proved the approximation realization ability of a k (≥3) – layers ANN in Reference [85]. The similar demonstration provides a mathematic insurance of the model possibility.

4.4 Back-Propagation Training of Multi-layer Perceptron

Models are born without any realization of the target problem before learning or training. The Process of training is the procedure of looking for best parameters set to approximate the mapping.

4.4.1 Supervised Learning and Unsupervised Learning

For machine learning, there are two general learning style, unsupervised learning and supervised learning. Unsupervised learning is a learning style operating without influence from outside. It usually provides output by competition. Supervised learning brings a target for each set of input to guide the system learning. System refers to the target and tries to fit itself closer to the preferred output.

4.4.2 Delta rule and Gradient Descent with Batch Learning

Gradient descent is an optimization algorithm with first-order differentiation targeting on the local minimum of a function. It is a widely applied optimization method. If there is a continuous function f(x), Equation 4.2 introduces this method for local minimum in a general case with this function. As it is shown, at the point X_n, the searching direction for new point X_{n+1} is the negative of the gradient. η is a small positive value that controls the step length.

$$X_{n+1} = X_n + \Delta X_n = X_n - \eta \frac{df(X_n)}{dX} \tag{4.2}$$

Delta rule, or so-call Widrow-Hoff rule, is the Gradient Descent application in Artificial Neural Network weights learning. The function whose minimum is interested is the least mean square error between real network output and the targets. The whole training process with Delta rule will be described in the following section. There are two training styles in ANN training, serial learning and batch learning. When using a training set, serial learning means that the error is calculated as every training element pass through the model. Compare to serial learning, batch learning groups all the training vectors into one epoch to calculate the average error. Though the serial learning cost less space and could achieve higher speed, it could not assure converging to the minimum. Batch learning operates with larger data amount, but it is easier to converge.

4.4.3 Back-Propagation (BP) Training Theory

Back-Propagation training is an expansion of Delta rule. With utilizing the negative gradient direction, the weights in ANN update from output side gradually to input side, so as to improve the system closer to the specific mapping. This method is held by

Werbos firstly in [86] and developed by Rumelhart in his book "Parallel Distributed Processing: Explorations in the Microstructures of Cognition" [86].

Each step of BP training is generally divided into two segments as Fig 4.6 shows:

- **Forward Calculation**: With the present weights and input, calculate the present output.
- **Error Propagation**: Error signal is generated by the differences between present output and target. With this signal, procedure is taken gradually from output side to input side. Following this procedure, weights are gradually adjusted by the error signal.

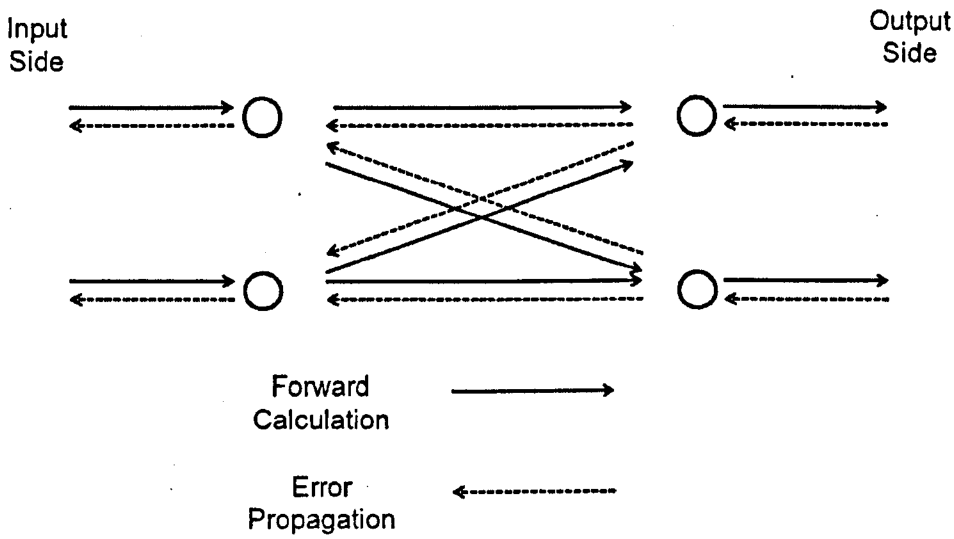


Fig 4.6: Forward Calculation and Error Propagation in BP

An ANN model with one input layer with input number I, one hidden layer with J neurons, and one output layer with output number P is selected to reveal details of Back-Propagation Training in the following section. Fig 4.7 reveals the selected model. Batch learning is selected for a better model converges.

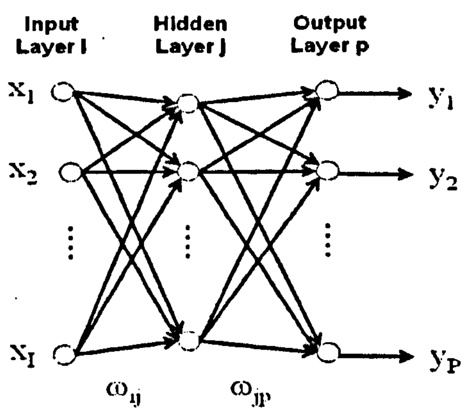


Fig 4.7: Typical 3 layers ANN

Back-Propagation Training Calculation Prepare

Assume the training set $X = [X_1, X_2, \cdots, X_N]$ contains N groups of input vectors. In each input vector $X_k = [x_{k1}, x_{k2}, \cdots x_{kI}]$ ($k = 1,2,\cdots,N$), there are I elements, which is the same as neurons' number in input layer i. The actual output corresponding to the k_{th} input vector $X_k = [x_{k1}, x_{k2}, \cdots x_{kI}]$ is $Y_k = [y_{k1}, y_{k2}, \cdots y_{kP}]$, the same as the model's output number. The expected output corresponding to the k_{th} input vector, so-call target, is $t_k = [t_{k1}, t_{k2}, \cdots t_{kP}]$. Assume the training process is at the n_{th} epoch.

The function that wished to be achieved minimum is batch average error E_{av} in equation 4.3.

$$\left.\begin{aligned} E_{av}(n) &= \frac{1}{N}\sum_{k=1}^{N} E_k(n) \\ E_k(n) &= \frac{1}{2}\sum_{p=1}^{P} e^2{}_{kp}(n) \\ e_{kp}(n) &= t_{kp} - y_{kp}(n) \end{aligned}\right\} \tag{4.3}$$

Equation 4.3 description: for the k_{th} input vector from training input set:

- $e_{kp}(n)$ *: It is the error between target and actual output at the p_{th} element in the output vector at the n^{th} iteration.*
- $E_k(n)$ *: It is the Least Mean Square error between the whole target vector and the output vector corresponding to the k_{th} input vector at the n^{th} iteration.*
- $E_{av}(n)$ *: It is the target function whose minimum is interested. It is the batch learning error considering the errors corresponding to all the training input vectors at the n^{th} iteration.*

Due to the Delta rule, the weights learning obey relationship in Equation 4.4.

$$\left.\begin{aligned} \omega_{ij}(n+1) &= \omega_{ij}(n) + \Delta\omega_{ij}(n) = \omega_{ij}(n) - \varepsilon\frac{\partial E_{av}(n)}{\partial \omega_{ij}} \\ \omega_{jp}(n+1) &= \omega_{jp}(n) + \Delta\omega_{jp}(n) = \omega_{jp}(n) - \varepsilon\frac{\partial E_{av}(n)}{\partial \omega_{jp}} \end{aligned}\right\} \tag{4.4}$$

Equation 4.4 description:

- ω_{ij} *: It is the weight between the i_{th} input element and the j_{th} neuron in the hidden layer.*

- *ω_{jp}: It is the weight between the j_{th} neuron in hidden layer and the p_{th} neuron in the output layer.*
- *The whole Equation 4.4 reveals that the variation direction of the weights is towards the negative gradient and the step length is controlled by a parameter ε.*

Forward Calculation

For batch learning, the whole training input set is passed to the ANN in every epoch. At the n_{th} epoch, when the k_{th} input vector arrives at the input layer, forward calculation of the ANN model is shown in Equation 4.5

$$\left.\begin{array}{ll} u_{kj}(n)=\sum_{i=1}^{I}\omega_{ij}(n)x_{ki} & v_{kj}(n)=f_j(u_{kj}(n))=f_j(\sum_{i=1}^{I}\omega_{ij}(n)x_{ki}) \\ u_{kp}(n)=\sum_{j=1}^{J}\omega_{jp}(n)v_{kj}(n) & v_{kp}(n)=f_p(u_{kp}(n))=f_p(\sum_{j=1}^{J}\omega_{jp}(n)v_{kj}(n)) \\ & y_{kp}(n)=v_{kp}(n) \end{array}\right\} \quad (4.5)$$

Equation (4.5) description: this is the forward calculation corresponding to the k_{th} input vector at the n_{th} epoch.

- *x_{ki}: It is the i_{th} element in the input vector.*
- *$\omega_{ij}(n)$ is the weight between the i_{th} neuron in input layer i and the j_{th} neuron in hidden layer j; $\omega_{jp}(n)$ is the weight between the j_{th} neuron in hidden layer j and the p_{th} neuron in output layer p.*
- *$u_{kj}(n)$: It is the input of the j_{th} neuron in hidden layer j, which is achieved by the sum of all the output in the previous layer multiplied with their weights; $u_{kp}(n)$ is the input of the p_{th} neuron in output layer p.*
- *$v_{kj}(n)$: It is the output of the j_{th} neuron in hidden layer j; $v_{kp}(n)$ is the output of the p_{th} neuron in output layer p; $y_{kp}(n)$ is the p_{th} element in the output vector.*
- *$f_j()$: It is the activation function in the hidden layer j; $f_p()$ is the activation function in the output layer p.*

Error Propagation

To look backward from the output side in Fig 4.7, weights between the hidden layer j and the output layer p are firstly updated. Considering Equation (4.4), the variation of weights is determined by the negative gradient and a small length step controlling variable ε as Delta-rule revealed in Equation (4.6).

$$\Delta\omega_{jp}(n) = -\varepsilon \frac{\partial E_{av}(n)}{\partial \omega_{jp}} \tag{4.6}$$

Note Equation (4.3):

$$\left.\begin{aligned} \frac{\partial E_{av}(n)}{\partial \omega_{jp}} &= \frac{1}{N}\sum_{k=1}^{N}\frac{\partial E_k(n)}{\partial \omega_{jp}} \\ \frac{\partial E_k(n)}{\partial \omega_{jp}} &= \frac{\partial E_k(n)}{\partial e_{kp}} \cdot \frac{\partial e_{kp}(n)}{\partial y_{kp}} \cdot \frac{\partial y_{kp}(n)}{\partial u_{kp}} \cdot \frac{\partial u_{kp}(n)}{\partial \omega_{jp}} \end{aligned}\right\} \tag{4.7}$$

Consider Equation (4.7) with (4.3) and (4.5), the second equation in Equation (4.7) could be solved in Equation (4.8).

$$\frac{\partial E_k(n)}{\partial e_{kp}} = e_{kp}(n); \frac{\partial e_{kp}(n)}{\partial y_{kp}} = -1; \frac{\partial y_{kp}(n)}{\partial u_{kp}} = f_p'(u_{kp}(n)); \frac{\partial u_{kp}(n)}{\partial \omega_{jp}} = \mathrm{v}_{kj}(n) \tag{4.8}$$

In (4.8), $f_p'()$ is the differentiation of $f_p()$ in Equation (4.5). With Equation (4.7) and (4.8), the variation $\Delta\omega_{jp}(n)$ could be calculated:

$$\left.\begin{aligned} \frac{\partial E_k(n)}{\partial \omega_{jp}} &= -e_{kp}(n) \cdot f_p'(u_{kp}(n)) \cdot \mathrm{v}_{kj}(n) \\ \delta_{kp}(n) &= -e_{kp}(n) \cdot f_p'(u_{kp}(n)) \\ \frac{\partial E_{av}(n)}{\partial \omega_{jp}} &= \frac{1}{N}\sum_{k=1}^{N}\left\{\delta_{kp}(n) \cdot \mathrm{v}_{kj}(n)\right\} \\ \Delta\omega_{jp}(n) &= -\varepsilon \frac{\partial E_{av}(n)}{\partial \omega_{jp}} \end{aligned}\right\} \tag{4.9}$$

When weights between output layer p and hidden layer j finish their updates, propagation procedure move to the weights between hidden layer j and input layer i. The variation of weights is still from Equation (4.4)

$$\Delta\omega_{ij}(n) = -\varepsilon\frac{\partial E_{av}(n)}{\partial \omega_{ij}} \tag{4.10}$$

Note Equation (4.10) with (4.3) and (4.5):

$$\left.\begin{aligned} \frac{\partial E_{av}(n)}{\partial \omega_{ij}} &= \frac{1}{N}\sum_{k=1}^{N}\frac{\partial E_k(n)}{\partial \omega_{ij}} \\ \frac{\partial E_k(n)}{\partial \omega_{ij}} &= \frac{\partial E_k(n)}{\partial \nu_{kp}}\cdot\frac{\partial \nu_{kj}(n)}{\partial u_{kj}}\cdot\frac{\partial u_{ki}(n)}{\partial \omega_{ij}} = \frac{\partial E_k(n)}{\partial \nu_{kp}}\cdot f_j^{'}(u_{kj}(n))\cdot x_{ki} \end{aligned}\right\} \tag{4.11}$$

Consider Equations (4.3) and (4.11)

$$\frac{\partial E_k(n)}{\partial \nu_{kp}} = \sum_{p=1}^{P} e_{kp}(n)\cdot\frac{\partial e_{kp}(n)}{\partial \nu_{kp}} = \sum_{p=1}^{P}[e_{kp}(n)\cdot\frac{\partial e_{kp}(n)}{\partial u_{kp}}\cdot\frac{\partial u_{kp}(n)}{\partial \nu_{kp}}] \tag{4.12}$$

From Equation (4.5), the two partial differentiations in Equation (4.12) could be transformed into (4.13):

$$\left.\begin{aligned} \frac{\partial e_{kp}(n)}{\partial u_{kp}} &= -f_p^{'}(u_{kp}(n)) \\ \frac{\partial u_{kp}(n)}{\partial \nu_{kj}} &= \omega_{jp}(n) \end{aligned}\right\} \tag{4.13}$$

With Equations (4.11), (4.12), (4.13) and (4.9), the variation $\Delta\omega_{ij}(n)$ could be calculated in Equation (4.14).

$$\left.\begin{aligned} \frac{\partial E_k(n)}{\partial \omega_{ij}} &= \delta_{kj}(n)\cdot x_{ki} \\ \delta_{kj}(n) &= f_j^{'}(u_{kj}(n))\cdot\sum_{p=1}^{P}(\delta_{kp}(n)\cdot\omega_{jp}(n)) \\ \Delta\omega_{ij}(n) &= -\varepsilon\frac{\partial E_{av}(n)}{\partial \omega_{ij}} \end{aligned}\right\} \tag{4.14}$$

Batch Training Steps summary

With the weights update methods introduced above, training steps of BP batch training is summarized as below [81]:

Step 1. Initialize all the weights into non-zero random value.

Step 2. Use Forward Calculation to calculate the $E_{av}(n)$ in Equation (4.3). Compare $E_{av}(n)$ and the error acceptable limit, goal. If $E_{av}(n)$ is larger than the goal, turn to Step 3. Otherwise turn to Step 5.

Step 3. Compare the iteration number n to its limit. If n is larger, turn to Step 5. Otherwise turn to Step 4.

Step 4. Use Error Propagation to calculate all $\omega_{ij}(n+1)$ and $\omega_{jp}(n+1)$ to update weights. Then turn to Step 2.

Step 5. Output the trained network and finish training.

Fig 4.8 summarize the above five steps.

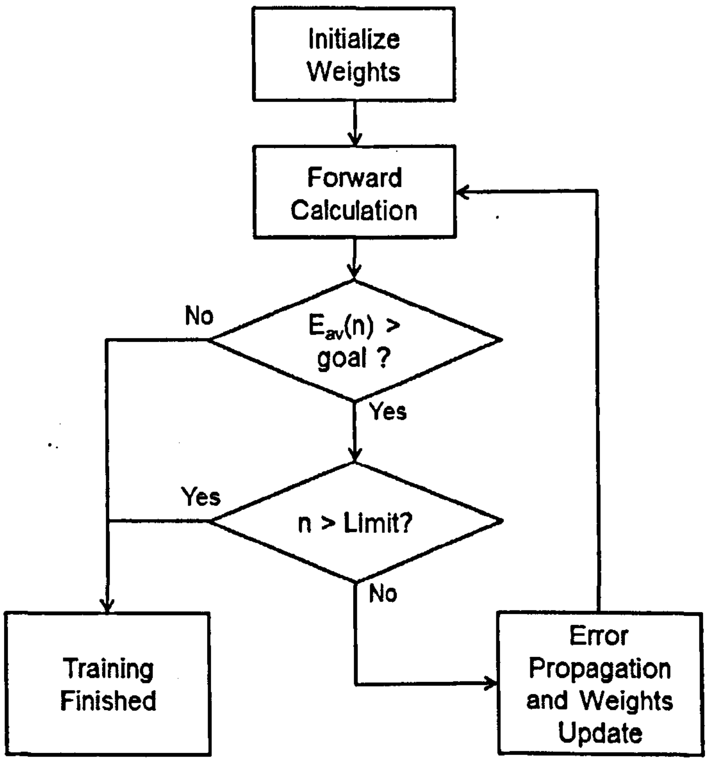


Fig 4.8: ANN Batch Training Step

4.4.4 Back-Propagation Trained ANN for Power Grid Load Forecasting

As previous sections mentioning, the hourly short-term load forecast problem in this chapter is the power demand forecast of Ontario province in Canada from Nov 11th, 2008 to Oct 31st, 2009. The ANN training data is from Nov 11th, 2005 to Oct 31st, 2008, which is 3 years data before the target period. 15% of training data is picked out as validation set for early stopping against over-fitting. The over-fitting problem will be discussed in section 4.6.4.

All the calculation will be performed in Matlab 7.11.0 (R2010b), with computer details listed in Table 4.2. Matlab is a programming environment for algorithm development, data analysis and numerical computation. With similar grammar as C++, Matlab has an affluent base function for various mathematical applications. Matlab will also integrate multiple toolboxes for specified area applications. Neural Network toolbox is the one used for load forecast in this thesis. It provides several typical neural networks including Multi-layer Perceptron. The toolbox provides a user friendly interface with easy-parameter-setting panel. The establishment and updates for ANN model is convenient. Moreover, once there is a requirement for a new ANN which is not included in the toolbox, the powerful mathematical based functions will help the user to build the model with coding easily. So though the Matlab may require more skill in programming, it is a good tool for ANNs in load forecast. For Multi-layer Perceptron, Chapter 4 deploys the ANN training with Multi-layer Perceptron Section in Matlab 7.11.0 (R2010b) Neural Network Toolbox.

Computer Hard Ware	Type
CPU	Intel(R) Core(TM)2 Quad CPU Q9650 @ 3.00GHz 2.99 GHz
Installed Memory	4.00 GB
Operating System	Windows 7 Ultimate Service Pack 1
Hard Disk	465 GB

Table 4.4: Computer details for ANN training

Architecture of Load Forecast System

In power system short-term load forecast, ANN achieves mapping between power demand and its influencing factors. The model assumption is listed below:

- Assume the influencing factors considered in Fig 4.9 are all the main factors affecting load pattern in the historical data for training and in the future forecast period.
- Assume the difference in data quality due to load change is acceptable between the past and in future forecast period

The above assumptions are necessary for ANN prediction as they promised the mapping stability between input and output space. Under these assumptions, the mapping learned by ANN from historical data could reflect the mapping in the future forecast period. But once the influencing factors changed or the mapping between influencing factors and the load changed, there will be unpredictable impacts occurred in the forecast accuracy.

Fig 4.9 reveals relevant influencing factors for power demand. As shown, the following factors are considered:

- **Weather Condition**: Weather status has a great impact on human comfort so that to impact on power devices selection and their utilization amount and time length.
- **Day Style**: People perform different lifestyle between in working days and in holidays.
- **Demand of the Previous Point**: Provide a reference for forecast.
- **Time Points Index**: Power consumption appears to be different at each specific hour in a day.

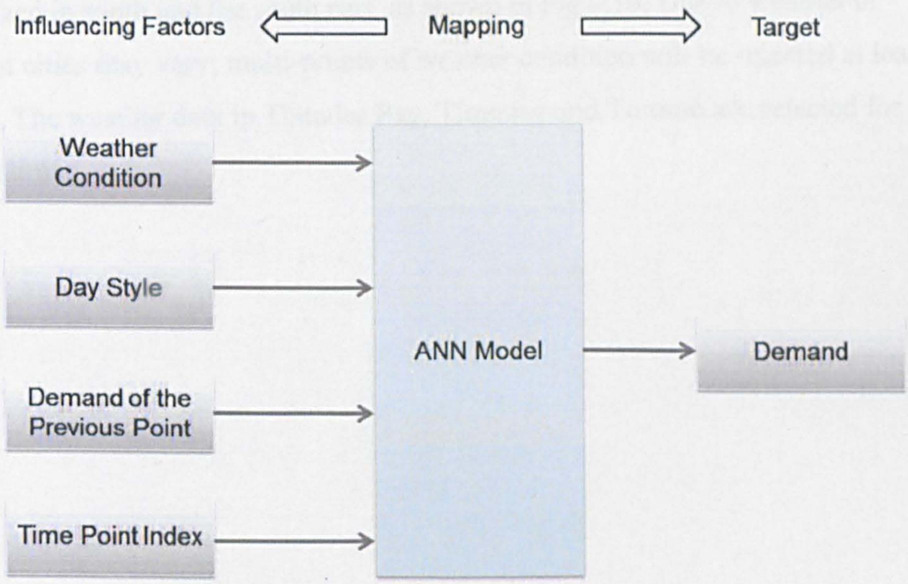


Fig 4.9: Influencing factors for power demand

In Weather Condition from Fig 4.9, human comfort is the main reason that leads to various demand requirements. There are 7 factors stimulus the human comfort [88]:

- Temperature
- Dew Point Temperature
- Relative Humidity
- Wind Speed
- Visibility

- Atmosphere Pressure
- Weather Status

In the 7 factors above, the first 6 factors are exact value variables directly related to human comfort. The last factor is a fuzzy data that influencing human's life style, which contains the following 6 indices:

- Clear Index: (0, 0.5, 1)
- Cloudy Index: (0, 0.5, 1)
- Fogy Index: (0, 0.5, 1)
- Rain Index: (0, 0.25, 0.5, 0.75, 1)
- Thunderstorms Index: (0, 0.5, 1)
- Snow Index: (0 0.3 0.6 1)

Each index is one of the values in the brackets above. They are value between 0 and 1, which describing how heavy the situation is.

Specific in Ontario Province, population and power consumption are mainly centralized in south and the south east, as shown in Fig 4.10. Due to weather in different cities may vary; multi-points of weather condition will be selected at load centres. The weather data in Thunder Bay, Timmins and Toronto are selected for ANN training.

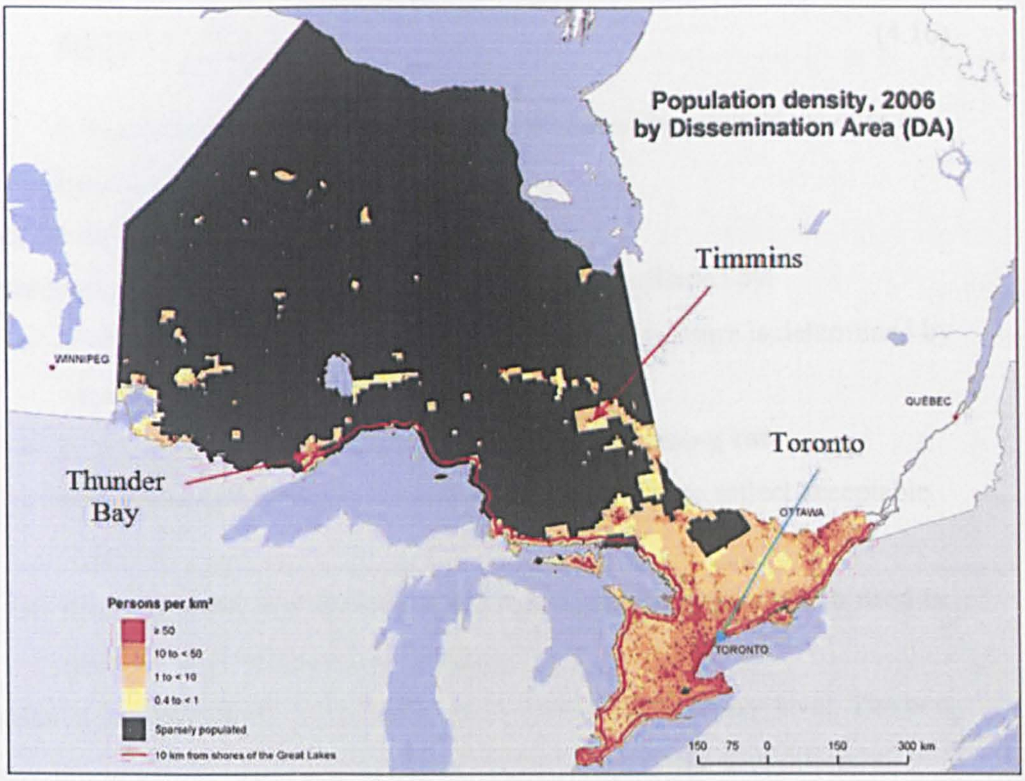


Fig 4.10: Canadian Population Distribution Map [88]

All the weather data is from [89].

Performance Measurement

Once networks are trained, indices reflecting the mapping approximation ability should be applied for performance measurement.

- **Root Mean Squared Error** (RMSE): Index quantifying difference between the prediction and the real demand by Equation (4.15). It has the same unit as the prediction output.

$$RMSE = \sqrt[2]{\frac{\sum_i^m (Out_i - Tar_i)^2}{m}} \tag{4.15}$$

  In Equation (4.15), m stands for total number of prediction outputs. Out_i *and* Tar_i *are the network output and the corresponding target to the* i^{th} *testing data.*

- **Mean Absolute Percentage Error** (MAPE): Index quantifying difference between the prediction and the real demand with reference percentage to the real demand. It is a percentage indicating how much percentage the difference is compare to the real demand. Equation (4.16) introduces the MAPE:

$$MAPE = \frac{1}{m}\sum_{i}^{m}\left|\frac{Tar_i - Out_i}{Tar_{mean}}\right| \tag{4.16}$$

In Equation (4.16), the variables have the same meaning as the ones in Equation (4.15).

ANN Training

Performance of ANN by Back-Propagation training is effected by:

- **ANN architecture:** For a 3-layer ANN, the architecture is determined by neurons number in the hidden layer.
- **Learning step length:** ε in Equation (4.4), is **learning rate.**
- **Error acceptable limit**: An index in training used to reflect acceptable solution.
- **Limit of batch training iteration:** n in Equation (4.4), is **epoch used to prevent** from an unlimited training.

For the best approximation, the goal is set to 0.0001 for accuracy level. The best epoch limit and learning rate are different between networks with different architecture. The ANN architecture and the learning step length should be optimized for the most suitable network selection. Table 4.5 reveals a process for ANN architecture selection.

Network Architecture (neurons in hidden layer)	Network Training Parameter	Training Performance
5	Goal: 0.0001;	Ave Train MSE: 0.0109
	Epoch Limit: 300	Ave CPU Time: 200.1066s
	Best Learning Rate: 0.05	Ave MAPE: 16.36%
10	Goal: 0.0001;	Ave Train MSE:0.0080
	Epoch Limit: 300	Ave CPU Time:261.4057s
	Best Learning Rate: 0.055	Ave MAPE: 13.99%
20	Goal: 0.0001;	Ave Train MSE:0.0059
	Epoch Limit: 300	Ave CPU Time:284.7701s
	Best Learning Rate: 1	Ave MAPE: 11.18%
50	Goal: 0.0001;	Ave Train MSE:0.0093
	Epoch Limit: 300	Ave CPU Time:420.6352s
	Best Learning Rate: 0.35	Ave MAPE: 11.97%

100	Goal: 0.0001;	Ave Train MSE:0.0091
	Epoch Limit: 300	Ave CPU Time:447.6930
	Best Learning Rate: 0.35	Ave MAPE: 11.15%

Table 4.5: ANN selection for BP training in load forecasting

Table 4.5 introduces 5 types of candidate networks whose differences are placed on the neurons' number in hidden layer. For each network, compare to the given goal and epoch limit, 10s of networks is investigated to looking for the best learning rate. After the best learning rate fixed, 10s of networks in the same type with different initial weights are trained by BP algorithm. The measurement is provided in the following three indices in the table:

- **Ave Train MSE**: The average value of all the Mean Squared Error performances of the same type networks corresponding to training data set.
- **Ave CPU Time**: The average CPU time spent by a certain type of networks.
- **Ave MAPE**: The average Mean Absolute Percentage Error achieved by a certain type of networks, corresponding to testing data set.

Fig 4.11 is another view on ANN selection of Table 4.5.

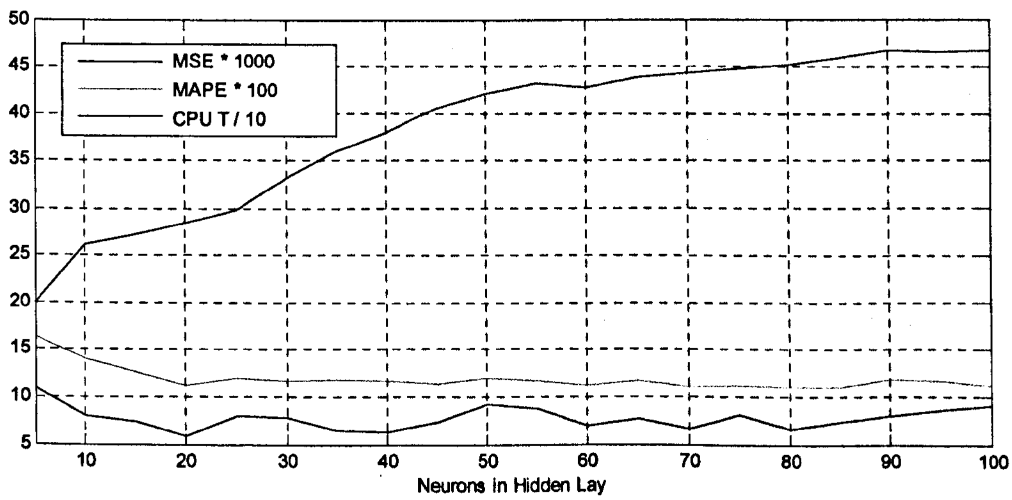


Fig 4.11: Measurement indices for ANN selection

In Fig 4.11, network performance, MSE and MAPE, decrease as increasing the neurons in hidden layer. As the neuron number is over 15, the performance appears to decrease not obviously. But the CPU time keeps increasing as more neurons in hidden layer increase the amount of calculation data. For achieving the best performance with less time, 20 neurons in hidden layer is selected. Table 4.6 provides result from an ANN load prediction by a network with 20 hidden neurons.

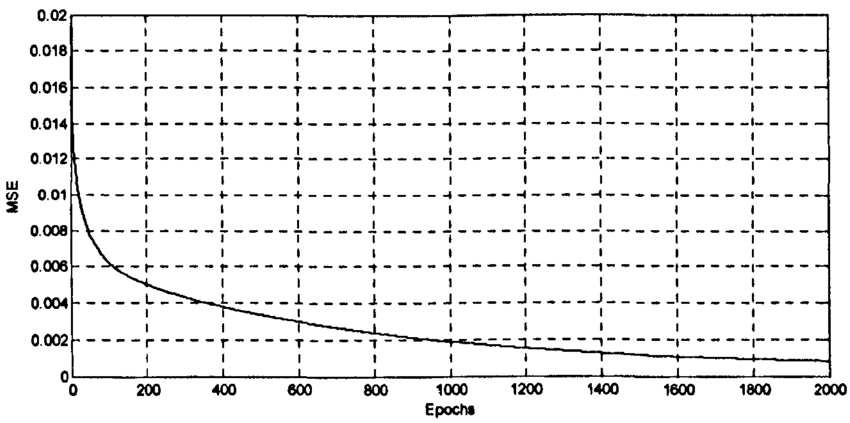
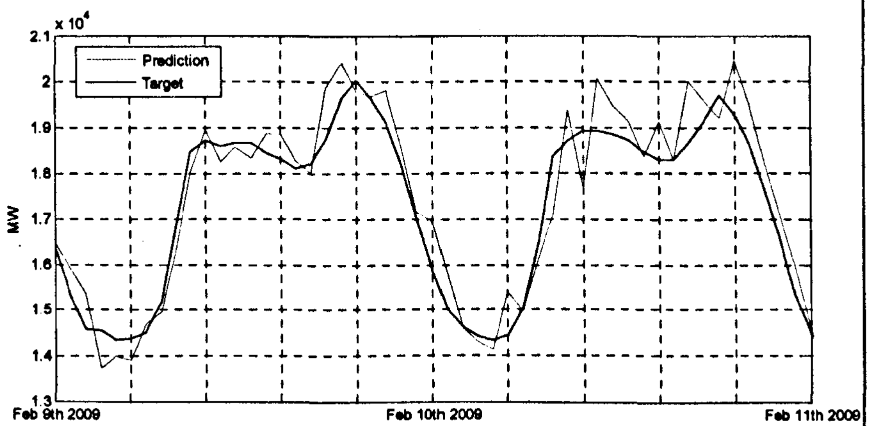
Network Architecture	Network Training Parameter	Training Performance
20	Goal: 0.0001;	Train MSE: 0.00081077
	Training Epoch: 2000	CPU Time: 1695.3s
	Best Learning Rate: 1	MAPE: 5.47%
Sample Training Process		
Sample Compare Between Target and Prediction of Test Data		

Table 4.6: Load forecasting by an ANN with 20 neurons in hidden layer

Table 4.6 reveals the training converging procedure. Back-propagation with delta-rule indeed figures out a way for ANN error minimization. Deploying with testing data, the predicted load mainly follows the trend of actual load in an MAPE 5.47%. But the CPU time is as long as 28.25 minutes.

4.5 Back-Propagation Training Improvement

Back-Propagation trained ANN has the capability to approximate the mapping between input space and output space by a long-enough training process. In some problems, there is requirement on training speed. A long training may lead to delay operation, so that decrease the response time and increase cost. Facing this situation, improvement on training speed appears.

4.5.1 Quasi-Newton Algorithm

Quasi-Newton algorithm is a method for function minimum or maximum based on stationary point of a function in optimization area. It is firstly introduced by W.C. Davidon at Argonne National Laboratory in 1959. The widespread modification of Quasi-Newton method is proposed independently by Broyden, Fletcher, Goldfarb, and Shannon in 1970, which is named BFGS method [90].

Consider an ANN network in Fig 4.7. Target function for the ANN is the error function $E_{av}(n)$ in Equation (4.3). For Quasi-Newton Algorithm, its expression could be re-write into Equation (4.17).

$$E_{av}(W_n) = E_{av}(\omega_{11}^{ij\,n}, \omega_{12}^{ij\,n}, \cdots, \omega_{IJ}^{ij\,n}, \omega_{11}^{jp\,n}, \omega_{12}^{jp\,n}, \cdots, \omega_{JP}^{jp\,n}) \tag{4.17}$$

Note from Equation (4.3) and (4.5), the target function is the function of all the weights between layers. So $E_{av}(n)$ could be rewrite into Equation (4.17). W_n is a vector whose elements are all the weights in ANN model. For each element of W_n, e.g. $\omega_{11}^{ij\,n}$, stands for the weight between the first neuron in the i^{th} layer and the first neuron in the j^{th} layer at the n^{th} iteration.

Expressed by Taylor Series, the target error is given in Equation (4.18):

$$\left.\begin{aligned} W_{n+1} &= W_n + \Delta W_n \\ E_{av}(W_n + \Delta W_n) &\approx E_{av}(W_n) + \nabla E_{av}(W_n)^T \cdot \Delta W_n + \frac{1}{2}\Delta W_n^{\,T} \cdot H_n \cdot \Delta W_n \end{aligned}\right\} \tag{4.18}$$

In equation 4.18,

- *H_n is the Hessian matrix of the target function.*
- *$\nabla E_{av}(W_n)$: It is the error gradient, which contains all the partial differentiation of each weight.*
- *ΔW_n: It is weights variation vector.*

Take the gradient of Equation (4.18). For achieving stationary point at $W_{n+1} = W_n + \Delta W_n$, the gradient of stationary point $E_{av}(W_n + \Delta W_n)$ is zero, the weights variation vector could be revealed in Equation (4.19) [91].

$$\left.\begin{aligned} 0 = \nabla E_{av}(W_n + \Delta W_n) &\approx \nabla E_{av}(W_n) + H_n \cdot \Delta W_n \\ \Delta W_n &= -H_n^{-1} \cdot \nabla E_{av}(W_n) \end{aligned}\right\} \tag{4.19}$$

The Quasi-Newton Algorithm in the ANN is trying to find the weights variance by the Hessian matrix and the target function gradient satisfying Equation (4.19).
As compared to the typical BP method introduced in section 4.4, the weights variation expression contains one more Hessian matrix. When complex calculation is processed for Hessian matrix, BFGS method offers an approximation calculation to the Hessian matrix, which is recognized a modification of Quasi-Newton Algorithm. Equation (4.20) introduces the Hessian matrix approximation by BFGS method [91].

$$\left.\begin{aligned} z_n &= \nabla E_{av}(W_{n+1}) - \nabla E_{av}(W_n) \\ H_{n+1} &= H_n + \frac{z_n \Delta W_n^{\;T}}{z_n^{\;T} \Delta W_n} - \frac{H_n \Delta W_n (H_n \Delta W_n)^T}{\Delta W_n^{\;T} \cdot H_n \cdot \Delta W_n} \end{aligned}\right\} \tag{4.20}$$

By Quasi-Newton Algorithm with BFGS modification in Equation (4.9) (4.14) (4.19) (4.20), the weights variance at each epochs could be calculated for training.

4.5.2 Levenberg-Marquardt (LM) Algorithm

The Levenberg-Marquardt Algorithm is firstly introduced by K. Levenberg in 1944 [92] and modified by D. Marquardt in 1963. It is a modification of Gauss-Newton method.

As required that the target function in Equation (4.3), $E_{av}(W_n) = \frac{1}{2N}\sum_{k=1}^{N}\sum_{p=1}^{P}(e_{kp}^n(W_n))^2$, is the sum of squares, LM method proposes another expression of gradient in Equation (4.21) [93].

$$\left.\begin{aligned} \frac{\partial E_{av}(W_n)}{\partial \omega_{ij}} &= \frac{1}{N}\sum_{k=1}^{N}\sum_{p=1}^{P} e_{kp}^n(W_n) \cdot \frac{\partial e_{kp}^n(W_n)}{\partial \omega_{ij}} \\ \nabla E_{av}(W_n) &= \frac{1}{N} J_e^{n\,T} \cdot \dot{e} \end{aligned}\right\} \tag{4.21}$$

In Equation (4.21):

- $e_{kp}^n(W_n)$ *: It is the re-write form of* $e_{kp}(n)$ *in Equation (4.3). k, p, n have the same meaning as in Equation (4.3).*
- $\dot{e}$ *: It is the error vector whose elements are all the* $e_{kp}^n(W_n)$.
- J_e^n *: It is the Jacobean matrix for the error vector* $\dot{e}$*, with respect to the weights vector* W_n *at the* n^{th} *iteration.*

- *For* $\frac{\partial E_{av}(W_n)}{\partial \omega_{jp}}$, *the expression only uses* ω_{jp} *to take the place of* ω_{ij}.

New approximation of Hessian matrix is also given in Equation (4.22) by ignoring second order partial derivative and adding a non-negative damping factor μ [91].

$$H_n \approx \frac{1}{N}(J_e^{n^T} \cdot J_e^n + \mu I) \tag{4.22}$$

In Equation (4.22), μ is the damping factor influencing the converging speed deeply. It changes until the step is a decrease step for target function.

Note Equations (4.19) (4.21) and (4.22), the weights variance of ANN by LM algorithm is given by Equation (4.23).

$$\Delta W_n = -\cdot (J_e^{n^T} \cdot J_e^n + \mu I) \cdot J_e^{n^T} \dot{e} \tag{4.23}$$

4.5.3 Load Forecasting by Improved BP Trained ANN

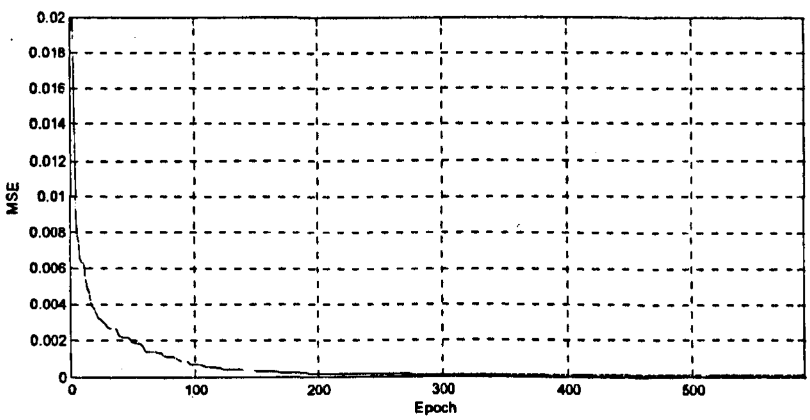
Quasi-Newton (BFGS) Training

Facing the same load forecasting problem in section 4.4, Quasi-Newton provides ANN architecture selection results in Table 4.7.

Network Architecture (neurons in hidden layer)	Average Training Performance	
5 Goal: 1×10^{-5}	Ave Training MSE: 1.23×10^{-4}	Ave Training CPU Time: 626.43s
	Ave MAPE: 1.55%	
10 Goal: 1×10^{-5}	Ave Training MSE: 1.03×10^{-4}	Ave Training CPU Time: 805.07
	Ave MAPE:1.53%	
20 Goal: 1×10^{-5}	Ave Training MSE: 0.87×10^{-4}	Ave Training CPU Time: 3056s
	Ave MAPE: 1.52%	
50 Goal: 1×10^{-5}	Ave Training MSE: 0.55×10^{-4}	Ave Training CPU Time: 66775s
	Ave MAPE:1.39%	

Table 4.7: ANN architecture selection for load forecast with Quasi-Newton (BFGS) Method

Titles in Table 4.7 are the same as that in Table 4.5. As networks' architecture increase, networks' MAPE and MSE decrease slowly, as compared to rapidly increasing calculation time. The network user should deeply consider the selection with their accuracy requirement, calculation time limit and the hardware limits. If the network could be trained less frequently than every 3 days or a high enough computer can be applied, the network with 50 neurons in hidden layer would be the best choice in this algorithm. If the network is trained hourly, then networks with 10 or 20 neurons in hidden layer would be the best choice. Table 4.8 provides results from a sample of network with 20 neurons in hidden layer by Quasi-Newton (BFGS) algorithm. Obviously from Table 4.8, performance of Quasi-Newton (BFGS) trained ANN is entirely better than delta-rule trained one. The higher accuracy is attracting but the only weak point is placed at the longer training time.

Network Architecture	Network Training Parameter	Training Performance
20	Goal: 1×10^{-5};	Train MSE: 7.74×10^{-5}
	Training Epoch: 591	CPU Time: 4716.3s
	Ave MAPE: 1.25%	
Sample Training Process		

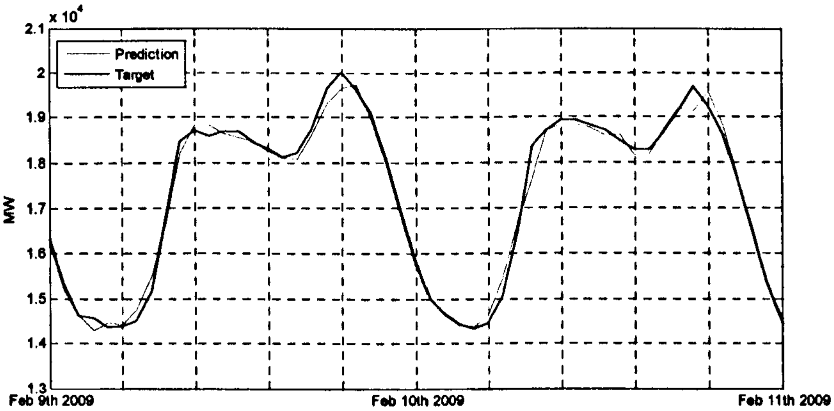
Sample Compare Between Target and Prediction of Test Data	

Table 4.8: Load forecast by ANN with 20 neurons in hidden layer with Quasi-Newton (BFGS) Algorithm

Levenberg-Marquardt Training

Facing the same load forecasting problem in Section 4.4, Levenberg-Marquardt method provides ANN architecture selection results in Table 4.9.

Network Architecture (neurons in hidden layer)	Average Training Performance	
5 Goal: 1×10^{-5}	Ave Training MSE: 0.549×10^{-4}	Ave Training CPU Time: 362.71s
	Ave MAPE: 1.08%	
10 Goal: 1×10^{-5}	Ave Training MSE: 0.403×10^{-4}	Ave Training CPU Time: 595.12s
	Ave MAPE:1.05%	
20 Goal: 1×10^{-5}	Ave Training MSE: 0.352×10^{-4}	Ave Training CPU Time: 646.12s
	Ave MAPE: 1.07%	
50 Goal: 1×10^{-5}	Ave Training MSE: 0.294×10^{-4}	Ave Training CPU Time: 1153.48s
	Ave MAPE:1.09%	

Table 4.9: ANN architecture selection for load forecast with LM Method

Table 4.9 reveals that Levenberg-Marquardt method is the best training algorithm in the three algorithms. It generally achieves the best performance with a not too long calculation time. Table 4.10 provides results from a sample of network with 10 neurons in hidden layer by Levenberg-Marquardt methods. From Table 4.10, LM trained ANN not only improves the accuracy but also less training epoch and training time comparing to Delta-rule training and Quasi-Newton (BFGS) training.

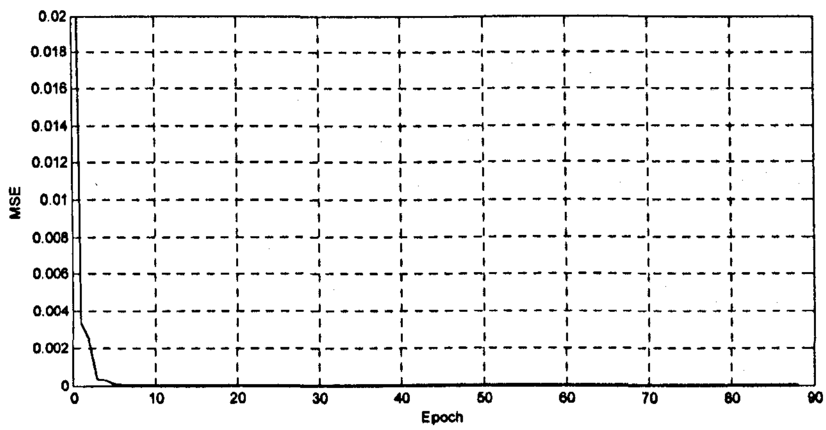
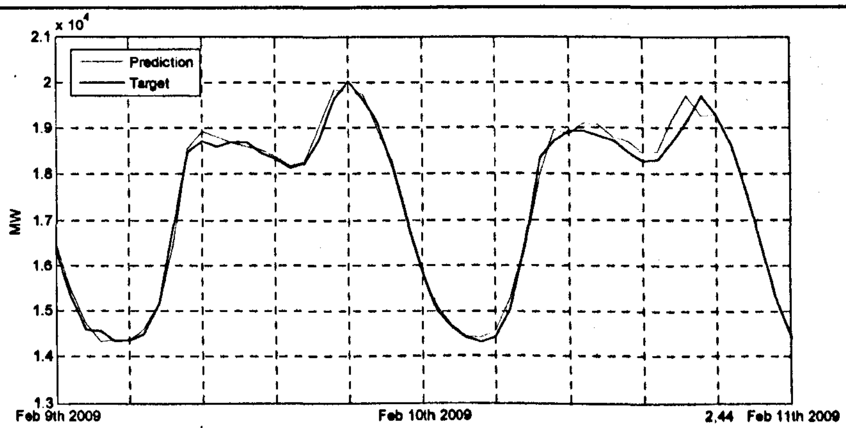
<table>
<tr><th>Network Architecture</th><th>Network Training Parameter</th><th>Training Performance</th></tr>
<tr><td rowspan="3">20</td><td>Goal: 1×10^{-5};</td><td>Train MSE: 3.815×10^{-5}</td></tr>
<tr><td>Training Epoch: 88</td><td>CPU Time: 398.5s</td></tr>
<tr><td colspan="2">Ave MAPE: 1%</td></tr>
<tr><td>Sample Training Process</td><td colspan="2">
</td></tr>
<tr><td>Sample Compare Between Target and Prediction of Test Data</td><td colspan="2">
</td></tr>
</table>

Table 4.10: Load forecast by ANN with 20 neurons in hidden layer with Levenberg-Marquardt Algorithm

4.6 Simulation Analysis

4.6.1 ANN Training Analysis

A sample ANN with 10 neurons in hidden layer is selected for compare the training effect between typical BP algorithm, Quasi-Newton (BFGS) algorithm and

Levenberg-Marquardt algorithm. Fig 4.12 shows the first 50 step of their training procedure.

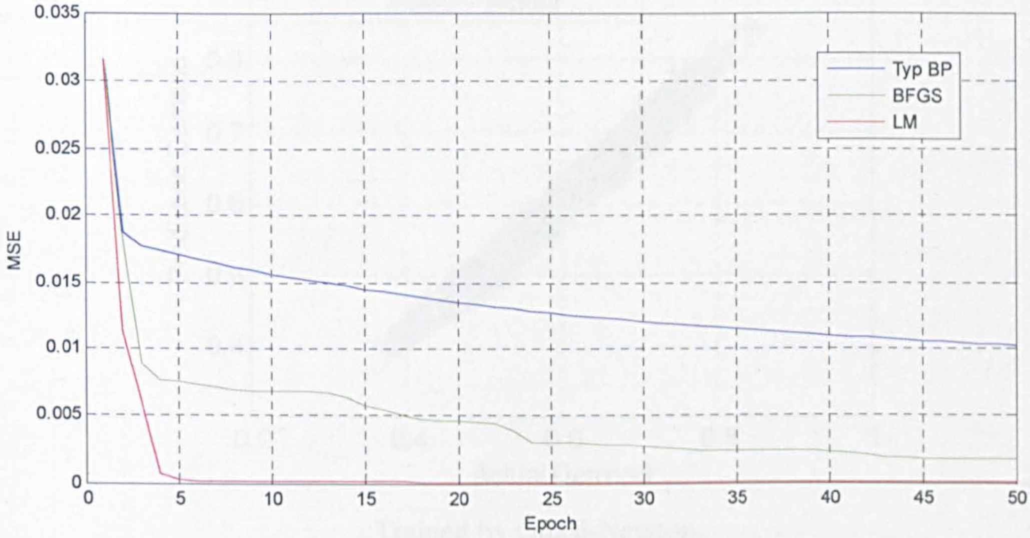
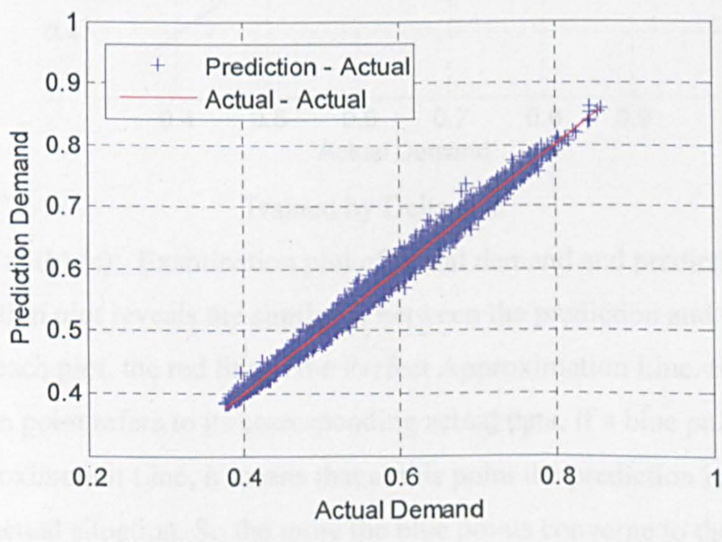


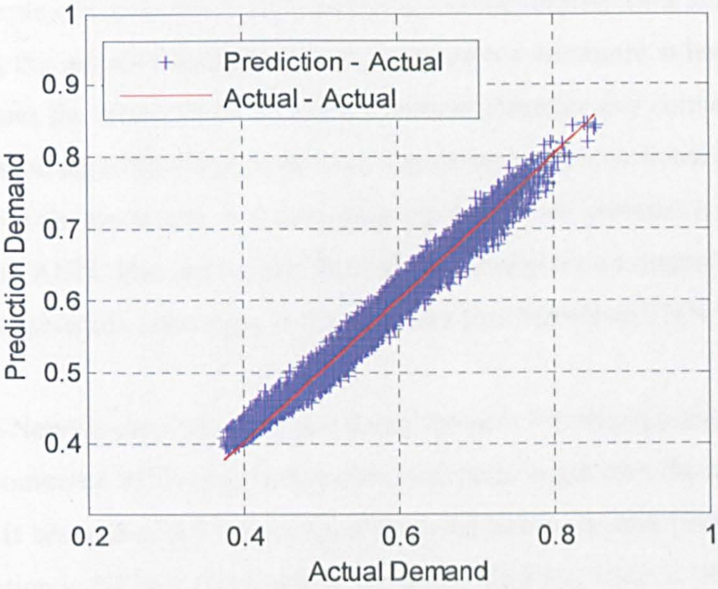
Fig 4.12: The first 50 steps of training with 3 different methods on a sample ANN

The Quasi-Newton algorithm converges faster than the pure Delta-rule in typical BP algorithm.

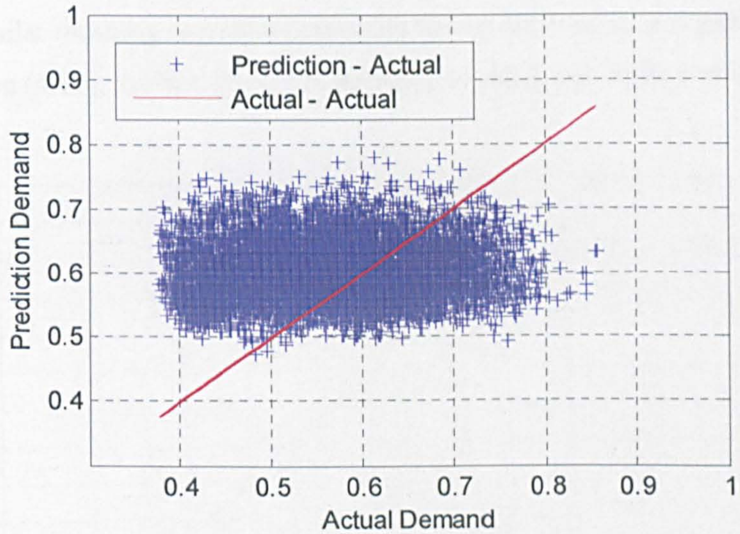
The converging ability could also be revealed by examination plot, which similar to regression plot, between actual demand and prediction demand. Fig 4.13 (a) (b) (c) has separately shown the plots of a 5-hidden neurons sample network trained by BP, Quasi-Newton (BFGS) and LM.



Trained by LM



Trained by Quasi-Newton



Trained by Delta rule

Fig 4.13 (a) (b) (c): Examination plot of actual demand and prediction demand

The examination plot reveals the similarity between the prediction and the actual situation. In each plot, the red line is the Perfect Approximation Line. Each blue point is a prediction point refers to its corresponding actual data. If a blue point is on the Perfect Approximation Line, it means that at this point the prediction is exactly the same as the actual situation. So the more the blue points converge to the red line, the better the network performance is. Fig 4.13 reveals that the LM algorithm achieves the best converging ability for its blue points are generally closer to the red line than the other two. This situation is caused by the converging force difference.

From the principle of Delta-rule in BP algorithm, the converging force is from the idea 'following the negative gradient will anyway reach a minimum at last'. This force only assures the arrival to the minimum without directing any converging path. But Quasi-Newton algorithm is derived from Taylor series with stationary point. The converging force always targets to the stationary point, which provides a converging path direction to ANN. The converging force differences place an impact that training with Delta-rule generally converges in a detour like path than Quasi-Newton algorithm.

Though Quasi-Newton generally selects a straighter path towards minimum than Delta-rule, it sometime still selects a direction with large angle than the correct one. This situation is because of the instability of inversed hessian matrix [90]. Suppose the target function is the blue line in Fig 4.14. Target function input is Wn. At the point shown in Fig 4.14, the gradient of function is negative but the Hessian matrix, which has similar meaning as partial derivative in one dimension, is negative, too. From Equation (4.19), the Wn change is positive, which is not in the correct direction [81].

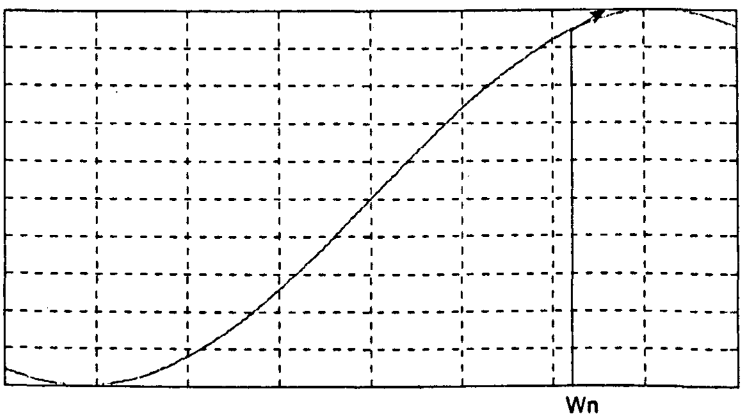


Fig 4.14: Example of instability of Quasi-Newton algorithm [81]

The Levenberg-Marquardt algorithm has solved this situation. The selection of damping factor has ensured all the steps to be downward steps. So it has the best converging ability in the three methods.

4.6.2 Performance Analysis

Mean Absolute Percentage Error (MAPE) is one of the indices to measure the network, which is shown in all the previous training result. It summarizes how far the prediction is from real power demand averagely by percentage. But the utilities may also prefer to have a smooth error distribution, for one thing, over-large error may cause huge lost and even damage.

ANN Hidden Layer Neurons (BFGS)	Performance Indices	
05	Ave MAPE (%): 1.55	Ave Largest (%): 11.34
	Ave APE STD (%): 1.21	Ave Largest (MW): 1470
10	Ave MAPE (%):1.53	Ave Largest (%): 12.49
	Ave APE STD (%):1.28	Ave Largest (MW): 1450
20	Ave MAPE (%):1.52	Ave Largest (%): 14.61
	Ave APE STD (%):1.32	Ave Largest (MW): 1665
50	Ave MAPE (%):1.39	Ave Largest (%): 15.23
	Ave APE STD (%):1.34	Ave Largest (MW): 1784

Table 4.11: Quasi-Newton (BFGS) trained ANN performances compare

Table 4.11 provides a performance of different ANN architecture by Quasi-Newton (BFGS) algorithm as example for performance analysis. In each type of architecture, 5 ANNs are selected to calculate the average performance. Titles description in Table 4.11 is following:

- **Ave MAPE:** The average Mean Absolute Percentage Error. It is the mean value of all the MAPE from the 5 ANNs, which indicate the performance of a certain network type.
- **Ave Largest (%)**: The average largest percentage error. It is the mean value of all the largest percentage error from the 5 ANNs, which indicate the largest percentage error of a certain network type.
- **Ave Largest (MW)**: The average largest absolute error. It is the mean value of all the largest absolute error from the 5 ANNs, which indicate the largest absolute error of a certain network type.
- **Ave APE STD**: Standard deviation for absolute percentage error (APE) is an index for measure the variation degree of the APE. The Ave APE STD is the mean value of all the APE standard deviation from the 5 ANNs, which indicate the error variation of a certain network type.

From Table 4.11, as the complexity of ANN architecture increases, MAPE decreases for the ability of ANN's approximation increases. But the other three indices for error variation increase, this indicates the increase of ANN complexity will also increase the instability of performance. This is because the higher complexity is the more capability for ANN to store redundancy. The redundancy is produced by that the training set does not perfectly represent the total set sufficiently and uniformly.

4.6.3 Calculation Time Analysis

Time limit for system calculation is a general requirement for short-term load forecasting. A over time consumed system will decrease the time in other coming work, e.g. planning.

For a 3 layers ANN whose hidden layer contains less than 100 neurons, each forward calculation consumes less than 0.1 second by the computer states in Table 4.4. So the main issue is placed at time limit on ANN training.

Training CPU Time Analysis

ANN training is basically constructed by two components:

- **Epoch Calculation Period**: Time cost in each epoch calculations. In Equation (4.24) is T_{epoch}.
- **Training Epoch Quantity**: Numbers of the epochs that required by training. In Equation (4.24) is Q_{epoch}.

$$T_{CPUT} = T_{epoch} \times Q_{epoch} \tag{4.24}$$

For Training Epoch Quantity, Fig 4.12 reveals that LM consumes the least epochs to a certain accuracy target. Quasi-Newton (BFGS) is the second and Delta-rule by BP training cost the most epochs for a target as it converge in a detour path mentioned in Section 4.6.1.

For Epoch Calculation Period, Table 4.12 offers a summary of Epoch Calculation Period by time consumption per epoch.

	Hidden Layer Neurons: 5	Hidden Layer Neurons: 10	Hidden Layer Neurons: 20
Delta-Rule	0.6670s	0.8714s	0.9492s
Quasi-Newton	1.8562s	2.6648s	8.0742s
LM	2.4238s	4.5312s	10.3068s

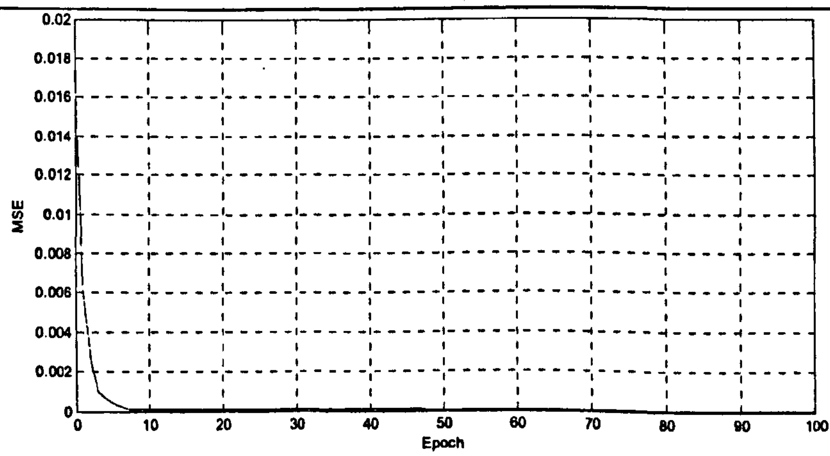
Table 4.12: Summary of ANN time consumption per epoch

From Table 4.12, as the complexity of ANN increases, the time consumption per epoch increase for more neurons in hidden layer will introduce more weights. To compare result by different methods, LM and Quasi-Newton algorithm have to calculate the approximation of hessian matrix, and make their time consumption per epoch much longer than that of Delta-rule. Delta-rule only needs to calculate the gradient. In LM, the Jacobean matrix calculation needs more time at each epoch.

Training Time Limit of Different Learning Structure

System of Pattern Recognition highly depends on the training set. On power system load forecasting, weather condition is one of the main training parameters. Due to weather may change as the time goes by, an off-line learning ANN system will anyway become invalid some day. A system that keeps trained by newest practical data could extent the working lifestyle. But once on-line learning is selected, there is a limit that the training time should not longer than the given period between two time-points that system updates. And it will influence the final ANN selection for load forecasting.

If the on-line learning plan is to train the system daily or even longer, the system will have enough time for training and planning. In this situation, system tends to find out a single ANN with the best performance. Table 4.13 shows the best trained ANN found in this load forecasting project.

<table>
<tr><th>Network Architecture</th><th>Network Training Parameter</th><th>Training Performance</th></tr>
<tr><td rowspan="3">05</td><td>Largest Error: 9.2%;</td><td>Train MSE: 5.20×10^{-5}</td></tr>
<tr><td>Training Epoch: 128</td><td>CPU Time: 303.95</td></tr>
<tr><td colspan="2">MAPE: 0.96%</td></tr>
<tr><td>Sample Training Process</td><td colspan="2">
</td></tr>
</table>

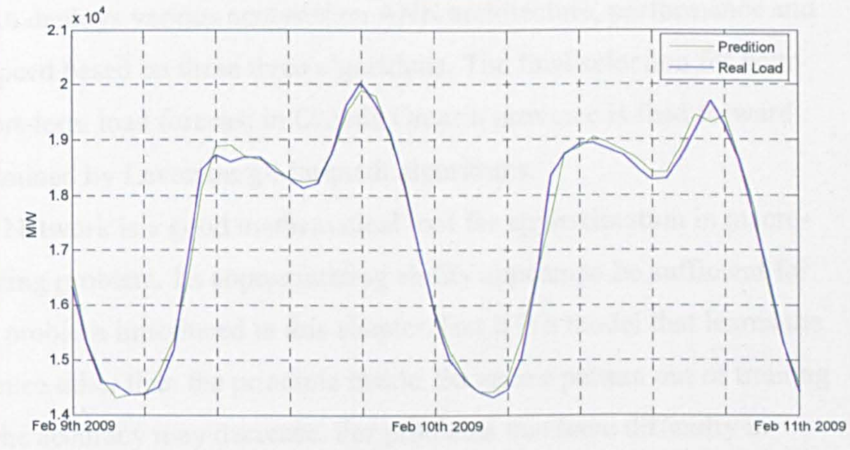
Sample Compare Between Target and Prediction of Test Data	

Table 4.13: Best performance ANN found in this load forecasting

If on-line training is hourly, it requires not only the time for training is less than 30 minutes but also tend to select networks by their general training effect for there is not enough time to create a large enough candidate ANN set. This situation will be more focus on the general training effect, as shown in Table 4.9. In the load forecasting problem considered in this thesis, a network type that contains 10 hidden layer neurons ANN with training algorithm Levenberg-Marquardt algorithm is selected.

4.7 Conclusion

Artificial Neural Network performs an excellent mapping between input space and output space. This chapter has applied this model into short-term load forecasting with an ANN based load forecasting framework.

- Section 4.3 introduces the principle theory of ANN and its training algorithm.
- Section 4.4 sketches the macro-grid load forecast problem on Canada Ontario province. Figure 4.9 reveals a black-box model for problem solving. The black-box input includes Weather Conditions, Day Style, Demand of the previous point and Time Point Index. The black-box is approximated by Delta-rule trained ANN. The best MAPE is 5.47%.
- Section 4.5 introduces two other training methods for ANN: Quasi-Newton method and Levemberg-Marquadt algorithm. These two methods achieve better performance than Delta-rule training. The best accuracy form Quasi-Newton method is 1.25% while LM achieves 0.96%.

- Section 4.6 deploys various analyses on ANN architecture, performance and training speed based on these three algorithms. The final selection for hour-ahead short-term load forecast in Canada Ontario province is feed forward network trained by Levemberg-Marquadt algorithms.

Artificial Neural Network is a good mathematical tool for approximation in macro-grid load forecasting problem. Its approximating ability appears to be sufficient for the load forecast problem introduced in this chapter. But it is a model that learns the mapping appearance other than the principle inside. So once a pattern out of training data set occurs, the accuracy may decrease. For problems that have difficulty in figuring out exact principles, like load forecast, ANN will be a good choice. But for problems which described well in maths or with high price for risk, ANN selection should be considered.

Chapter 5

Artificial Neural Network in Load Forecasting of Micro-grid

5.1 Micro-grid Load Forecasting

5.1.1 Micro-grid

Micro-grid is born for the new development of Distributed Generation and the Self-sufficient Concepts in Smart Grid. There is still not a standard definition for Micro-grid. Generally compare to traditional power grid, micro-grid is a localized and small-scale power system that integrate multi electricity generation types, energy storage types, several specific loads and a specific designed control & management system [94].

Micro-grid achieves a high potential in future development for the coming decade for several reasons. Distributed Generations, including wind energy, fuel cells, hydropower, biomass energy, is promoted strongly for their advantages on relieving transmission load in traditional power grid with lower carbon emission [95]. Micro-grid as the best platform for integrating and managing Distributed Generation is attracting attention from public and governments worldwide. Reliability is another reason as reliability of the centralized traditional power supply may be no longer . suitable for future applications. Micro-grid possesses the function to disconnect from large grid when disturbances occur, so as to protect the reliability and security of power grid [96]. Moreover, compatibility of new applications promotes micro-grid development as well. E.g. the application of Electric Vehicle has raised the burden of traditional power transmission and distribution as the application brings energy from grid to vehicle instead of from fuel to vehicle. Micro-grid and its Distributed Generation helps traditional grid to handle this application so as preventing the heavy work in transmission conjunction management. Fig 5.1 introduces an example of micro-grid.

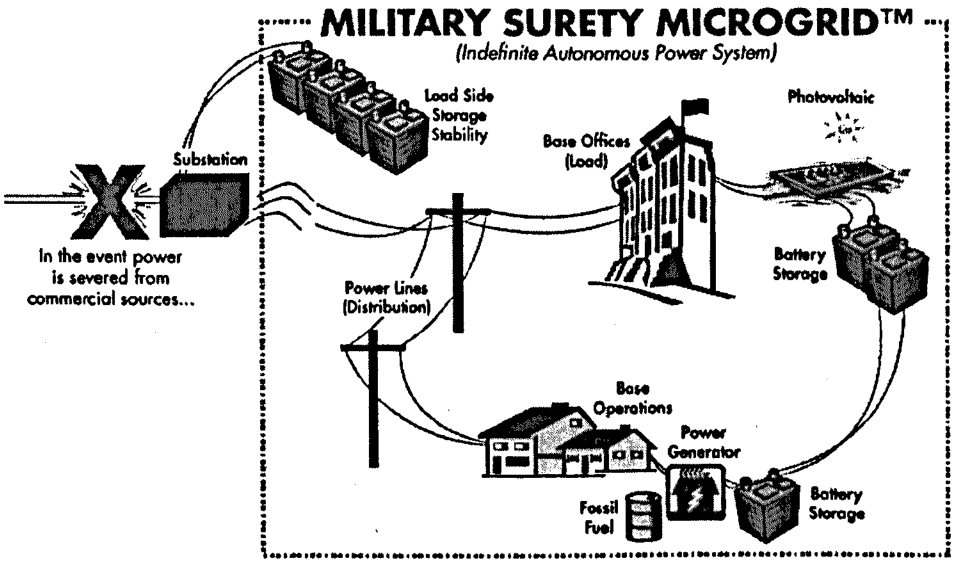


Fig 5.1: Sample of Microgrid [97]

5.1.2 Micro-grid Load Forecasting

To achieve an optimal generation planning and the best utilization of multi-types applications, it is necessary to find out the micro-grid demand. Short-term load forecasting is basic information for micro-grid management and the trade of electric energy with the grid [96]. Accurate load information may help the micro-grid manager to arrange their assets utilization and maintenance. Moreover, utilities may benefit from better reliability and support in decision making from micro-grid load forecast. However, as compared to the demand in large grid, micro-grid demand appears to be volatile and high variations. This is because the localization of micro-grid brings local consumption feature. Randomly occurring incidents may influence the demand of micro-grid. In this Chapter, a demand of a Micro-grid chiller-system is predicted with a new designed Micro-grid load forecast framework. The chillers system is installed as a core section of building energy management system in a typical University in Hong Kong, providing air conditioning for the whole university usage. The requirement is to achieve a chiller demand forecasting every half hour. All the calculation will be performed in Matlab 7.11.0 (R2010b), with a computer details same as that in chapter 4.

5.2 Back-Propagation Trained ANN for Micro-grid Load Forecasting

5.2.1 System Design

Main assumptions are listed below:

- Assume the influencing factors considered in Fig 5.2 are the main factors affecting load pattern in the historical data for training and in the future forecast period.
- Assume the difference in data quality due to load change is acceptable between the past and in future forecast period

Same as macro-grid load forecast, the above assumptions are necessary for ANN prediction in micro-grid load forecast as ANN needs the promise of the mapping stability between input space and output space. Under these assumptions, the mapping learned by ANN from historical data could reflect the mapping in the future forecast period. But in micro-grid, some unpredictable randomly factors take more proportion than in macro-grid, though they are still not the main influencing factors. The impact of these factors will be revealed in section 5.4.4.

The demand of chillers is highly depended on humans' behaviour and feelings. The main influencing factors following:

- **Weather conditions**: The variation of weather conditions is the original promotion to the utilization of chillers system. So different weather conditions definitely influence the system demand.
- **Human lifestyle in one day**: Human tends to work in the daytime and rest at night. At **each hour time point** of a day, people target at different work, so as manipulating the demand.
- **Day style**: Demand in working days and in weekend is totally different, due to the lifestyle.
- **Calendar period**: This is a university specialized factors as university organizes their staffs and students into different activities in diverse calendar period.
- **Load of Previous Time Point**: Provide a reference for forecast.

With the factors the prediction system is designed in Fig 5.2 .

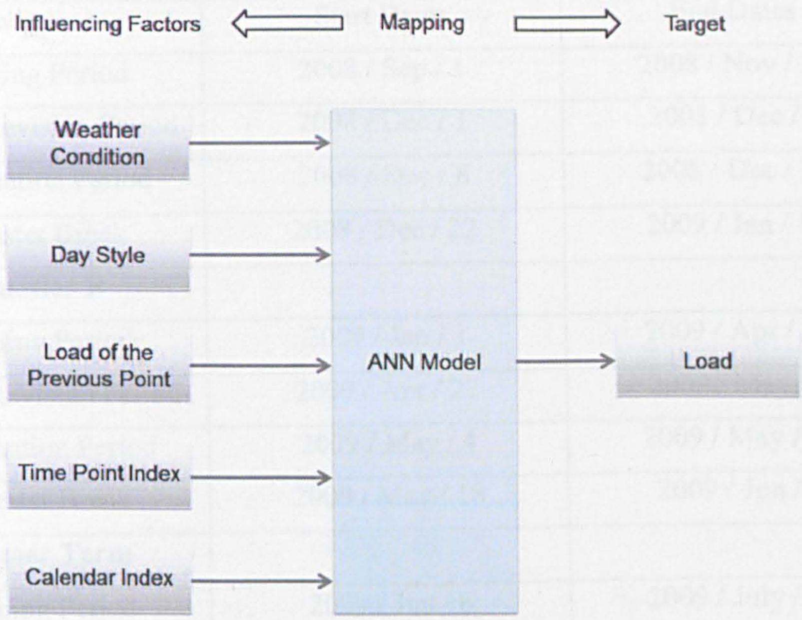


Fig 5.2: Chillers system demand prediction structure

In Weather Condition from Fig 5.2, human comfort is the main reason that leads to various demand requirements. There are 9 factors stimulus the human comfort [88];

- Temperature
- Dew Point Temperature
- Relative Humidity
- Wind Speed
- Visibility
- Atmosphere Pressure
- Weather Status

Same as chapter 4, Weather Status is fuzzy indices with the same structure.

University lifestyle is deeply related to its calendar. The specified university has 4 periods in one term [98]:

- Teaching Period.
- Student Revision Period.
- Examination Period.
- Semester Break Period.

Table 5.1 introduce its calendar of 2008-2009:

Semester A	Start Dates	End Dates
Teaching Period	2008 / Sep / 1	2008 / Nov / 29
Student Revision Period	2008 / Dec / 1	2008 / Dec / 6
Examination Period	2008 / Dec / 8	2008 / Dec / 20
Semester Break	2008 / Dec / 22	2009 / Jan / 10
Semester B		
Teaching Period	2009 / Jan / 1	2009 / Apr / 25
Student Revision Period	2009 / Apr / 27	2009 / May / 2
Examination Period	2009 / May / 4	2009 / May / 16
Semester Break	2009 / May / 18	2009 / Jun / 6
Summer Term		
Teaching Period	2009 / Jun / 8	2009 / July / 25
Student Revision Period	2009 / July / 27	2009 / Aug / 1
Examination Period	2009 / Aug / 3	2009 / Aug / 8
Semester Break	2009 / Aug / 10	2009 / Aug / 29

Table 5.1: 2008-2009 Academic Calendars [98]

Each period will be identified by a logic variable, which 1 represent the calendar period status and 0 for opposite.

5.2.2 BP Trained ANN Training and Performance

Back-Propagation trained ANN and the training improvement is introduced in Chapter 4 with its theory and application in power grid load forecasting. In this micro-grid demand prediction case, the Back-Propagation trained feed-forward network is still selected for prediction.

Typical BP Training

To achieve the best performance, network architecture selection is necessary. Table 5.2 reveals the performance of different ANN architecture.

Network Architecture (neurons in hidden layer)	Network Training Parameter	Training Performance
5	Goal: 0.0001;	Ave Train MSE: 0.0042
	Ave Epoch: 1000	Ave CPU Time: 501.73s
	Best Learning Rate: 0.8	Ave MAPE: 43.04%
10	Goal: 0.0001;	Ave Train MSE:0.0041
	Ave Epoch: 1000	Ave CPU Time:572.38s
	Best Learning Rate: 0.8	Ave MAPE: 48.27%
20	Goal: 0.0001;	Ave Train MSE:0.0040
	Ave Epoch: 1000	Ave CPU Time:903.34s
	Best Learning Rate: 1	Ave MAPE: 44.70%

Table 5.2: Network architecture selection for BP training

Table 5.2 introduces 3 types of candidate networks whose differences are placed on the neurons' number in hidden layer. For each network, compare to the given goal and epoch limit, 10s of networks is investigated. After the best learning rate fixed, 10s of networks in the same type with different initial weights are trained by BP algorithm. The measurement is provided in the following three indices in the table:

- **Ave Train MSE**: The average value of all the Mean Squared Error performances of the same type networks corresponding to training data set.
- **Ave CPU Time**: The average CPU time spent by the certain type of networks.
- **Ave MAPE**: The average Mean Absolute Percentage Error achieved by the certain type of networks corresponding to testing data set.

Fig 5.3 reveals the performance of different architecture network.

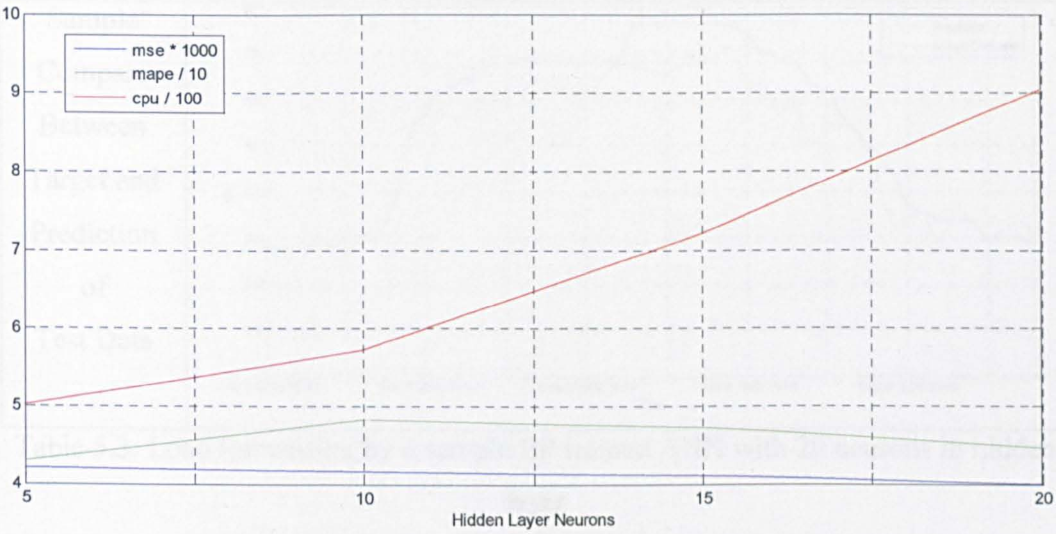
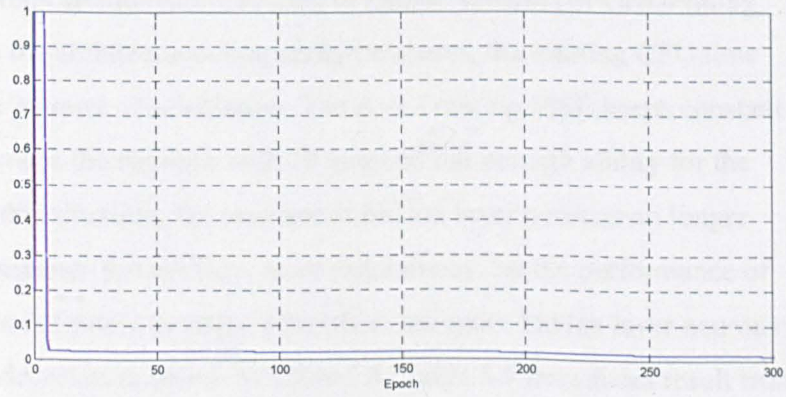


Fig 5.3: Performance of different architecture network

From Fig 5.3, as there are more neurons in hidden layer, it leads to larger data set for calculation. So the CPU time increases with the addition of hidden layer neurons. And the more complex the architecture is, the more ability for the network to approximate to the training set, so the Mean Square Error decreases. But the Mean Absolute Percentage Error (MAPE) of the testing data is not definitely decrease. This is due to the generalization problem and noise of micro-grid. This situation will be discussed in the analysis section. Table 5.3 introduces result from ANN load prediction by a network with 20 hidden neurons.

Network Architecture	Network Training Parameter	Training Performance
20	Goal: 0.0001;	Train MSE: 0.0047
	Training Epoch: 1000	CPU Time: 903.4s
	Best Learning Rate: 1	MAPE: 42.26%
Sample Training Process		

Sample Compare Between Target and Prediction of Test Data	

Table 5.3: Load forecasting by a sample BP trained ANN with 20 neurons in hidden layer

Quasi-Newton (BFGS) Training

Details of Quasi-Newton (BFGS) algorithm is introduced in Section 4.5.1. The architecture selection of Quasi-Newton (BFGS) training is in Table 5.4:

Network Architecture (neurons in hidden layer)	Average Training Performance	
5 (Ave Epochs: 164)	Ave Training MSE: 0.0028	Ave Training CPU Time: 294.53s
	Ave MAPE: 14.01%	
10 (Ave Epochs: 148)	Ave Training MSE: 0.0028	Ave Training CPU Time: 386.57
	Ave MAPE:13.96%	
20 (Ave Epochs: 147)	Ave Training MSE: 0.0028	Ave Training CPU Time: 1225.6s
	Ave MAPE: 14.25%	

Table 5.4: ANN architecture selection of Quasi-Newton (BFGS) training

From Table 5.4, as the architecture complexity increases, the training CPU time increases for larger amount of calculation. The Ave Training MSE keeps constant as complexity increases as the network with 10 neurons has enough ability for the approximation. In this situation, the increase of hidden layer neurons no longer improves the performance but contains more redundancy. So the performance of testing data will not definitely increase when there are more hidden layer neurons; in fact, it could even decrease as shown in Table 5.4. Table 5.5 introduces result from

Quasi-Newton (BFGS) trained ANN load prediction by a network with 10 hidden neurons.

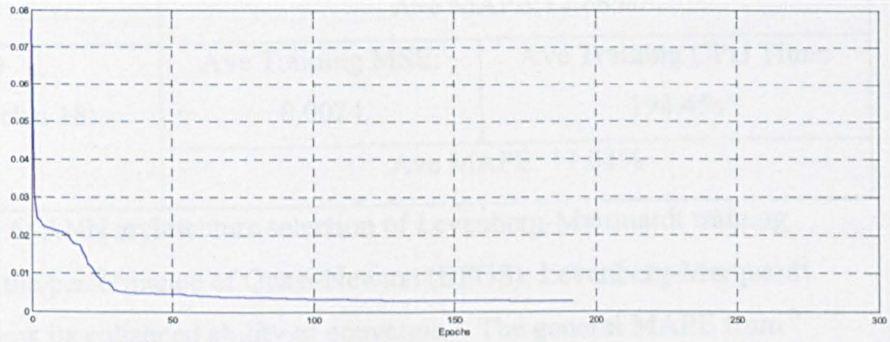
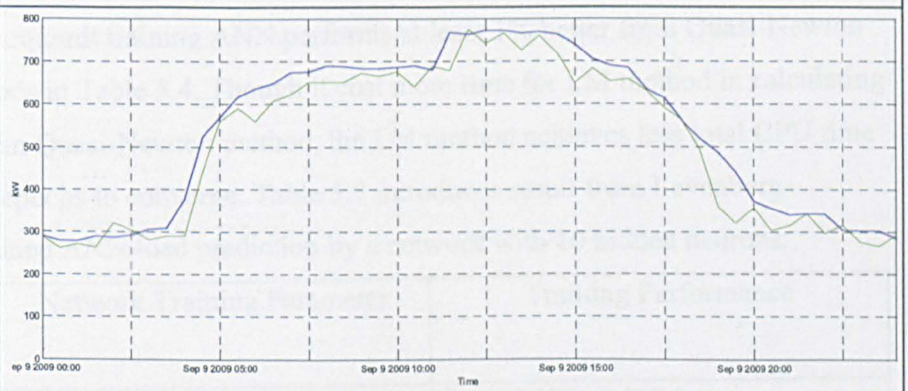
Network Arch	Network Training Parameter	Training Performance
10	Goal: 0.0001;	Train MSE: 0.0028
	Training Epoch: 192	CPU Time: 499.8s
	MAPE: 13.29%	
Sample Training Process		
Sample Compare Between Target and Prediction of Test Data		

Table 5.5: Load forecasting by a sample Quasi-Newton (BFGS) trained ANN with 10 neurons in hidden layer

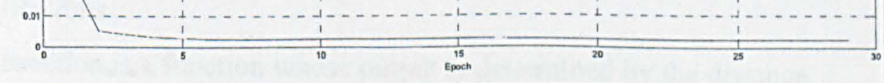
Levenberg Marquardt Training

Details of Levenberg Marquardt algorithm is introduced in Section 4.5.2. The architecture selection is shown in Table 5.6:

Network Architecture (neurons in hidden layer)	Average Training Performance	
5 (Ave Epochs: 86)	Ave Training MSE: 0.0027	Ave Training CPU Time: 176.64s
	Ave MAPE: 13.23%	
10 (Ave Epochs: 22)	Ave Training MSE: 0.0024	Ave Training CPU Time: 101.14s
	Ave MAPE:12.88%	
20 (Ave Epochs: 18)	Ave Training MSE: 0.0024	Ave Training CPU Time: 194.45s
	Ave MAPE: 13.01%	

Table 5.6: ANN architecture selection of Levenberg-Marquardt training

Compare with the performance of Quasi-Newton (BFGS), Levenberg-Marquardt algorithm express its enhanced ability of converging. The general MAPE from Levenberg-Marquardt training ANN performs at least 1% better from Quasi-Newton (BFGS) methods in Table 5.4. Though it cost more time for LM method in calculating each epoch than Quasi-Newton method, the LM method achieves less total CPU time for it use less epochs to converge. Table 5.7 introduces result from Levenberg-Marquardt trained ANN load prediction by a network with 10 hidden neurons.

Network Arch	Network Training Parameter	Training Performance
10	Goal: 0.0001;	Train MSE: 0.0024
	Training Epoch: 29	CPU Time: 132.6s
	MAPE: 12.62%	
Sample Training Process		

Sample Compare Between Target and Prediction of Test Data	

Table 5.7: Load forecasting by a sample Levenberg-Marquardt trained ANN with 10 neurons in hidden layer

By summarising load prediction by BP trained ANN, the simulation needs more than 200 seconds. It means that in the 10 minutes limit, it only allows 3 candidate ANNs for training. The small sample set cannot provide enough space for optimization. In all the BP training algorithms, Levenberg-Marquardt algorithm has the best performance and the least training time, still over 100 seconds. To meet the requirement of time limit with an enough large network candidate set, new method should be applied.

5.3 Radial Basis Function Network for Micro-grid Load Forecasting

5.3.1 Radial Basis Function (RBF) Network

Radial Basis Function is a traditional interpolation technique in hyperspace. In 1988, based on that biological neurons comprise local response, Broomhead and Lowe has transfer RBF into Artificial Neural Network [99]. In the following, Tomaso Poggio and Federico Girosi have demonstrated RBF network has good approximation ability in non-linear approximation [100]. RBF network attracts attentions as a result of his ability on function approximation, interpolation, pattern recognition and other intelligent applications.

A radial basis function is a function whose output is determined by the distance between the input and the centre point. It is a sensor recognizing that how far the input is from the centre. Usual functions that could be select for RBF is introduced in Table 5.8:

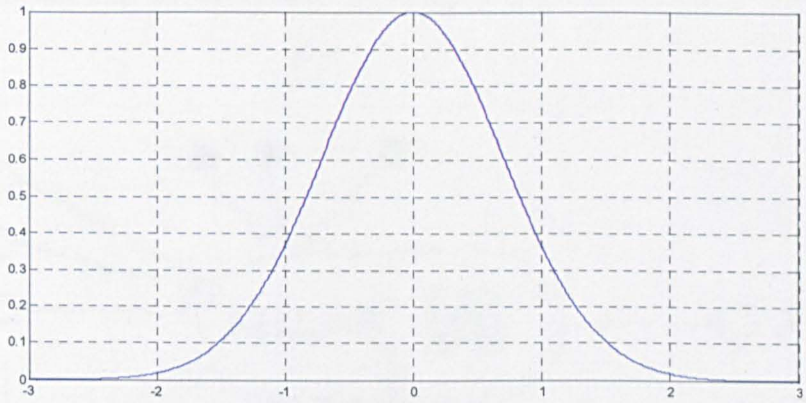
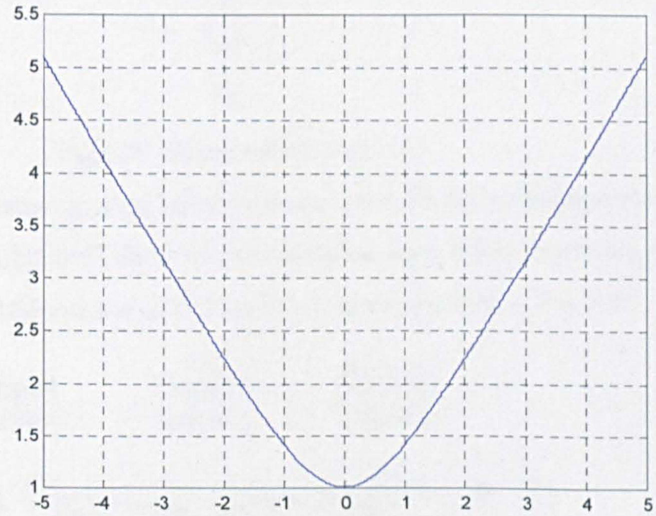
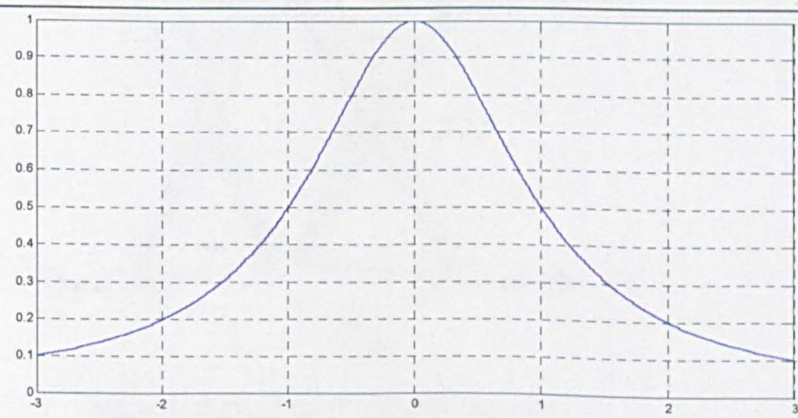
Name	Expression Graph
Gaussian	$b \cdot e^{-r^2}$
	
Multiquadric	$\sqrt{1 + b \cdot r^2}$
	
Inverse Quadratic	$\frac{1}{1 + b \cdot r^2}$
	

Table 5.8: Candidate functions for Radial Basis Function selection

Radial Basis Function network is a 3-layer feed-forward ANN. Comparing to typical BP trained network, the only difference is that the hidden layer neurons are RBF neurons and all the weights between input layer and hidden layer is 1. Output layer is the same as BP network. Fig 5.4 introduces the model of RBF neuron [101].

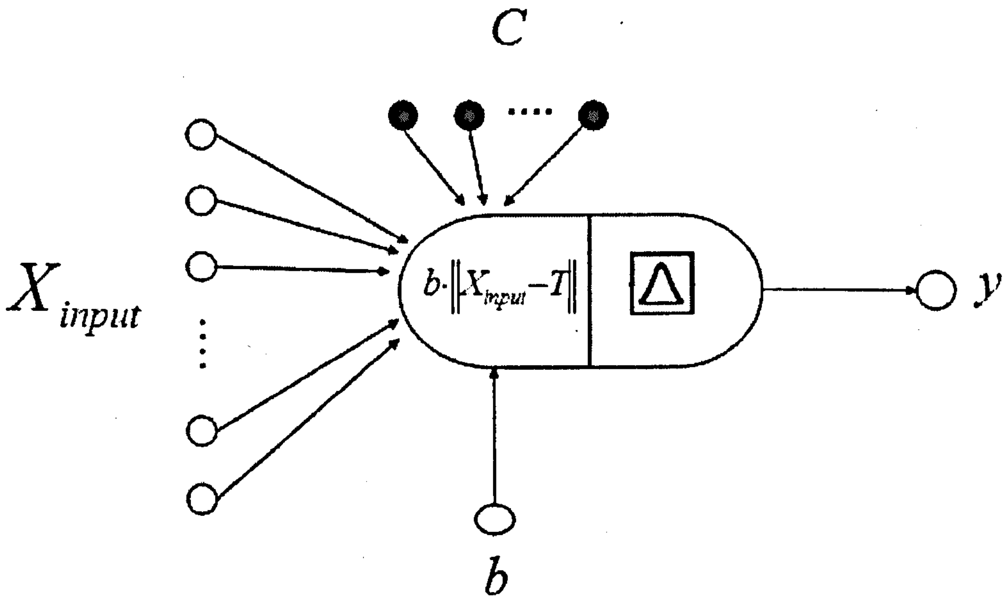


Fig 5.4: Model of RBF neuron

The RBF neuron calculates the similarity between input vector and centre vector. If the distance between input and centre become smaller, then the neuron output becomes larger. With RBF neurons, RBF network is established in Fig 5.5:

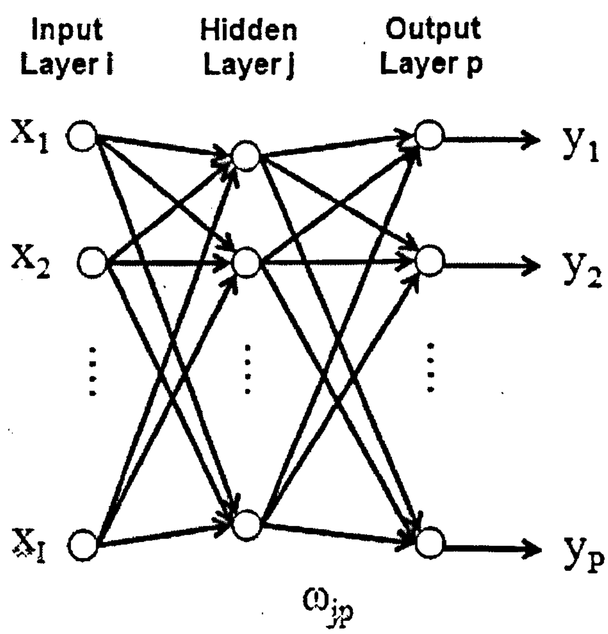


Fig 5.5: Typical RBF network structure

Tomaso Poggio and Federico Girosi have demonstrated RBF network has good approximation ability in non-linear approximation [100]. As they introduced, the main principle of RBF network is to use the similarity between inputs and centres as basis to project to the output space. The Radial Function is actually the sensitivity function that measures the similarity degree between inputs and centres, as Fig 5.6 introduces. When a specific point is set as input, all the centres contribute their similarity degree for this point to map to a corresponding point in output space.

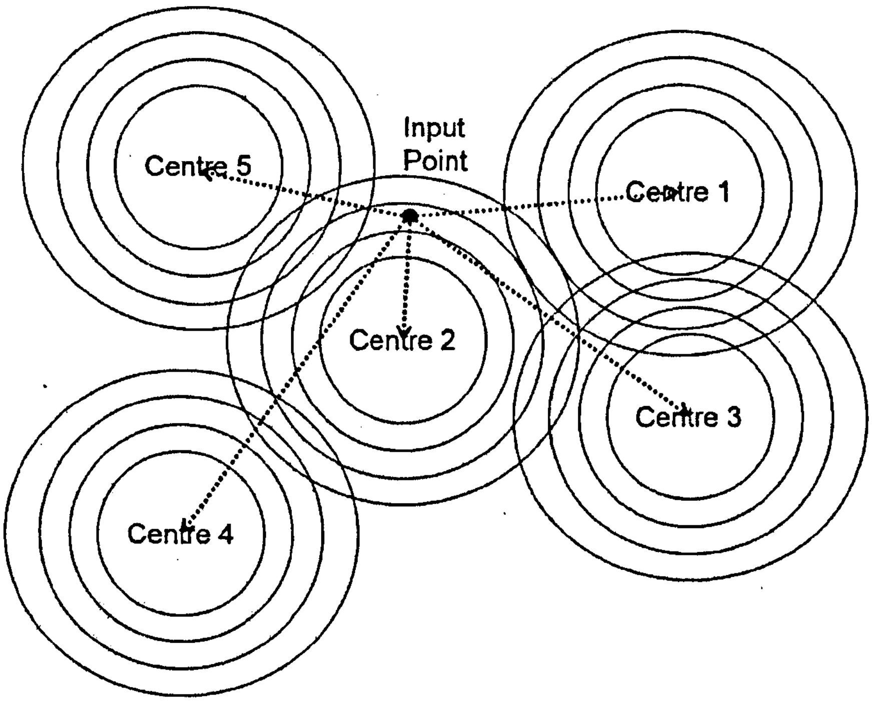


Fig 0.6: RBF network sense the input point in hyper panel by distance combination

If there are K input vectors in training set $X = [X_1, X_2, \cdots, X_k, \cdots, X_K]^T$. Each input vector contains I elements $X_k = [x_{k1}, x_{k2}, \cdots, x_{ki}, \cdots, x_{kI}]^T$. Then forward calculation from input side to output side in Fig 5.6 could be revealed in Equation (5.1) [101].

$$\left.\begin{aligned} \nu_{kj} &= G(b \cdot \|X_k - C_j\|) \\ y_{kp} &= \sum_{j=1}^{J} \omega_{jp} \cdot \nu_{kj} \end{aligned}\right\} \tag{5.1}$$

In Equation (5.1):

- *There are J RBF neurons in hidden layer. Each neuron contains a centre vector C_j. The centre space is $C = [C_1, C_2, \cdots, C_j, \cdots, C_J]$.*

- *Function $G(X)$ is the Radial Basis Function in RBF neurons. b is the bias.*
- v_{kj}: *It is the output of the j^{th} neuron in hidden layer corresponding to the k^{th} input in the training set.*
- ω_{jp}: *It is the weight between the j^{th} neuron in hidden layer and the p^{th} output element.*
- y_{kp}: *It is the p^{th} output element corresponding to the k^{th} input in the training set.*

5.3.2 Radial Basis Function (RBF) Network Training

RBF network can perfectly fit into the training set data. This case occurs when number of RBF neurons equals to the number of training set input vectors, which all input vectors are used as centres as well. This centres setting promotes input pattern recognition by ensuring all the training input will be sensed by at least one RBF neurons. In this case, the weights between hidden layer and output layer are unique determined in Equation (5.2). Rewrite Equation (5.1) in matrix form into Equation (5.2).

$$\left.\begin{array}{c} v_{kj} = G(b\cdot\|X_k - C_j\| \\ W = V^{-1}Y \\ W = \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1P} \\ \omega_{21} & \ddots & \ddots & \omega_{2P} \\ \vdots & \ddots & \ddots & \vdots \\ \omega_{J1} & \omega_{J2} & \cdots & \omega_{JP} \end{bmatrix} \\ V = \begin{bmatrix} v_{11} & v_{11} & \cdots & v_{1J} \\ v_{21} & \ddots & \ddots & v_{2J} \\ \vdots & \ddots & \ddots & \vdots \\ v_{K1} & v_{K2} & \cdots & v_{KJ} \end{bmatrix} \quad Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1P} \\ y_{21} & \ddots & \ddots & y_{2P} \\ \vdots & \ddots & \ddots & \vdots \\ y_{K1} & y_{K2} & \cdots & y_{KP} \end{bmatrix} \end{array}\right\} \tag{5.2}$$

In Equation (5.2):

- *W: It is weights matrix.*

- ω_{jp}: *It is the weight between the j^{th} neuron in hidden layer and the p^{th} output element in W.*
- *V: It is RBF neuron output matrix.*
- v_{kj}: *It is the output of the j^{th} neuron in hidden layer corresponding to the k^{th} input in the training set in V.*
- *Y: It is network output matrix.*
- y_{kp}: *It is the p^{th} output element corresponding to the k^{th} input in the training set in Y.*

The training set perfect fit method can 100% output the exact target by passing through any training input. But when training set or the system parameter appear to be large, this method not only seize large space but also have problems in low speed and over-fitting.

Facing this case, an architecture-based training algorithm for achieving a network with less complexity is selected. This method increases one RBF neuron once a time instead of putting all the input space into the centres concurrently. The training method flow chart is shown in Fig 5.7.

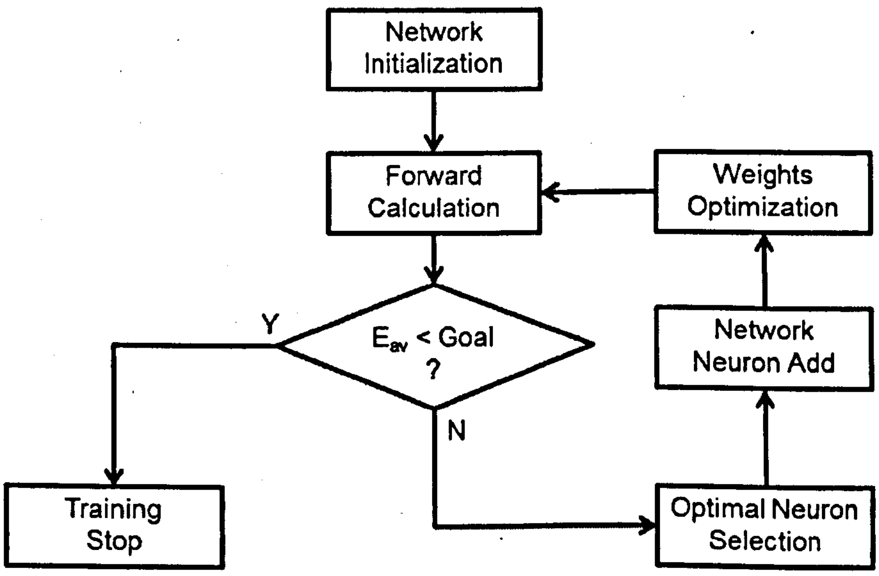


Fig 5.7: Flow chart of RBF network training

In the training method introduced by Fig 5.7, when a new neuron is added, the weights dimension will be altered. There is more than one solution for weights matrix when network architecture variation occurs. The optimal weights matrix should be selected as follow. Assume the target set from training data set is $T=[T_1,T_2,\cdots,T_K]$, each $T_k=[t_{k1},t_{k2},\cdots,t_{kP}]$. The error expression is in Equation (5.3).

$$\left.\begin{aligned} E_{av} &= \frac{1}{K}\sum_{k=1}^{K} E_k \\ E_k &= \frac{1}{2}\sum_{p=1}^{P} e_{kp}^2 \\ e_{kp} &= t_{kp} - y_{kp} \end{aligned}\right\} \tag{5.3}$$

Equation 5.3 description: for the k_{th} input vector from training input set:

- *e_{kp}: It is the error between target and actual output at the p_{th} element in the output vector.*
- *E_k: It is the Least Mean Square error between the whole target vector and the output vector corresponding to the k_{th} input vector.*
- *E_{av}: It is the target function whose minimum is interested. It is the batch learning error which considers the errors corresponding to all the training input vectors.*

Noting Equations (5.1) and (5.3), gradient of E_{av} by weights is expressed in Equation (5.4).

$$\left.\begin{aligned} \frac{\partial E_{av}}{\partial \omega_{jp}} &= \frac{1}{K}\sum_{k=1}^{K}\frac{\partial E_k}{\partial \omega_{jp}} \\ &= \frac{1}{K}\sum_{k=1}^{K}\frac{\partial E_k}{\partial e_{kp}}\cdot\frac{\partial e_{kp}}{\partial y_{kp}}\cdot\frac{\partial y_{kp}}{\partial \omega_{jp}} \\ &= \frac{1}{K}\sum_{k=1}^{K} e_{kp}\cdot(-1)\cdot \mathrm{v}_{kj} \\ &= \frac{1}{K}\sum_{k=1}^{K}(y_{kp}-t_{kp})\cdot \mathrm{v}_{kj} \end{aligned}\right\} \tag{5.4}$$

At the stationary point, gradient is zero. Equation (5.4) turns into Equation (5.5) [102].

$$\left.\begin{array}{l} \frac{1}{K}\sum_{k=1}^{K}(y_{kp}-t_{kp})\cdot \mathrm{v}_{kj}=0 \\ \Rightarrow \sum_{k=1}^{K} y_{kp}\cdot \mathrm{v}_{kj}=\sum_{k=1}^{K} t_{kp}\cdot \mathrm{v}_{kj} \\ \Rightarrow V_j^T\cdot Y_p = V_j^T\cdot T_p \\ V_j=\begin{bmatrix} \mathrm{v}_{1j} \\ \mathrm{v}_{2j} \\ \vdots \\ \mathrm{v}_{Kj} \end{bmatrix}_{K\times 1} \quad Y_p=\begin{bmatrix} y_{1p} \\ y_{2p} \\ \vdots \\ y_{Kp} \end{bmatrix}_{K\times 1} \quad T_p=\begin{bmatrix} t_{1p} \\ t_{2p} \\ \vdots \\ t_{Kp} \end{bmatrix}_{K\times 1} \end{array}\right\} \tag{5.5}$$

Equation (5.6) expands Equation (5.5) into the whole network in matrix form.

$$\left.\begin{array}{l} V^T\cdot Y = V^T\cdot T \\ V=[V_1,V_2,\cdots V_J] \\ Y=[Y_1,Y_2,\cdots,Y_P] \\ T=[T_1,T_2,\cdots,T_P] \end{array}\right\} \tag{5.6}$$

Noted from Equation (5.1), relation between matrix Y and V is:

$$\left.\begin{array}{l} Y^T = W^T\cdot V^T \\ Y=V\cdot W \end{array}\right\} \tag{5.7}$$

Consider Equations (5.6) and (5.7) together, the optimized weights W is shown in Equation (5.8):

$$\left.\begin{array}{l} V^T\cdot V\cdot W = V^T\cdot T \\ \Rightarrow W=(V^T\cdot V)^{-1}\cdot V^T\cdot T \end{array}\right\} \tag{5.8}$$

5.3.3 Radial Basis Function Network for Micro-grid Load Forecasting

For RBF Network training, this chapter selects Radial Basis Network Section in Matlab 7.11.0 (R2010b) Neural Network Toolbox. It is a specified tool for RBF network training in Neural Network Toolbox.

Use RBF to train the network with same data set as BP training in Section 5.2.2. Samples of results are shown in Table 5.9.

	Goal 0.001	Goal 0.002	Goal 0.004	Goal: 0.006
Sensitivity 1.6651 (0.5)	Ave MAPE 0.3645	Ave MAPE 0.3607	Ave MAPE 0.3647	Ave MAPE 0.3836
	Ave CPU Time 1162.9s	Ave CPU Time 799.27s	Ave CPU Time 465.32s	Ave CPU Time 303.58s
	Ave Epochs 690	Ave Epochs 534	Ave Epochs 385	Ave Epochs 285
Sensitivity 0.8326 (1)	Ave MAPE 0.2172	Ave MAPE 0.2064	Ave MAPE 0.2173	Ave MAPE 0.2444
	Ave CPU Time 528.16s	Ave CPU Time 218.38s	Ave CPU Time 44.49s	Ave CPU Time 19.12s
	Ave Epochs 392	Ave Epochs 208	Ave Epochs 64	Ave Epochs 33
Sensitivity 0.4163 (2)	Ave MAPE 0.2237	Ave MAPE 0.1765	Ave MAPE 0.1495	Ave MAPE 0.2160
	Ave CPU Time 500.93s	Ave CPU Time 140.6s	Ave CPU Time 6.57s	Ave CPU Time 3.20s
	Ave Epochs 404	Ave Epochs 153	Ave Epochs 11	Ave Epochs 6
Sensitivity 0.2775 (3)	Ave MAPE 0.2483	Ave MAPE 0.1717	Ave MAPE 0.1336	Ave MAPE 0.1336
	Ave CPU Time 666.58s	Ave CPU Time 155.88s	Ave CPU Time 7.14s	Ave CPU Time 6.88s
	Ave Epochs 487	Ave Epochs 170	Ave Epochs 11	Ave Epochs 11
Sensitivity 0.1665 (5)	Ave MAPE 0.2637	Ave MAPE 0.1789	Ave MAPE 0.1717	Ave MAPE 0.2039
	Ave CPU Time 758.04	Ave CPU Time 207.25s	Ave CPU Time 10.65s	Ave CPU Time 8.17s
	Ave Epochs 491	Ave Epochs 209	Ave Epochs 16	Ave Epochs 13
Sensitivity 0.1189 (7)	Ave MAPE 0.2583	Ave MAPE 0.1880	Ave MAPE 0.1637	Ave MAPE 0.2420
	Ave CPU Time 834.73s	Ave CPU Time 406.133s	Ave CPU Time 14.65s	Ave CPU Time 10.86s
	Ave Epochs 531	Ave Epochs 334	Ave Epochs 25	Ave Epochs 17
Sensitivity 0.0925 (9)	Ave MAPE 0.2791	Ave MAPE 0.1835	Ave MAPE 0.1476	Ave MAPE 0.2436
	Ave CPU Time	Ave CPU Time	Ave CPU Time	Ave CPU Time

	769.00s	213s	12.4s	11.18s
	Ave Epochs 585	Ave Epochs 216	Ave Epochs 21	Ave Epochs 17

Table 5.9: Samples of RBF network training results.

In Table 5.9, titles descriptions are listed below:

- **Ave MAPE**: The average value of all the Mean Absolute Percentage Error performances of the same type networks corresponding to training data set.
- **Ave CPU Time**: The average CPU time spent by a certain type of networks.

Example of a RBF network for Micro-grid load forecast is introduced in Table 5.10.

Network Architecture	Network Training Parameter	Training Performance
11	Goal: 0.004;	Train MSE: 0.0039
	Training Epoch: 11	CPU Time: 7.14s
	MAPE: 13.36%	
Sample Training Process		
Sample Compare Between Target and Prediction		

Table 5.10: Example of RBF network for Micro-grid load forecast

5.4 Micro-grid Load Forecasting Analysis

5.4.1 Analysis on Prediction with RBF Networks

There are two main influencing factors for RBF training:

- **Network Architecture**: Number of RBF neurons in hidden layer.
- **Radial Basis Function Sensitivity**: the coefficient b in Equation (5.1).

In the process of RBF network training, as the epochs increase, the number of RBF neurons increase. The network will have stronger ability in approximation and is more and more similar to the network set by perfect fit method. In this case the network error of training data set is decreasing. So if the accuracy goal is set to lower, the trained network will have higher accuracy for training data set and will have more RBF neurons. Fig 5.8 introduces a sample case of network result variation when training accuracy goal is changing with sensitivity 0.1665.

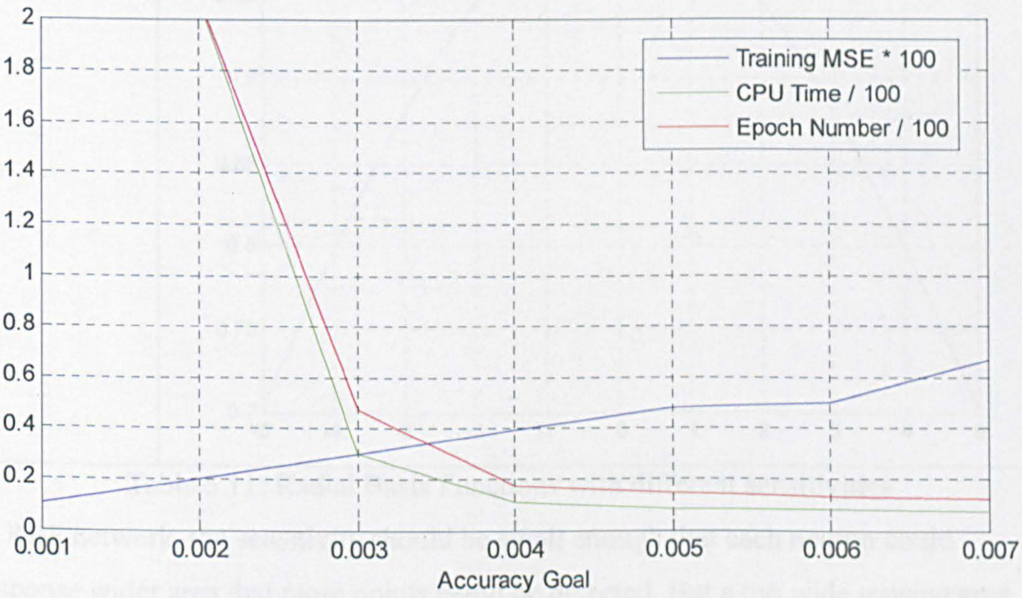


Fig 5.8: a sample case of network result variation as goal change with sensitivity 0.1665

The sensitivity coefficient controls the shape of Radial Basis Function. Large sensitivity RBF can sense wider range of its input space. Table 5.11 compares RBF with different sensitivities.

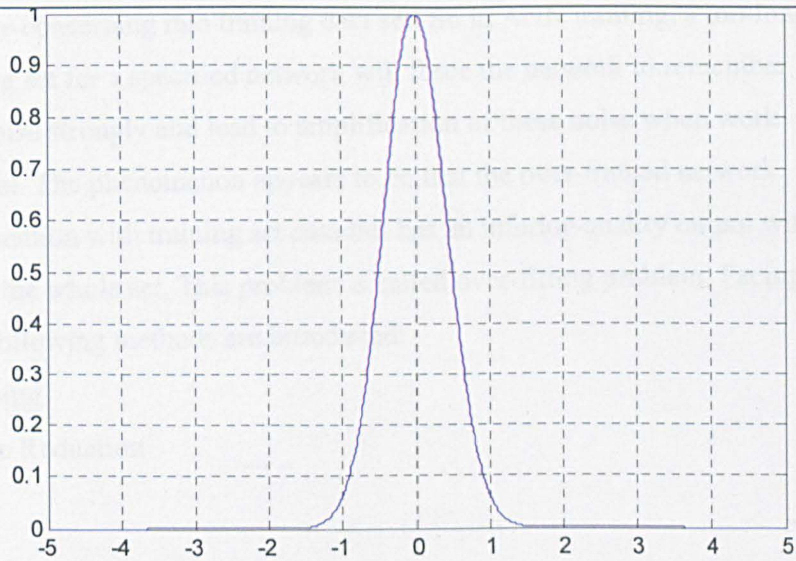
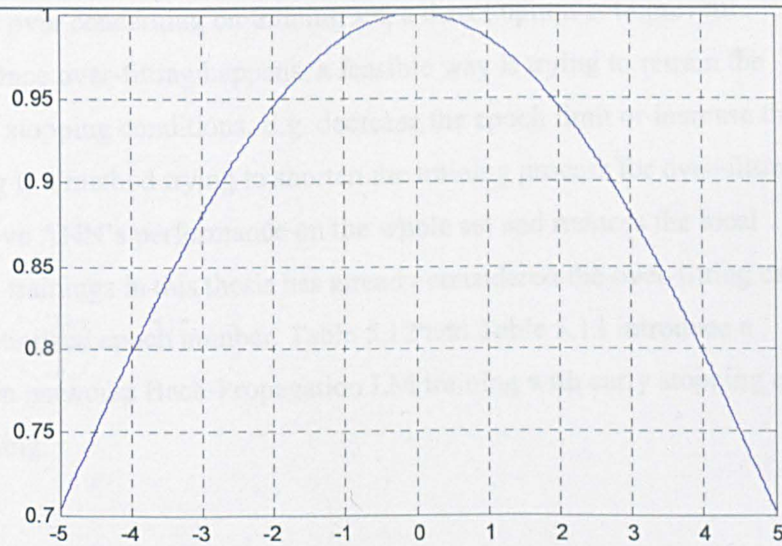
<table>
<tr><th>Sensitivity</th><th></th></tr>
<tr><td>1.6651</td><td>
</td></tr>
<tr><td>0.1189</td><td>
</td></tr>
</table>

Table 5.11: Radial Basis Functions with different sensitivities

In RBF network, the sensitivity should be small enough that each neuron could response wider area that more points could be detected. But a too wide sensing area will lead to fewer differences between outputs from each neuron. So as in Table 5.10, an optimal sensitivity should be selected.

5.4.2 Analysis on ANN Generalization

In training of Artificial Neural Network, there will be a problem on over training, which also named as over-fitting. The training data set is always supposed to be perfect and averagely representing the whole set. But the actual situation is that the training set more or less may be lacking of certain patterns or bias to certain patterns involuntary or voluntary, even contains error. These unpredictable noisy factors will be amplified by over-concerning into training data set. So in ANN training, a too-long training on a training set for a specified network will force the network to remember those training set noise strongly and lead to amplification of these noise when work back to the whole set. The phenomenon appears to be that the over-trained network has a good approximation with training set data but has an inferior-quality output with other data set from the whole set. This problem is called over-fitting problem. Facing this problem, two following methods are introduced:

- Early Stopping
- Architecture Reduction

Early Stopping

To prevent an ANN over concerning on training set, a direct option is to prevent overlong training. Once over-fitting happens, a feasible way is trying to retrain the network with lower stopping conditions. E.g. decrease the epoch limit or increase the goal. Early stopping is a method trying to shorten the training process for over-fitting case, so as to improve ANN's performance on the whole set and reduces the local noise. All the ANN trainings in this thesis has already considered the over-fitting case and trained with optimized epoch number. Table 5.12 and Table 5.13 introduce a comparison between networks Back-Propagation LM training with early stopping and without early stopping.

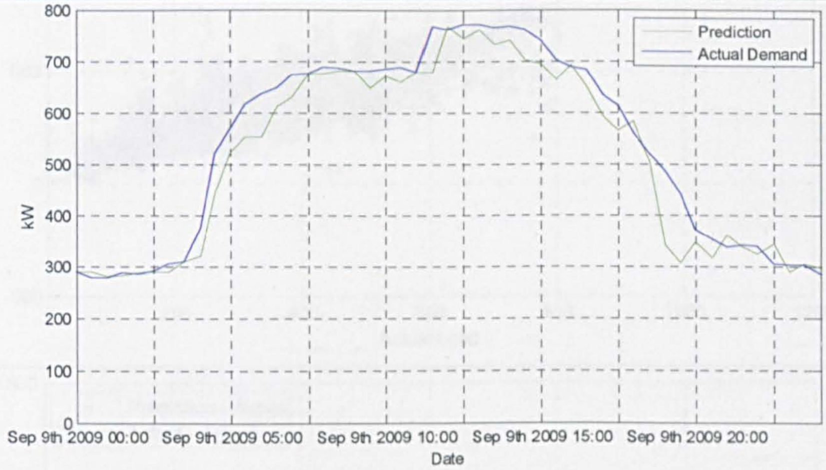
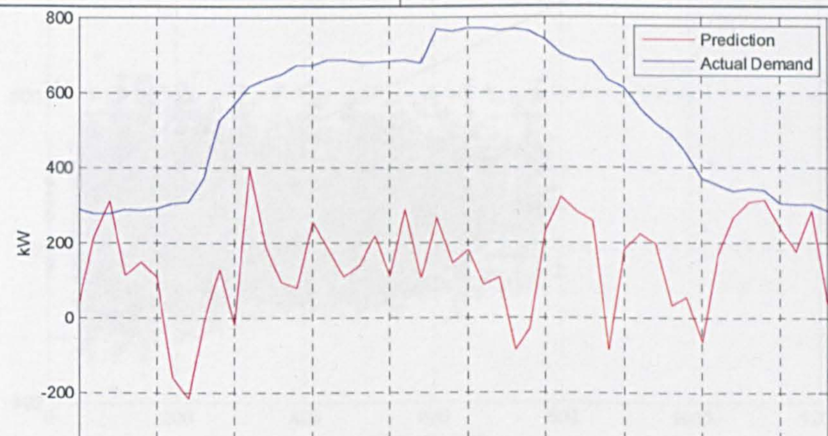
<table>
<tr><th>Network Architecture</th><th>Network Training Parameter</th><th>Training Performance</th></tr>
<tr><td rowspan="3">Early Stopping</td><td>MAPE: 12.60%</td><td>Train MSE: 2.294×10^{-3}</td></tr>
<tr><td>Training Epoch: 15</td><td>CPU Time: 150.3s</td></tr>
<tr><td colspan="2">
</td></tr>
<tr><td rowspan="3">Over-Fitting</td><td>MAPE: 78.93%</td><td>Train MSE: 2.256×10^{-3}</td></tr>
<tr><td>Training Epoch: 201</td><td>CPU Time: 2430.8</td></tr>
<tr><td colspan="2">
</td></tr>
</table>

Table 5.12: Training comparison between early stopping and over-fitting

From Table 5.12, the same network is trained with different iteration. The more trained case though achieve better performance in MSE with more CPU time, when test with data set other than training data, the MAPE emerges rising up. The comparison of Prediction and Actual Demand reveals that the over – fitting case do not fit to data set from the whole set other than training data.

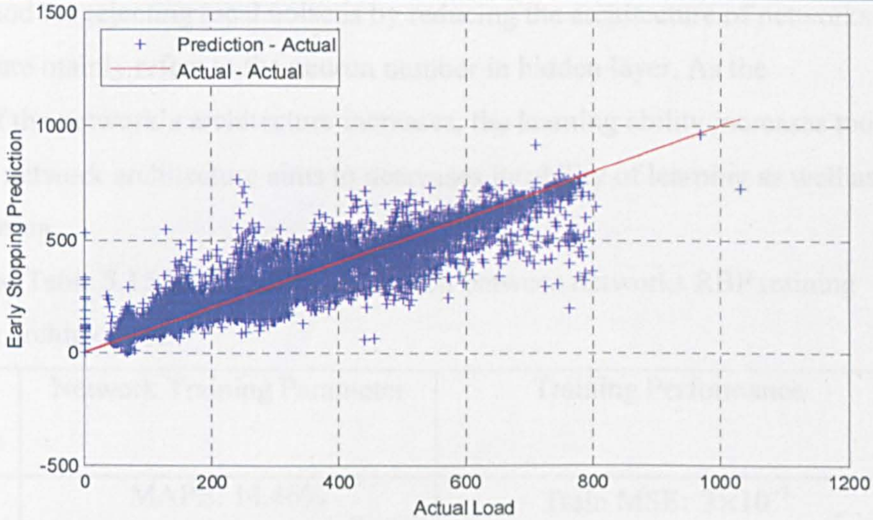
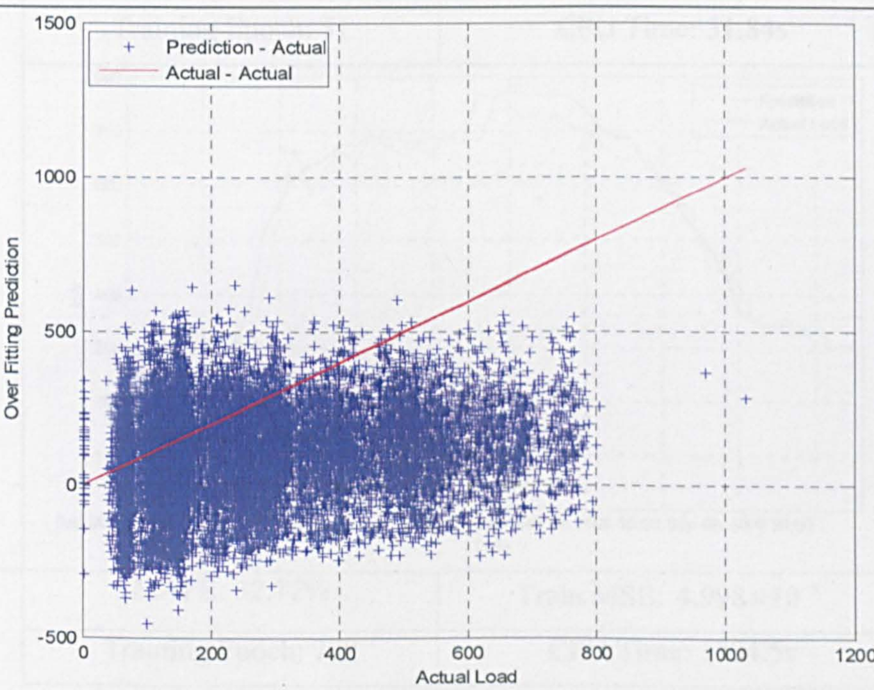
<table>
<tr><td></td><td>Examination Plot of Prediction – Actual Load</td></tr>
<tr><td>Early Stopping</td><td>
</td></tr>
<tr><td>Over Fitting</td><td>
</td></tr>
</table>

Table 5.13: Examination plot compare the differences between early stopping and over-fitting case.

Table 5.13 uses examination plot to find out the performance difference between early stopping and over-fitting. The point on the red line represents that prediction at this point is exactly the same as the actual load. So in examination plot, the more blue points converge to the red line, the better the prediction is. Obviously early stopping helps the network to achieve a better converging to the perfect line, so as a better prediction.

Architecture Reduction

Another method for rejecting local noise is by reducing the architecture of networks. The architecture mainly refers to the neuron number in hidden layer. As the complexity of the network's architecture increases, the learning ability increases too.. Reducing the network architecture aims to decreases its ability of learning as well as the noise learning.

Table 5.14 and Table 5.15 introduce a comparison between networks RBF training with different architectures.

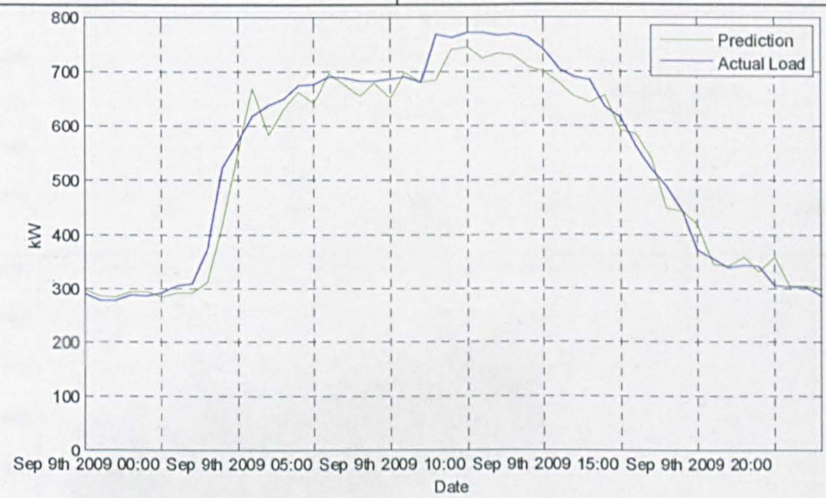
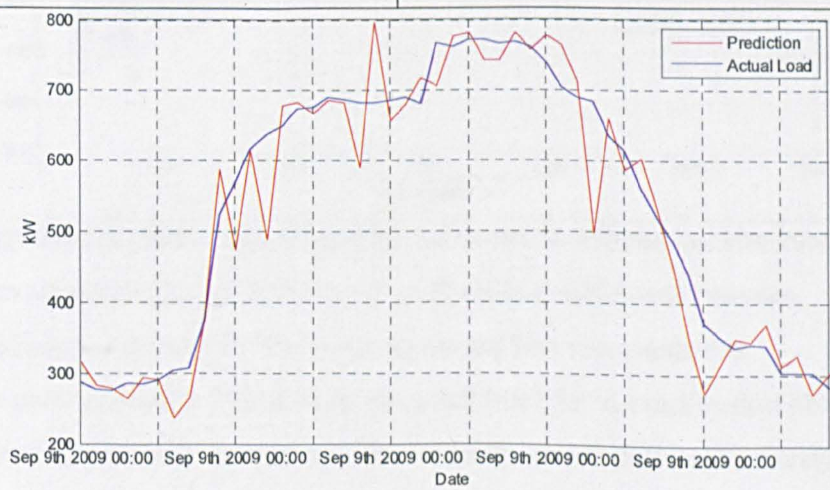
<table>
<tr><td>Network Architecture</td><td>Network Training Parameter</td><td>Training Performance</td></tr>
<tr><td rowspan="3">51 RBF Neurons</td><td>MAPE: 14.46%</td><td>Train MSE: 3×10^{-3}</td></tr>
<tr><td>Training Epoch: 51</td><td>CPU Time: 31.84s</td></tr>
<tr><td colspan="2">
</td></tr>
<tr><td rowspan="3">779 RBF Neurons</td><td>MAPE: 32.77%</td><td>Train MSE: 4.998×10^{-4}</td></tr>
<tr><td>Training Epoch: 779</td><td>CPU Time: 1704.5s</td></tr>
<tr><td colspan="2">
</td></tr>
</table>

Table 5.14: Training comparison between different architecture

From Table 5.14, networks with different architecture are compared. The network with too much RBF neurons appears to have not only more training CPU time but also an over-fitting case. The compare graph of Prediction and Actual Demand also reveals that the over – fitting case caused by too much RBF neurons do not fit to data set from the whole set other than training data.

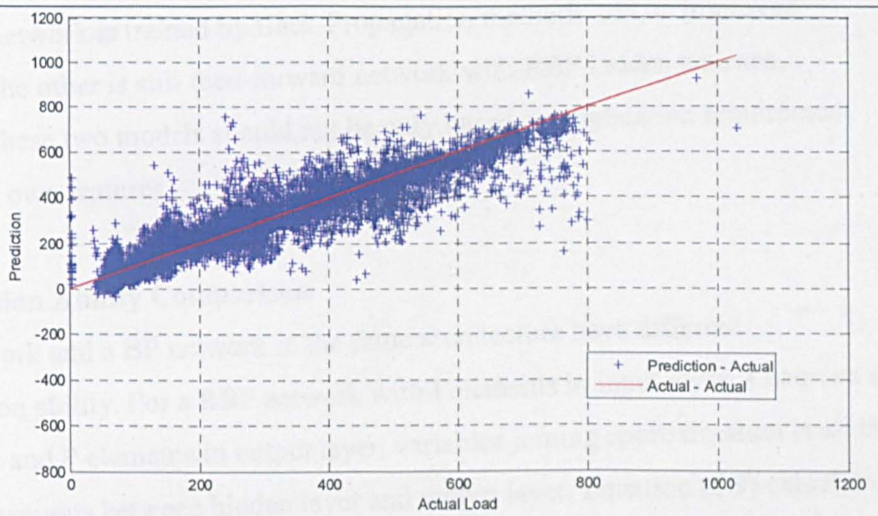
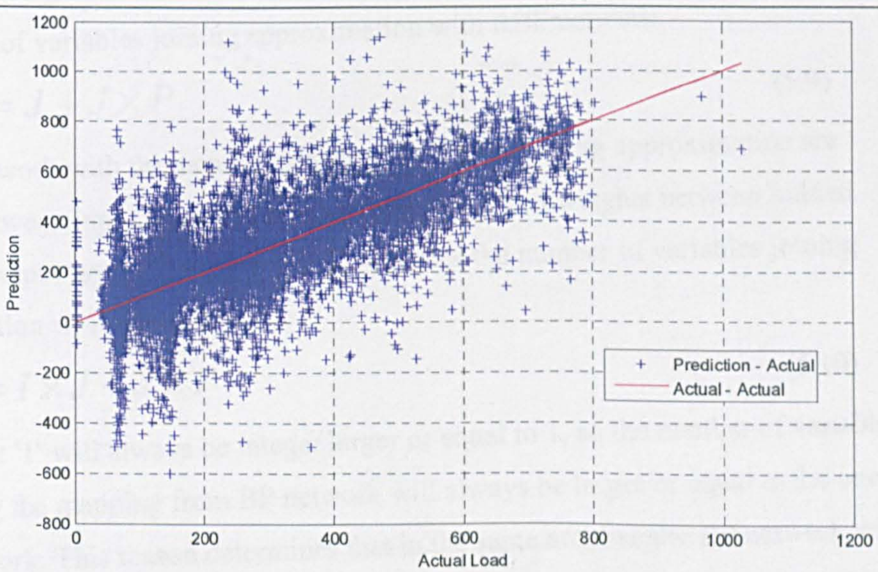
<table>
<tr><td></td><td>Examination Plot of Prediction – Actual Load</td></tr>
<tr><td>51 RBF Neurons</td><td>
</td></tr>
<tr><td>779 RBF Neurons</td><td>
</td></tr>
</table>

Table 5.15: Examination plot compare between networks in different architectures

Table 5.15 uses examination plot to find out the performance difference between networks with different architecture. The point on the red line represents that prediction at this point is exactly the same as the actual load. So in examination plot, the more blue points converge to the red line, the better the prediction is. Obviously a network with too much RBF neurons contains more noise than the one with less neuron.

5.4.3 Analysis on BP Trained ANN and RBF Trained ANN for Micro-grid Load forecast

For Micro-grid load forecasting problem, two types of networks are applied. One is feed-forward network whose hidden-layer neurons have sigmoid activation function. This type of network is trained by Back-Propagation methods and its improved algorithms. The other is still feed-forward network with RBF hidden neurons. Selection of these two models should not be only based on application requirement but also their own features.

Approximation Ability Comparison

A RBF network and a BP network in the same architecture have different approximation ability. For a RBF network with I elements in input layer, J neurons in hidden layer and P elements in output layer, variables joining approximation is all the centres and weights between hidden layer and output layer. Equation (5.9) calculates the number of variables joining approximation with RBF network.

$$NO._{RBF} = J + J \times P \tag{5.9}$$

In a BP network with the same architecture, variables joining approximation are weights between input layer and hidden layer, as well as weights between hidden layer and output layer. Equation (5.10) calculates the number of variables joining approximation with BP network.

$$NO._{BP} = I \times J + J \times P \tag{5.10}$$

Due to that 'I' will always be integer larger or equal to 1, so the number of variables expressing the mapping from BP network will always be larger or equal to the one in RBF network. This reason determines that in the same architecture BP network will have better approximation ability than RBF network. Table 5.16 introduces compared results on a network with 10 hidden neurons from BP training and RBF training. Both networks are trained with – fitting rejection. Though the difference in examination plot is not obvious, the MAPE proves that BP achieves a better approximation.

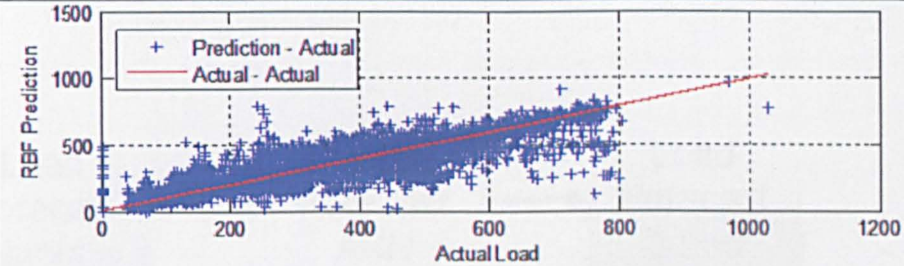
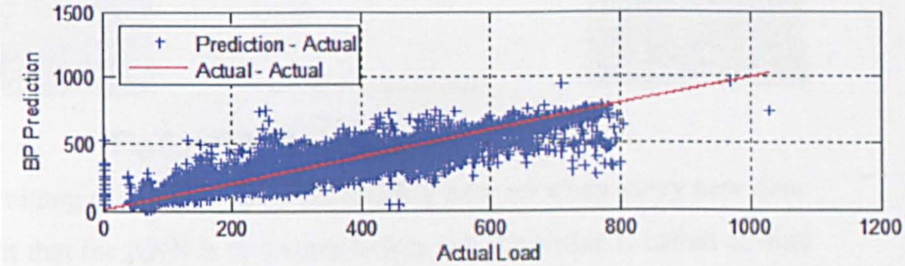
	Examination Plot of Prediction – Actual Load
RBF 11 Neus (13.36%)	
BP 10 Neus (12.27%)	

Table 5.16: Approximation Ability compare between RBF network and BP network

Training Speed Comparison

Though BP network can achieve better performance, it also spends quite a lot of time in network training. Due to the variables revised in training in BP network is more than the ones in RBF training, RBF usually achieve a faster training procedure. One example is that BP training usually requires a fix architecture fixed network but RBF does not. This is because the training for one architecture of RBF network is quite fast that the system could integrate the architecture selection in the training as well. In Table 4.16, the BP network spends 83.3 seconds for training, comparing to 7.14 seconds from RBF training.

Application Framework for BP Network and RBF network

Load forecasting system requires trained ANN. But as the time goes by, load pattern of the Micro-grid load may be varying slowly or rapidly. Working with network trained by antiquated data may leads to large error. So network should be re-trained by updated training set.

One re-training plan is to replace the old network by a new trained network every certain period. When the network is working with the whole system, no training occurs. This plan is named off-line learning. Fig 5.9 introduces the off-line learning. The time cost for off-line learning system is only the forward calculation time.

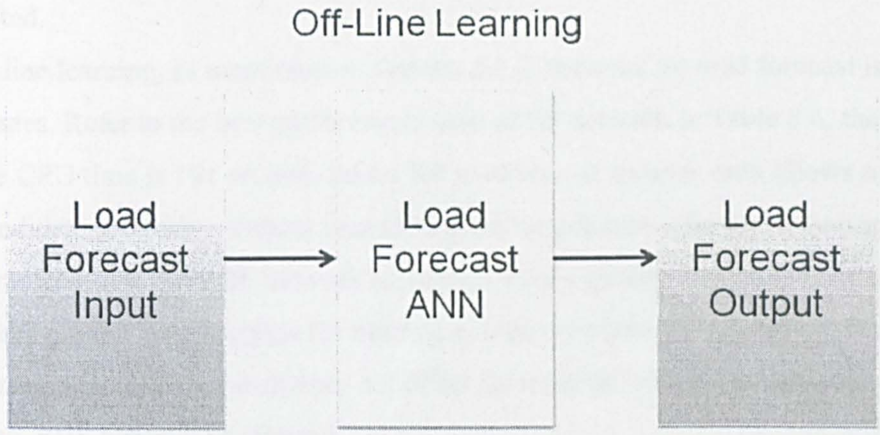


Fig 5.9: Off-line learning for load forecast.

Another re-training plan is to update the training data set when every new data appears. After that the ANN is re-trained before use. This plan is called on-line learning. Fig 5.10 reveals the on-line learning. The on-line learning can guarantee an always fresh system. But the time cost includes not only the forward calculation but also the time for training.

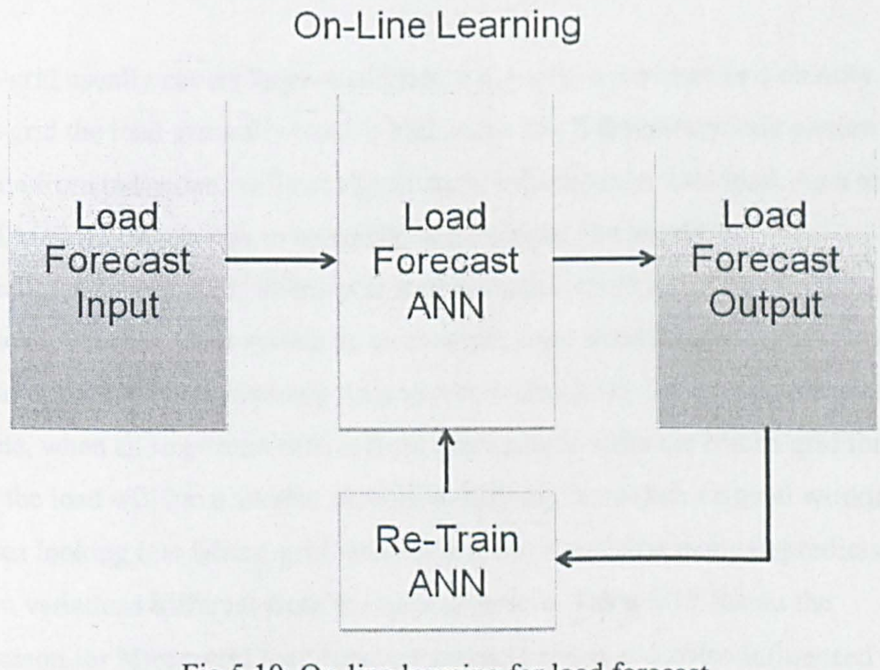


Fig 5.10: On-line learning for load forecast

ANN's forward calculation usually cost small time period. So for off-line learning, both BP networks and RBF networks are suitable. The competition is only placed on their performance. In this case, due to BP networks achieves a better performance in this Micro-grid load forecasting than RBF networks, the optimal BP trained network is selected.

For on-line learning, as mentioned in Section 5.1.2, the time for load forecast is only 10 minutes. Refer to the best performance case of BP network in Table 5.6, the average CPU time is 101 second. So for BP network, 10 minutes only allows training of 6 candidate networks, without considering training failure. The performance cannot be guaranteed. But the RBF network in Table 5.9 only spends 7 seconds for training, providing enough time margins for training a large networks set. Moreover, the performance for RBF network does not differ far from the one of BP networks. So in this case, RBF network is selected.

5.4.4 Analysis on Compare between Micro-grid and Macro-grid Load Forecast

Load Forecast is critical for grid management and planning in power grid with whatever scale. But the differences of load features between Micro-grid (small scale) and Macro-grid (large scale) produce significant impact on their load forecasting work.

Macro-grid usually covers large customers, e.g. a city, a province or a country. In Macro-grid the load generally reach a high value that a temporary load pattern variation from individual will not significantly influence the total load. As a result, load of Macro-grid appears to be stable, as demanded in Chapter 4.

Comparing to Macro-grid, Micro-grid load has much more random influencing noise on its load. Set the chiller system as an example, once maintenance work is operated, the load of demand will suddenly drop to zero without any indication. Another example, when an important officer from government visits the Micro-grid for a small while, the load will have another alteration differing from their original working plan. So when looking into Micro-grid load feature, there could be many unpredictable random variations different from the normal pattern. Table 5.17 shows the comparison for Micro-grid load between normal pattern and noise-influenced pattern.

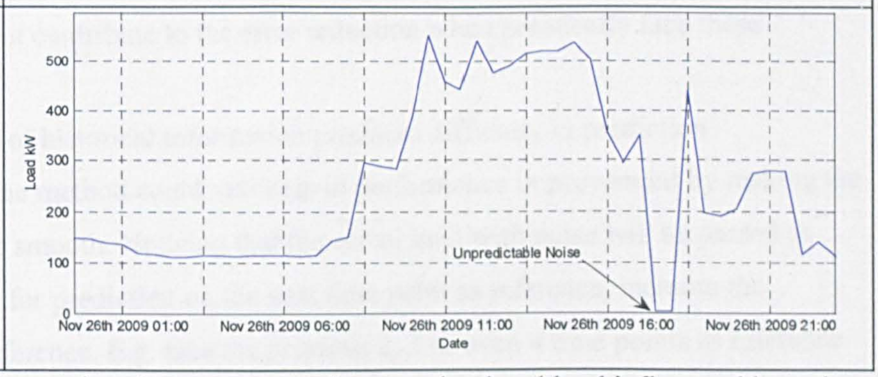
<table>
<tr><td>Normal Case (23/11/2008)</td><td>
</td></tr>
<tr><td>Influenced by Noise (26/11/2009)</td><td>
</td></tr>
</table>

Table 5.17: Comparison between normal pattern load and load influenced by noise

Table 5.17 held a one-day load comparison between two days whose weather condition is similar. The normal case, as people start working in the morning, the load of chiller increases to a peak value. It decreases in the evening as people gradually leave the university. But in the noise case, the chillers' load decreases into 0 between 18:00 to 19:00 on Nov 26^{th} 2009, and then rises up again. It may be a serious chiller system faults, or a maintenance plan, or even heavy disaster that all the people had left the university. Relevant information of that time point was lost with only the load data left. When load forecasting system comes to this time point, as the input does not vary much as usual, the output error will be very large. And this factor is the main reason that the accuracy of prediction in Micro-grid tends to be larger than prediction in Macro-grid in most cases. The examination plot also reveals this factor in Fig 5.11.

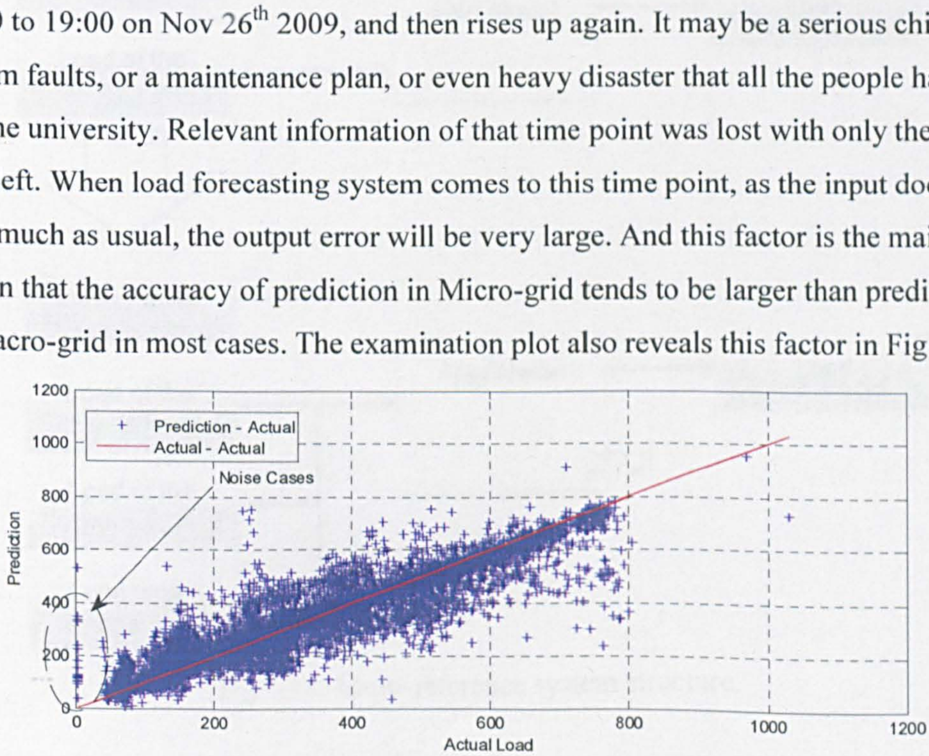


Fig 5.11: Noise case in examination plot

The data provider has verified that actual load never touch zero in the chiller system normally. But in Fig 5.11 there are some points in the circle whose actual load is zero but the prediction is far more. These unpredictable noises distort the prediction not only at the time point of noise, but also the time point next to it for the actual load with noise will be passed as one of the input for prediction on the next time point.
A general idea is to reject the data looks as if noise. But due to the unclear reasons for these tough data points, rejection algorithm may eliminate a segment of normal pattern and do not contribute to the error reduction when practically face these situations.
Though the loss of historical information produces difficulty in prediction improvement, one method could still help in performance improvement by making the prediction curve smooth. Noticing that the actual load with noise will be passed as one of the input for prediction on the next time point as reference, increase the dimension of reference. E.g. take the previous 2, 3 or even 4 time points as reference instead of only 1 time point as Fig. 5.12 shows. This multi-reference method is selected when error occurs on one reference, the impact of this error will be relaxed by other normal references.

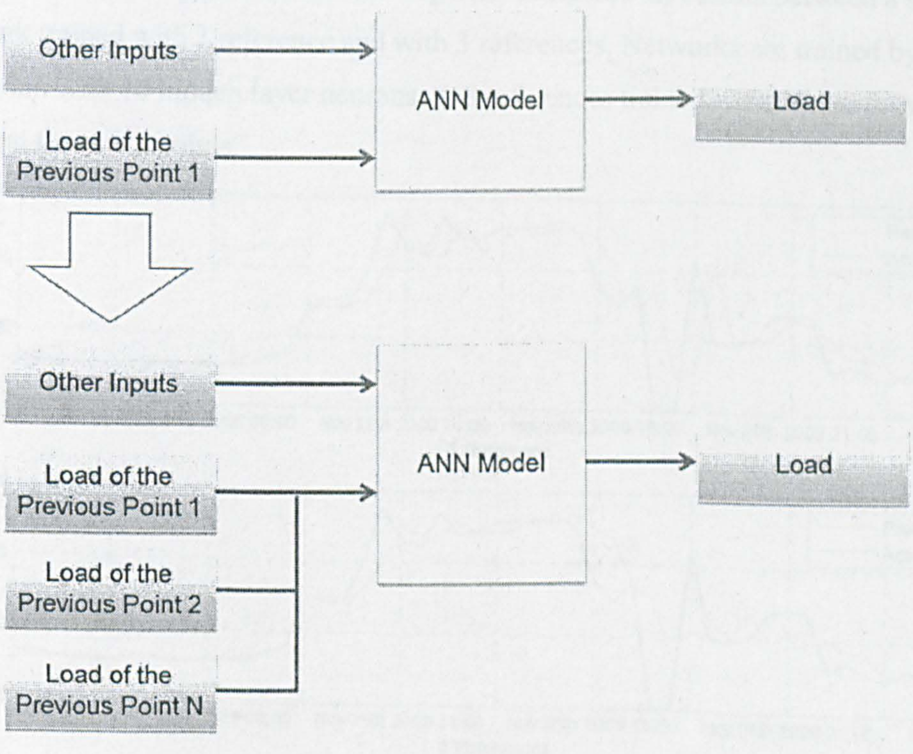


Fig 5.12: Multi-reference system structure.

Multi-reference input does not mean that improvement can be achieved by unlimited addition of reference point. As in normal case in Table 5.17, the variation of load at 6 o'clock to 10 o'clock belongs to normal load pattern as people are gradually come to university for study or work. A system with too many references will relax these normal cases as well as the noise. Table 5.18 introduces the results compare example among different multi-reference system with 10 hidden- neuron ANN trained by LM.

Reference Quantity	Ave MAPE	Ave CPU Time	Ave Epoch
1	12.88%	101.14s	22
2	11.87%	156.53s	24
3	11.68%	129.71s	20
4	11.87%	185.34s	33
5	11.89%	155.84s	31

Table 5.18: Result compare between different multi-reference systems

When more references are added into input space, though improve the performance, it also increases the training CPU time. User should consider selection of reference quantity deeply with the requirement. Fig 5.13 compares the results between a sample network trained with 1 reference and with 3 references. Networks are trained by LM algorithm with 10 hidden layer neurons. A 3 references trained network performs better at the noisy points.

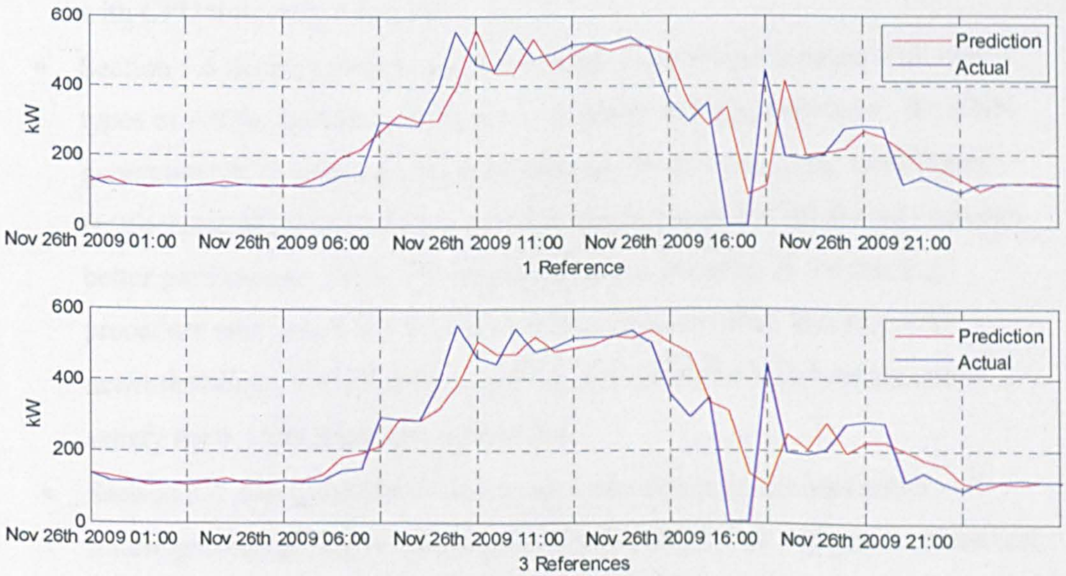


Fig 5.13: result compare of networks trained with different reference quantity.

A better average performance should be based on input space inclusion of the bad data point patterns.

5.5 Conclusion

Feature of Micro-grid load is not completely the same as load in Macro-grid. The lower capability of noise from random incidences determines that Micro-grid load is born with unpredictable distortion. This Chapter mainly focuses on micro-grid load forecast in the following sections:

- Section 5.2 sketches the micro-grid load forecast problem on a chiller system in City University of Hong Kong. Figure 5.2 reveals a black-box model for problem solving which similar to Figure 4.9 in macro-grid load forecast. The black-box input includes Weather Conditions, Day Style, Demand of the previous point and Time Point Index. One more factor that different from macro-grid load forecast in chapter 4 is the model for micro-grid consider the university calendar as a typical influencing factor. The black-box is approximated by multi-layer perceptron trained in the three algorithms introduced in chapter 4. The best MAPE is 12.62% with CPU time 132s.
- Section 5.3 introduces a new type of ANN, Radial Basis Function Network, for micro-grid load forecast. It is trained in an algorithm shown in Figure 5.7. The performance of RBF networks on chiller system load forecast is 13.36% with CPU time only 7 seconds.
- Section 5.4 deploys various analyses on micro-grid load forecast with two types of ANNs. Including analyses on network self-characteristics, the ANN generalization in training is also introduced. What's more, due to different requirement of on-line learning and off-line learning, BP MLP could achieve better performance but fail to deploy in on-line learning as the training procedure take long CPU time for one candidate network training. RBF network though with a bit worse performance, but the high training speed satisfy the on-line learning requirement.
- Section 5.4 also investigates on the more obvious noise in load pattern of micro-grid comparing to macro-grid. A new model type with more references points is introduced for accuracy improvement. But the micro-grid load

forecast accuracy is still worse than macro-grid load forecast for difficulties in significantly decreasing impact from random incidences.

Chapter 6

Conclusion and Future Work

6.1 Thesis Conclusion

Facing the serious climate change and energy terrain, Smart Grid is the trend for sustainable development worldwide. Providing various solutions to the traditional problems, Smart Grid not only integrates new applications like multi-types generations, storage, electric vehicle, but also provides new services like metering automation, dynamic pricing, Demand Side Management, as well as promotion for a greener consuming way. Various works are placed at Smart Grid definition, characteristics summarization, standardization and Smart Grid test bedding. But seldom people have organized the above work in a reasonable scope of developing a procedure. This thesis has organized the scope design into a four-step procedure as follows:

1) Summarizing the aims of national development to form an object for Smart Grid.
2) Based on the object, find out the new contributions to these aims from new power system. The contribution develops the characteristics so each characteristic could reflect the aims.
3) With the characteristics, find out what technologies and standards that could help achieving these characteristics.
4) With the above three steps, a scope design will be completed.

The procedure has successfully integrated most research and application works of Smart Grid into a framework system during project development. This procedure will be a suggestive reference for those districts that prefer to establish their Smart Grid in a practical way.

As Smart Grid develops, new services and technologies deployment will need an optimal planning and coordination. Load forecast is one of the necessary technologies for providing information as all deployments are based on demand load. . In Smart Grid demand response, customers will receive the predicted price from utilities at smart meters that help users to manage their consumption. The predicted price is based on load forecast with the latest load information. So load forecast is a basic element for demand response in Smart Grid.

Targeting to this point, this thesis introduces smart metering as the platform for load forecast in Smart Grid, and the ANN based load forecast technology in macro-grid and micro-grid.

Smart metering has attracted more and more attentions worldwide and its deployment gradually covers more districts. So it provides a platform for customers' access to load forecasting information.

With the possibility to have service of load forecast, technology for load forecast is a need. This thesis has figured out ANN-based models for macro-grid load forecast and micro-grid load forecast. Macro-grid has more stability in mapping between main influencing factors and load for unpredictable factors are ignorable under large requirement. So the thesis applies several training algorithms for Multi-layer Perceptron Neural Network, like Delta-rule, Quasi-Newton (BFGS) and Levemberg-Marquadt, and use load forecast problem of Canada Ontario province as macro-grid forecast. The final error for macro-grid load forecast problem in Canada is 0.96%.

Micro-grid is one of the new concepts introduced by Smart Grid. It constructs a significant segment of demand response integrated with load forecast techniques. Due to that micro-grid contains local feature and unpredictable influencing factors, the prediction is more complex than macro-grid. This thesis selected a chiller system load forecast in City University of Hong Kong as micro-grid load forecast problem. Based on the input-output model, MLP and RBF networks are selected for the mapping approximation. Analyses are applied on ANNs accuracy, speed, training algorithms and generalization. MLP is found to be more suitable for off-line learning as its training speed is low but higher accuracy. RBF network is better for on-line training as it has a significant small training time. Research is also placed on the more obvious noise problem in load pattern of micro-grid as compared to macro-grid. A new model type with more references points was introduced for accuracy improvement. But the micro-grid load forecast accuracy is still worse than macro-grid load forecast as there are difficulties in significantly decreasing impact due to random incidences.

Generally speaking, Artificial Neural Network is a good approximator for macro-grid load forecasting problem. Its approximating ability appears to be sufficient for the load forecast problem introduced in this chapter. But it is a model that learns the mapping appearance other than the principle inside. So once a pattern is outside the training data set occurs, the accuracy may decrease significantly. For problems that have difficulty in figuring out exact principles, like load forecast, ANN will be a good choice. But for problems which described not well with mathematics, ANN selection should be considered..

6.2 Future Work

Smart Grid is a large concept covering various areas and organizations. So Smart Grid scope design procedure may have differences due to different views. Chapter 2 introduces a general procedure with a general national development aims. But more work should be done to research on that when the aims are bias to a specific country, what direction will the characteristics and technologies change on and how much they will change. This future work is much more than a concept model of scope design but trying to find out an optimal scope design way.

Smart Grid also introduces new services and markets. Dynamic pricing will be one of the critical. In the environment of real-time price, macro-grid will face different load pattern as customers will be influenced by the price, which traditional power system do not contain. It is a pity that real-time pricing is only deployed in few countries and for specific customers like some industrial users. A large area deployment of dynamic pricing covering all type of users still does not exist. So the macro-grid load forecast for real-time price environment could only be studied at the end of Smart Grid pricing system deployments. It will be an essential future task.

For micro-grid load forecast, the significant impact from localized feature and random incidents are the main factors for low accuracy. It is difficult to apply bad data rejection as no evidence on that the abnormal data is not a section of normal pattern. To enhance the accuracy of micro-grid load forecast, possible ways are introduced as follows:

a. More data and information support on the incidents happened are needed at the abnormal points. With the new information, pattern of incidents may be selected as new input to improve the accuracy. Also the information could provide evidence for bad data rejection.
b. New problem solving model could be applied for micro-grid load forecast.

References

[1]. '*20% Wind energy by 2030: Increasing Wind Energy's Contribution*', U.S. Department Of Energy, Report, [on-line]. Available: www.nrel.gov/docs/fy08osti/41869.pdf, visited on Sep 9th 2011.

[2]. Charles Goldman, Michael Reid, Roger Levy and Alison Silverstein, '*Coordination of energy Efficiency and Demand Response*', Ernest Orlando Lawrence Berkeley National Laboratory, Report, Jan 2010, [on-line]. Available: http://eetd.lbl.gov/ea/ems/reports/lbnl-3044e.pdf, visited on Sep 9th 2011.

[3]. Eric Bonneville and Anne Rialhe, AERE, '*Demand Side Management for residential and commercial end-users*', Leonardo ENERGY, Report, [on-line]. Available: http://www.leonardo-energy.org/Files/DSM-commerce.pdf, visited on Aug 14th 2011.

[4]. Huibin Sui, Honghong Wang, Ming-Shun Lu and Wei-Jen Lee, '*An AMI System for the Deregulated Electricity Markets*', IEEE Transactions on Industry Applications, Volume 45, pp. 2014 – 2018, 2009,.

[5]. Dallinger, D., Krampe, D. and Wietschel, M., '*Vehicle-to-Grid Regulation Reserves Based on a Dynamic Simulation of Mobility Behaviour*', IEEE Transactions on Smart Grid, Volume 2, pp. 302 – 313, 2011.

[6]. '*The Smart Grid: an Introduction*', U.S. Department of Energy, Report, [on-line]. Available: www.**smartgrid**.gov/sites/default/files/pdfs/sg_**introduction**.pdf, visited on Sep 29th, 2010.

[7]. A. Battaglini, J. Lilliestam, C. Bals and A. Haas, '*The Super Smart Grid*', European Climate Forum, June 18th 2008, Report, [on-line]. Available: http://www.germanwatch.org/klima/ssg08.pdf, visited on Oct 2nd 2010.

[8]. '*Technology Roadmap Smart Grids*', International Energy Agency, Apr 2011, Report, [on-line]. Available: http://www.iea.org/papers/2011/smartgrids_roadmap.pdf, visited on May 6th 2011.

[9]. 'What is Smart Grid', IEEE Smart Grid website, [on-line]. Available: http://smartgrid.ieee.org/ieee-smart-grid, visited on Aug 5th 2011.

[10]. '*Functionalities of smart grids and smart meters, Draft Report*', EU Commission Task Force for Smart Grids, 5th Steering committee meeting Brussels, Jun 22 2010, [on-line]. Available:

http://ec.europa.eu/energy/gas_electricity/smartgrids/doc/expert_group1.pdf, visited on Apr 7th 2011.

[11]. '*Position Paper on Smart Grids, An ERGEG Conclusions Paper*', European Regulators Group for Electricity & Gas (ERGEG), Jun 10th 2010, [on-line]. Available: http://www.energy-regulators.eu/portal/page/portal/EER_HOME/EER_PUBLICATIONS/CEER_PAPERS/Electricity/2010/E10-EQS-38-05_SmartGrids_Conclusions_10-Jun-2010_Corrigendum.pdf, visited on Aug 5th 2011.

[12]. Antti Silvast and Joe Kaplinsky, '*Project UNDERSTAND White Paper on Security of European Electricity Distribution*', Jun 20th 2007, [on-line]. Available: www.understand.se/docs/White_Paper_EN.doc, visited on Jun 4th 2009.

[13]. Fang yuan Xu, Long Zhou, Yi Lin Wu and Yingnan Ma, '*Standards, Policies and Case studies in smart metering*', Power and Energy Society General Meeting, 2010 IEEE, pp. 1-5, Sep 2011, Print ISBN: 978-1-4244-6549-1.

[14]. Fang yuan Xu and Loi Lei Lai, '*A Study on Design and Functionalities of Smart Grid*', Power and Energy Society General Meeting, 2011 IEEE, pp. 1-5, Oct 2011, Print ISBN: 978-1-4577-1000-1.

[15]. '*Smart Grid System Report*', U.S. Department of Energy, Report, July 2009.

[16]. '*NIST Smart Grid Conceptual Framework*', IEEE Smart Grid website, [on-line]. Available: http://smartgrid.ieee.org/smart-grid-framework-diagram, visited on Feb 23rd, 2011.

[17]. Don Von Dollen, '*Report to NIST on the Smart Grid Interoperability Standards Roadmap*', Electric Power Research Institute (EPRI), Jun 17th 2009, [on-line]. Available: http://www.nist.gov/smartgrid/upload/Report_to_NIST_August10_2.pdf, visited on Jun 19th 2011.

[18]. Electricity Networks Strategy Group, '*A Smart Grid Vision*', Nov 2009, [on-line]. Available: http://webarchive.nationalarchives.gov.uk/20100919181607/http:/www.ensg.gov.uk/assets/ensg_smart_grid_wg_smart_grid_vision_final_issue_1.pdf, visited on Aug 2010.

[19]. '*Operation of NIST Smart Grid Conceptual Framework*', IEEE Smart Grid website, [on-line]. Available: http://smartgrid.ieee.org/smart-grid-framework-diagram/operations, visited on Feb 23rd, 2011.

[20]. '*Market of NIST Smart Grid Conceptual Framework*', IEEE Smart Grid website, [on-line]. Available: http://smartgrid.ieee.org/smart-grid-framework-diagram/markets, visited on Feb 23rd, 2011.

[21]. '*Service Provider of NIST Smart Grid Conceptual Framework*', IEEE Smart Grid website, [on-line]. Available: http://smartgrid.ieee.org/smart-grid-framework-diagram/service-provider, visited on Feb 23rd, 2011.

[22]. '*Introduction to IEEE*', IEEE Home website, [on-line]. Available: http://www.ieee.org/about/index.html, visited on Jul 5th 2011.

[23]. '*IEEE Technology Working Groups*', IEEE website, [on-line]. Available: http://www.ieee.org/about/technologies/emerging/emerging_working_groups.html, visited on Jul 5th 2011.

[24]. '*Introduction of IEEE Transactions on Smart Grid*', [on-line]. Available: http://assets.pes-store.org/pdfs/09IEEEFlyersGMREVHi-RES-9-11-09.pdf, visited on Jul 6th 2011.

[25]. '*IEEE Smart Grid Standard List*', IEEE website, [on-line]. Available: http://smartgrid.ieee.org/standards, visited on Jul 6th 2011.

[26]. '*Introduction to 2030 Smart Grid Interoperability Series*', IEEE Standards Association, [on-line]. Available: http://grouper.ieee.org/groups/scc21/dr_shared/2030/, visited on Jul 6th 2011.

[27]. IEC Home website, '*Introduction to IEC*', [on-line], Available: http://www.iec.ch/about/profile, visited on Jul 7th 2011.

[28]. '*IEC Smart Grid Standardization Roadmap*', IEC Smart Grid Strategic Group (SG3), Report, June 2010, Edition 1.0, [on-line]. Available: http://www.iec.ch/smartgrid/downloads/sg3_roadmap.pdf, visited on Jul 7th 2011.

[29]. '*Core IEC Smart Grid Standards*', IEC, [on-line], Available: http://www.iec.ch/smartgrid/standards, visited on Jul 7th 2011.

[30]. '*Introduction to ZigBee*', ZigBee Alliance, [on-line]. Available: http://www.zigbee.org/About/AboutTechnology/ZigBeeTechnology.aspx, visited on Jul 9th 2011.

[31]. '*ZigBee FAQ*', ZigBee Alliance, [on-line]. Available: http://www.zigbee.org/About/FAQ.aspx, visited on Jul 9th 2011.

[32]. '*ZigBee Member List*', ZigBee Alliance, [on-line]. Available: http://www.zigbee.org/About/OurMembers.aspx, visited on Jul 9th 2011.

[33]. 'FAQ', HomePlug Powerline Alliance Home website, [on-line], Available: http://www.homeplug.org/about/faqs/, visited on Jul 12th 2011.

[34]. '*Introduction to HomePlug AV*', HomePlug Powerline Alliance, [on-line], Available: https://www.homeplug.org/tech/tech_page2, visited on Jul 12th 2011.

[35]. 'Introduction to HomePlug Global Standards', HomePlug Powerline Alliance, [on-line], Available: https://www.homeplug.org/tech/global_standards/, visited on Jul 12th 2011.

[36]. 'HomePlug utilization in Smart Grid', HomePlug Powerline Alliance, [on-line], Available: http://www.homeplug.org/tech/app_smart_energy, visited on Jul 12th 2011.

[37]. '*Modern Grid Benefits*', U.S. National Energy Technology Laboratory, Report, Aug 2007, [on-line]. Available: http://www.netl.doe.gov/smartgrid/referenceshelf/whitepapers/Modern%20Grid%20Benefits_Final_v1_0.pdf, visited on Jul 23rd 2011.

[38]. Illinois Smart Grid Initiative, '*Summary of Smart Grid Benefits and Issues*', [on-line]. Available: http://www.cnt.org/news/media/isgi-summary-of-benefits-and-issues-6-08.pdf, visited on Jul 23rd 2011.

[39]. '*Pacific Northwest Smart Grid Demonstration Project Quarterly Newsletter*', Summer 2010, [on-line]. Available: http://www.pnwsmartgrid.org/docs/newsletter_Summer2010.pdf, visited on Jul 25th 2011.

[40]. '*Pacific Northwest Smart Grid Demonstration Project Quarterly Newsletter*', Fall 2010, [on-line]. Available: http://www.pnwsmartgrid.org/docs/newsletter_Fall2010.pdf, visited on Jul 25th 2011.

[41]. '*EPRI Smart Grid Demonstration Initiative*', U.S. Electric Power Research Institute, [on-line]. Available: http://smartgrid.epri.com/doc/EPRI%20Smart%20Grid%20Demonstration%202-Year%20Update_final.pdf, visited on Mar 17th 2011.

[42]. Vincenzo Giordano, Flavia Gangale and Gianluca Fulli, Manuel Sanchez Jimenez, '*Smart Grid projects in Europe: lessons learned and current developments*', European Commission Joint Research Centre, 2011, [on-line].

Available:
http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/22212/1/smart gridprojects-report_final_pubsy.pdf, visited on Sep 1st 2011.

[43]. '*Annual Energy Outlook 2011 With Projections to 2035*', U.S. Energy Information Administration, Reports, 2011, [on-line]. Available: http://www.eia.gov/forecasts/aeo/pdf/0383(2011).pdf, visited on Oct 2011.

[44]. '*GRID 2030, A National Vision for Electricity's Second 100 Years*', U.S. Department of Energy, Reports, July 2003, [on-line]. Available: http://www.climatevision.gov/sectors/electricpower/pdfs/electric_vision.pdf, visited on Feb 2nd 2011.

[45]. '*Smart Grid Research & Development, Multi-Year Program Plan (MYPP) 2011 - 2014*', U.S. Office of Electricity Delivery & Energy Reliability, Report, 2010.

[46]. '*Overview of the Smart Grid – Policies, Initiatives, and Needs*', ISO New England Inc., Report, Feb 17th 2009, [on-line]. Available: http://www.iso-ne.com/pubs/whtpprs/smart_grid_report_021709_final.pdf, visited on May 5th 2010.

[47]. Cecati, C., Citro, C., and Siano, P., '*Combined Operations of Renewable Energy Systems and Responsive Demand in a Smart Grid*', IEEE Transactions on Sustainable Energy, Volume 2, pp. 468 – 476, 2011.

[48]. Saber, A.Y., and Venayagamoorthy, G.K., '*Plug-in Vehicles and Renewable Energy Sources for Cost and Emission Reductions*', IEEE Transactions on Industrial Electronics, Volume 58, pp. 1229 – 1238, 2011.

[49]. Xiong Liu, Peng Wang, and Poh Chiang Loh, '*A Hybrid AC/DC Microgrid and Its Coordination Control*', IEEE Transactions on Smart Grid, Volume 2, pp. 278 – 286, 2011.

[50]. Metke, A.R. and Ekl, R.L., '*Security Technology for Smart Grid Networks*', IEEE Transactions on Smart Grid, Volume 1, pp. 99 – 107, 2010.

[51]. Moslehi, K. and Kumar, R., '*A Reliability Perspective of the Smart Grid*', IEEE Transactions on Smart Grid, Volume 1, pp. 57 – 64, 2010.

[52]. Pei Zhang, Fangxing Li and Bhatt, N., '*Next-Generation Monitoring, Analysis, and Control for the Future Smart Control Centre*', IEEE Transactions on Smart Grid, Volume 1, pp. 186 – 192, 2010.

[53]. '*Assessment of Demand Response and Advanced Metering*', U.S. Federal Energy Regulatory Commission, Staff Report, Dec 2008, [on-line]. Available:

http://www.ferc.gov/legal/staff-reports/12-08-demand-response.pdf, visited on Feb 10th 2010.

[54]. '*Annual Report on the Progress in Smart Metering 2009*', European Smart Metering Alliance, Report, Version 2.0, Jan 2010.

[55]. '*Advanced Metering Infrastructure*', U.S. National Energy Technology Laboratory, White Paper, Feb 2008.

[56]. '*Definition of Smart Metering and Applications and Identification of Benefits*', European Smart Metering Alliance, Report, May 12th 2008.

[57]. '*Report on Drives and barriers and Recommendations for action*', European Smart Metering Alliance, Report, Version 1.1, Aug 2008.

[58]. '*Smart Metering Guide: Energy Saving and the Customer*', European Smart Metering Alliance, Report, Edition 2010, Nov 16th 2009.

[59]. '*NIST Framework and Roadmap for Smart Grid Interoperability Standards*', U.S. National Institute of Standards and Technology, Report, Release 1.0, Jan 2010.

[60]. Google Smart Meter Projects Map, [on-line]. Available: http://maps.google.com/maps/ms?ie=UTF8&oe=UTF8&msa=0&msid=115519311058367534348.00000113 62ac6d7d21187, visited on Aug 4th 2011.

[61]. '*What smart meters will do for you*', BBC News on Smart Meter, [Online].Available: http://news.bbc.co.uk/1/hi/business/8023507.stm, visited on 6 December 2009.

[62]. Long Zhou, Fang yuan Xu and Ying nan Ma, '*Impact of smart metering on energy efficiency*', 2010 International Conference on Machine Learning and Cybernetics, Volume 6, pp. 3213 – 3218.

[63]. '*Impacts Analysis on European Metering Industry Stakeholders*', European Smart Metering Alliance, Report, Version 1.0, March 2008.

[64]. Benzi, F., Anglani, N., Bassi, E. and Frosini, L., '*Electricity Smart Meters Interfacing the Households*', IEEE Transactions on Industrial Electronics, Volume 58, pp. 4487 – 4494, 2011.

[65]. Saele, H. and Grande, O.S., '*Demand Response From Household Customers : Experiences From a Pilot Study in Norway*', IEEE Transaction on Smart Grid, Volume 2, pp. 102 – 109, 2011.

[66]. Parvania, M., and Fotuhi-Firuzabad, M., '*Demand Response Scheduling by Stochastic SCUC*', IEEE Transactions on Smart Grid, Volume 1, pp. 89 - 98, 2010.

[67]. Palensky, P. and Dietrich, D., '*Demand Side Management: Demand Response, Intelligent Energy Systems, and Smart Loads*', IEEE Transactions on Industrial Informatics, Volume 7, pp. 381 – 388, 2011.

[68]. Billinton, R. and Dange Huang, '*Effects of Load Forecast Uncertainty on Bulk Electric System Reliability Evaluation*', IEEE Transactions on Power Systems, Volume 23, pp. 418 – 425, 2008.

[69]. Fay, D., and Ringwood, J.V., '*On the Influence of Weather Forecast Errors in short-Term Load Forecasting Models*', IEEE Transactions on Power Systems, Volume 25, pp. 1751 – 1758, 2010.

[70]. Abedi, A., '*Signal detection in passive wireless sensor networks based on back-propagation neural networks*', IET Wireless Sensor Systems, Volume 1, pp. 48 – 54, 2011.

[71]. Suresh, S., Omkar, S.N. and Mani, V., '*Parallel implementation of back-propagation algorithm in networks of workstations*', IEEE Transactions on Parallel and Distributed Systems, Volume 16, pp. 24 – 34, 2005.

[72]. Bhotto, M.Z.A. and Antoniou, A., '*Improved Quasi-Newton Adaptive-Filtering Algorithm*', IEEE Transactions on Circuits and Systems I, Volume 57, pp. 2109 – 2118, 2010.

[73]. Saini, L.M. and Soni, M.K., '*Artificial neural network based peak load forecasting using Levenberg-Marquardt and quasi-Newton methods*', IEE Proceedings Generation, Transmission and Distribution, Volume 149, pp. 578 – 584, 2002.

[74]. Erdogmus, D., Fontenla-Romero, O.; Principe, J.C., Alonso-Betanzos, A., and Castillo, E., '*Linear-least-squares initialization of multilayer perceptron through back-propagation of the desired response*', IEEE Transactions on Neural Networks, Volume 16, pp. 325 – 337, 2005.

[75]. Savran, A., '*Multifeedback – Layer Neural Network*', IEEE Transactions on Neural Networks, Volume 18, pp. 373 – 384, 2007.

[76]. Christopher M. Bishop, '*Neural Networks for Pattern Recognition*', Oxford University Press, ISBN 978-0-19-853864-6, 2008.

[77]. Hudson, D. and Cohen, M., '*Neural Networks*', Wiley-IEEE Press, ISBN 9780470545355, 2000.

[78]. Gupta, M., Jin, L. and Homma, N., '*Continuous Time Dynamic Neural Networks*', Wiley-IEEE Press, ISBN 9780471427957, 2003.

[79]. Independent Electricity System Operator (IESO) Home website, [on-line], Available: http://www.ieso.ca/imoweb/siteShared/demand_price.asp, visited on Mar 6th 2010.

[80]. Jun Gao, '人工神经网络原理及仿真实例', Edition 2, China Machine Press, ISBN 978-7-111-12591-4, Feb 2007.

[81]. Sandhya Samarasinghe, '*Neural Networks for Applied Sciences and Engineering*', Auerbach Publications, ISBN: 978-0-8493-3375-0, 2007.

[82]. 'Introduction to Neural Network', Neural Power, [on-line]. Available: http://www.neuralpower.com/technology.htm, visited on Aug 8th 2010.

[83]. McCulloch, W.S. and Pitts, W. '*A logical calculus of ideas immanent in nervous activity*', Bulletin of Mathematical Bio-Physica, Volume 5, pp. 115, 1943.

[84]. G. Cybenko, '*Approximation by Superposition of a Sigmoidal Function*', Mathematics of Control, Signals, and Systems, Volume 2, pp. 303 – 314, 1989.

[85]. KEN-ICHI FUNAHASHI, 'On the Approximate Realization of Continuous Mappings by Neural Networks', Neural Network, Volume 2, pp. 183 – 192, 1989.

[86]. Werbos, P.j. '*The Roots of Back-propagation from ordered derivative to Neural Networks and Political Forecasting*', Wiley, ISBN 0-471-59897-6, 1993.

[87]. Agnes Gulyas, Janos Unger, and Andreas Matzarakis, '*Assessment of the microclimatic and human comfort conditions in a complex urban environment: Modelling and measurements*', Building and environment, Volume 41, pp. 1713 – 1722, 2006.

[88]. Population Density of Ontario Canada, Statistics Canada, [on-line]. Available: http://www12.statcan.ca/census-recensement/2006/as-sa/97-550/vignettes/m2o-eng.htm, visited on Dec 5th 2009.

[89]. Weather Data from National Climate Data and Information Archive, Canada, [on-line]. Available: http://climate.weatheroffice.gc.ca/climateData/menu_e.html?Prov=ON&StationI

D=9999&Year=2011&Month=8&Day=29&timeframe=1, visited on Dec 6th 2009.

[90]. Bonnans, J. F., Gilbert, J.Ch., Lemaréchal, C. and Sagastizábal, C.A., '*Numerical optimization, theoretical and numerical aspects*', Second edition. Springer, ISBN 978-3-540-35445-1, 2006.

[91]. Jorge Nocedal and Stephen J. Wright, '*Numerical Optimization*', Second Edition, Springer Verlag, ISBN 9780387987934, 2006.

[92]. K. Levenberg, '*A method for the solution of certain problems in least squares*', Quart. Appl. Math., Volume 2, pp. 164–168, 1944.

[93]. Ananth Ranganathan, '*The Levenberg-Marquardt Algorithm*', 8th June 2004, [on-line]. Available: http://www.scribd.com/doc/10093320/Levenberg-Marquardt-Algorithm, visited on Jan 2010.

[94]. Lasseter, R.H. and Paigi, P., '*Microgrid: a conceptual solution*', Power Electronics Specialists Conference 2004, Volume 6, pp. 4285 – 4290.

[95]. Patrick Chan, Wei Chun Chen, Wing NG and Daniel S. Yeung, '*Multiple Classifier System for Short Term Load Forecast of Microgrid*', International Conference on Machine Learning and Cybernetics 2011, Volume 3, pp. 1268 – 1273.

[96]. Amjady, N.; Keynia and F.; Zareipour, H., '*Short-Term Load Forecast of Microgrids by a New Bilevel Prediction Strategy*', IEEE Transactions on Smart Grid, Volume 1, pp. 286 – 294, 2010.

[97]. Microgrid Image, [on-line]. Available: https://share.sandia.gov/news/resources/releases/2006/microgrid.html, visited on Aug 20th 2011.

[98]. '*City University of Hong Kong: Calendar 2006-2007*', [on-line]. Available: http://lbms03.cityu.edu.hk/calendar/calendar_2006_2007.pdf, visited on Aug 2010.

[99]. D.S. Broomhead and D. Lowe. '*Multivariate functional interpolation and adaptive networks*', Complex Systems, Volume 2, pp. 321-355, 1988.

[100]. Poggio, T. and Girosi, F., '*Networks for approximation and learning*', Proceedings of the IEEE, Volume 78, pp. 1481 – 1497, 1990.

[101]. Lampariello, F., and Sciandrone, M., '*Efficient training of RBF neural networks for pattern recognition*', IEEE Transactions on Neural Networks, Volume 12, pp. 1235 – 1242, 2001.

[102]. Mark J. L. Orr, '*Introduction to Radial Basis Function Networks*', 1996, [on-line]. Available: http://anc.ed.ac.uk/rbf/intro/intro.html, visited on Aug 2010.

[103]. Kyung-Bin Song, Seong-Kwan Ha, Jung-Wook Park, Dong-Jin Kweon and Kyu-Ho Kim, '*Hybrid Load Forecasting Method With Analysis of Temperature Sensitivities*', IEEE Transactions on Power Systems, Volume 21, No. 2, pp. 869 – 876, May 2006.

[104]. E. A. Feinberg and D. Genethliou, '*Optimization, Control, and Computational Intelligence: Load Forecasting*', Applied Mathematics for Restructured Electric Power Systems, New York: Springer, pp. 269-285, 2005.

[105]. Ying Chen, Peter B. Luh, Chen Guan, Yige Zhao, Laurent D. Michel, Matthew A. Coolbeth, Peter B. Friedland and Stephen J. Rourke, '*Short-Term Load Forecasting: Similar Day-Based Wavelet Neural Networks*', IEEE Transactions On Power Systems, Volume 25, No. 1, pp. 322 - 330, Feb 2010.

[106]. Methaprayoon K., Wei-Jen Lee, Rasmiddatta S., Liao J.R. and Ross R.J., '*Multistage Artificial Neural Network Short-Term Load Forecasting Engine With Front-End Weather Forecast*', IEEE Transactions on Industry Applications, Volume 43, pp. 1410 – 1416, Nov. 2007.

[107]. A.Battaglini, J. Lilliestam, C. Bals and A. Haas, '*The SuperSmart Grid*', European Climate Forum, June 18th 2008.