# MEDIUM TERM LOAD FORECASTING USING STATISTICAL FEATURE SELF ORGANIZING MAPS (SOM)

**NIK NUR ATIRA BINTI NIK IBRAHIM**

**BACHELOR OF ELECTRICAL ENGINEERING**

**(INDUSTRIAL POWER)**

**UNIVERSITI TEKNIKAL MALAYSIA MELAKA**

"I hereby declare that I have read this report and in my opinion this project is sufficient in terms of scope and quality for the award of the Bachelor of Electrical Engineering (Industrial Power) with Honours"

Signature :

Supervisor‟s Name : En Zul Hasrizal bin Bohari

Date : 2nd June 2017

**MEDIUM TERM LOAD FORECASTING USING STATISTICAL FEATURE SELF ORGANIZING MAPS (SOM)**

**NIK NUR ATIRA BINTI NIK IBRAHIM**

**A thesis submitted in fulfilment**

**of the requirements for the degree of Bachelor of Electrical Engineering (Industrial Power)**

**Faculty of Electrical Engineering**

**UNIVERSITI TEKNIKAL MALAYSIA MELAKA**

**2017**

"I declare that this report entitled "Medium Term Load Forecasting Using Statistical Feature Self Organizing Maps (SOM)" is the results of my own research except cited in references. The report has not been accepted for any degree and is not currently submitted in candidate of any degree".

Signature :

Name : Nik Nur Atira binti Nik Ibrahim

Date : 2nd June 2017

Dedicates to my parents *Nik Ibrahim bin Nik Abdullah* and *Wan Noor Mawani binti Wan Ahmad*. Not forgotten my siblings *Nik Mohd Heezrad, Nik Muhamad Heedham* and *Nik Mohd Haqeem.*

## ACKNOWLEDGEMENT

Praise to Allah S.W.T The Most Gracious and The Most Merciful for His Blessing for me to finish the Final Year project 1. First of all, I would like to take this opportunity to thank my supervisor, En Zul Hasrizal bin Bohari for all guidance and encouragement for me in completing my Final Year Project.

I would like to extend my thankfulness to the most cherished persons in my life, my parents ( Nik Ibrahim bin Nik Abdullah and Wan Noor Mawani binti Wan Ahmad ) for all their moral, financial support and for never ending reminding me to always be honest and trustworthy.

Not forgotten, I would like to thank my friends (Nurul Farhana, Nor Asiilah and Nurul Natasha) for all helps and moral support during my studies in Universiti Teknikal Malaysia Melaka.

# ABSTRACT

Load forecasting is an essential tool for power system activity and planning. With increasing in development and the expansion of power system, it is important for the electrical utility to make a decision in ensuring that there would be enough supply of electricity to deal with the increasing demand. This research presents the Medium Term Load Forecasting using the artificial neural networks: Kohonen"s Self-organizing Maps. The main purpose of this project is to understand the ability Self-Organizing Maps in forecasting the load demand, and to train and test via Self-Organizing Maps method using the selected features (average temperature, K; holiday list; seasons). The data are provided by the Global Energy Forecasting Competition (GEFCom2012). This project will focus on the missing data from year 2005 and 2006 for the load forecasting. The total power and average temperature are calculated for each month in the year 2004, 2005 and 2006. The data from the year 2004 will be trained to test and forecast the data for the year 2005 while data from 2005 will be used to train for testing and forecasting the year 2006. The load data will be train, test and forecast using SOM Toolbox in MATLAB software. The accuracy of the forecasted data will be determined by calculating error of each forecasted data by comparing with the actual data. Then the Mean Absolute Percentage Error is compute to determine the accuracy of the results.

## ABSTRAK

Beban ramalan merupakan alat penting bagi aktiviti sistem perancangan dan kuasa. Dengan peningkatan dalam bahagian pembangunan dan pengembangan sistem kuasa, ia adalah penting bagi utiliti elektrik untuk membuat keputusan agar memastikan bahawa bekalan elektrik yang dibekalkan mencukupi bagi menangani permintaan yang semakin meningkat. Kajian ini membentangkan Beban Ramalan Jangka Sederhana menggunakan rangkaian neural tiruan: Kohonen Diri menganjurkan Peta. Tujuan utama projek ini adalah untuk memahami keupayaan Penganjur Peta dalam meramal permintaan beban, dan untuk melatih Kohonen Diri menganjurkan Peta menggunakan ciri-ciri yang terpilih (suhu purata, K; senarai percutian; jenis musim). Data yang disediakan oleh Ramalan Pertandingan Energy Global (GEFCom2012). Projek ini akan fokuskan kepada data yang hilang pada tahun 2005 dan 2006 bagi ramalan beban. Jumlah kuasa dan suhu purata dikira bagi setiap bulan pada tahun 2004, 2005 dan 2006. Data daripada tahun 2004 akan dilatih untuk menguji dan meramal data bagi tahun 2005 manakala data daripada tahun 2005 akan digunakan untuk melatih bagi ujian dan ramalan pada tahun 2006. Data yang akan dilatih, diuji dan diramalkan menggunakan SOM Toolbox di dalam perisian MATLAB. Ketepatan data yang diramalkan akan ditentukan dengan mengira ralat pada setiap data yang telah diramalkan dengan membandingkannya dengan data yang sebenar. Kemudian Kadar Peratusan Min Ralat digunakan dalam pengiraan bagi menentukan ketepatannya.

## TABLE OF CONTENT

# LIST OF TABLE

## LIST OF FIGURES

## ABBREVIATION

| | |
|---|---|
| ANN | Artificial Neural Network |
| SOM | Self-Organizing Maps |
| STLF | Short Term Load Forecasting |
| MTLF | Medium Term Load Forecasting |
| LTLF | Long Term Load Forecasting |
| MAPE | Mean Absolute Percentage Error |
| SVR | Support Vector Regression |

# CHAPTER 1

# INTRODUCTION

## 1.1 Motivation

Forecasting is a development of making a decision about an event which actual outcome has not yet been observed and it is the basic facet of making a decision. Load forecasting is an critical device for power system activity and planning [1], [2]. With power system development and the expansion in their complexity, various aspects have given affect to the electric power generation and consumption; such load management, energy exchange, spot pricing, and etc. In this way, the forecasting procedure has turned out to be much more complex, and more precise forecasts are required [3]. For electric utilities, it is essential to have an authentic load forecasting for variety time measure. By deregulation of energy industries, forecasting is significantly more imperative for a dispatcher to settle on a superior choice and conform to them. Consequently, electric utilities decrease events of equipment failure and power outage [4].

Load forecasting can be classified into three division which; short term, medium term and long term. The forecast for variety time horizons are essential for contrary application within a utility company and the natures of this forecast are contrast as well [5]. For the short-term load forecasting, it is important for the operation of a power system; such as unit commitment, economic dispatch, security assessment and etc. The long term load forecasting is often used in power system expansion and planning such as the construction of new power generator while the mid-term load forecasting are normally involved in the operative planning of power systems such as schedule of maintenance and power generation coordination [2].

For the medium term load forecasting, lots of variables are contributing to the load causes an exact prediction of load forecast becomes a complicated process since the variables are characterized to be a non-linear and non-stationary process. The process is complicated since the load can encounter rapid changes due to many factors and variables such as weather, seasonal and macroeconomic variations thus the load forecasting using the classical prediction models are not suitable [6]. Artificial Neural Networks (ANNs) methods are considered to be other more advanced forecasting methods which are useful for a multi-variable model. The ability of it forecast the non-linear and non-stationary load make it widely used in electricity load forecasting since 1990 [7].

The Self-Organizing Maps (SOMs) are being used since other architecture required the supervised training and do not have a favourable ability to disclose data outside of the domain of trained data. Thus the SOMs have been designed to overcome these shortcomings.

## 1.2 Problem Statement

All countries undergo the stages of economic development from „under-developed" country to „developed" country. Thus the amount of energy demand for the countries is keeping increasing over the years. It is essential for the electrical utility to make a decision in ensuring that there would be enough supply of electricity to deal with the increasing demand. It is important for the forecasting to be emphasized at all level as the aftereffect of under and over forecasting will affect all stakeholders of electricity utilities. In consequence, detailed research on forecasting method is required to forecast the load which will minimize the aftereffect of under and over forecasting.

## 1.3 Objectives

The main objectives of this research are:

1. To understand the ability Self-Organizing Maps in forecasting the load demand.
2. To train and testing via the Self-Organizing Maps method using selected features (average temperature, K; holiday list; seasons).
3. To analyse and predict the missing data and calculate the MAPE value.

## 1.4 Scope of Work

The project will focus on medium-term load forecasting using Self-Organizing Maps (SOM). The data are provided by the Global Energy Forecasting Competition (GEFCom2012). The data that had been obtained are the load history, temperature history, and holiday list. The data that had been provided are in-term of Kilowatts for the load history and Fahrenheit for the temperature history. The data that had been used for this project are from the January until December for 2004, 2005 and 2006. This project will focus on the missing data from year 2005 and 2006 for the load forecasting.

# CHAPTER 2

# LITERATURE REVIEW

## 2.1 Introduction

This chapter introduces the theoretical background for the load forecasting and the methods of load forecasting based on the past researches.

## 2.2 Theoretical Background

### 2.2.1 Load Forecasting

It is critical for electrical utilities to have an explicit data of load forecasting for the distant duration of time. With the self-regulating market of energy industries, forecasting has been essential especially for the dispatcher to make a choice and accede with them [4]. Forecasting is a process of prediction about an event which the real outcomes is yet not been occurs. Load forecasting is the divination of electrical load using the historical load data for a certain geographical area [1].

### 2.2.2 Important of load forecasting

Load forecasting acted as an essential role in helping electrical utility to make a decision which including on acquiring and generating electric power, load switching, contract evaluation, network planning and development of infrastructure [5]. It also has many applications such as in maintenance scheduling, medium term hydrothermal coordination, and competence assessment, managements of limited energy resources and agreement of progressive contracts and building of cost adequate fuel purchasing strategies [2]. Although it plays as the main factor in economic operation, the load forecasting is a two-dimensional concept which based on the consumer and load forecasting. Thus the significance for both is can be handled separately [8].

### 2.2.3 Categories of load forecasting

Load forecasting can be classified into three divisions which are short term, medium term, and long term. Although the period of the categories is not declared clearly in the research, the short-term load forecasting (STLF) covers hour to weekly forecast [8]. The STLF is important for the action of power systems such as a unit of commitment, economic dispatch, and security dispatch. The medium term load forecasting (MTLF) covers the prediction from weekly to a year. The MTLF normally used in the power system operational planning such as hydrothermal coordination and maintenance of scheduling [2]. The long term load forecasting (LTLF) normally covers the load forecasting for few years ahead and primarily purposed for bulk expansion plans, investments and corporate estimation [2], [8].

### 2.2.4 Factors and variable affect the load forecasting

The STLF usually take into account a few factors which; time influence, weather data and classification of the consumer. The MTLF and LTLF usually consider using the historical load data and climate data, customers in various classes, the appliances at a particular field and their distinctive, the economic and analytical data of the forecast, the data of appliances sales, and other factors [5].

The factors that affect the MTLF in term of load demand are the index of industrial, the price index of consumer, climate influence at that particular area and holiday influence. The weather or climate influence normally controlled by the maximum and minimum temperature, rainfall, humidity and wind speed at the particular area [9].

The electric usage or load is quite different for the each class of the customer which; residential, commercial, and industrial. The load pattern for each class has a large difference in the consumption of demand and load is much higher for the industrial compared to the residential [1], [5]. The holiday influence may affect the load usage and demand especially during major holidays [4].

### 2.2.5 Methods of load forecasting

Many forecasting designs have been refined over the last few decades. The researchers had been classified the methods into two which are econometric approach and artificial neural network (ANN) [1].

The econometric approach is using the statistical approach which usually uses the mathematical model that use the load as a function of various factors; time data, climate data, and customer classification as the load. However, the traditional econometric approach often assuming the linear relationship which adapting the functional relationship between weather factors and load demand. According to Park et al. indicates that the econometric approach might not give an accurate result due to the non-linear and non-static relationships between the load and the climate factor. Thus the ANN methods are preferable compared the econometric approach [2], [10].

The ANN can adapt and model any complicated non-linear relationship and mostly preferred to solve the load forecasting. However, the MTLF have a much difficult problem as it deals with different and various factors for a long time period. Theoretically, the MTLF can still be solved using the ANN methods since it has the non-linear problems which the ANN way could be utilized to identify any complicated non-linear relationships [10]. Figure 2.1 shows the methods of load forecasting that had been used for the STLF, MTLF, and LTLF from the past researches.

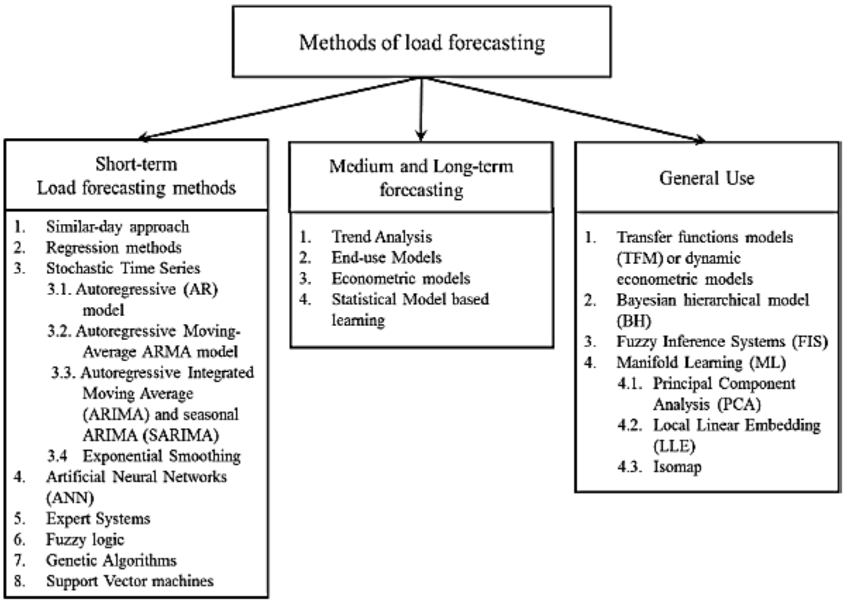


Figure 2.1: The methods for load forecasting [7].

### 2.2.6 Artificial Neural Networks

The Artificial neural network has been broadly experimental for load forecasting methods since 1990. ANNs are electronic clone based on the structure of neural in the brain as shown in Figure 2.2. The artificial neuron process are motivated by models of the neural which recognized the pattern, and then used the pattern to utilize and affect the formation of huge parallel networks, and coaching those networks to solve specific problems [8], [11]. Figure 2.2Figure 2. shows the similarities between the brain neuron with the artificial neuron as compared to Figure 2.3.

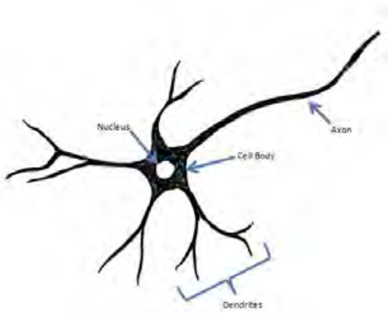


Figure 2.2: The components of neuron

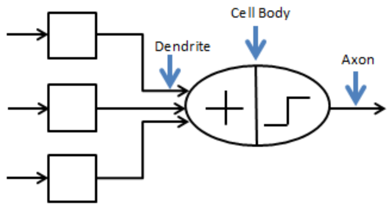


Figure 2.3: The neuron model

The ANNs is essentially non-linear circuits its output in linear or non-linear mathematical functions its inputs. The input of the data may be outputs or inputs of the other network elements. The ANNs normally have three layers as shown in Figure 2.4. The first layer is connected to the input variables known as the input layer. The third layer is connected to the output variables known as the output layer. The layer in-between the input and output layer known as the hidden layer. The hidden layer can be existed more than one layer [8], [12]. The processing elements in each layer are called nodes. Every node is connected to other neighbouring layers. Thus the parameter that involves with each connection is called weight [12].

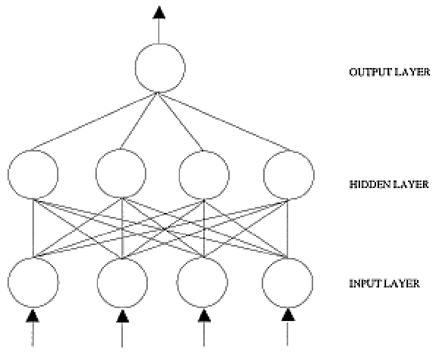


Figure 2.4: A typical ANN layer [12].

Before the network can be used, the network must learn the information beforehand. After coaching, it can be applied for pragmatic function. In broad, there are two types of learning which; supervised learning and unsupervised learning.

- Supervised learning means that the exact answer is well-known and the data is used to train the network for a given problems. This learning utilized both input and output variables. The input variables are used to accommodate initial data while the output variables can be used to differentiate with input data to determine the fault.

- Unsupervised learning means the exact answer are not known. The network needs to discover its own pattern based on input data which it purely depends on the input variables. The output generated will not use to learn from. The learning also does not need human interaction and can be handled with a broad and/or complicated dataset [13].

### 2.2.7 Classification of ANN

ANN techniques have many types of method approach. To apply the ANN technique into the load forecasting, one must choose one of the architecture approaches; Back-propagation, Hopfield, Kohonen‟s self-organizing maps, etc. [5]. The Back-propagation (1970) adopts continuously valued function and supervised learning. Through supervised learning, the real numerical weight is actuated by matching the past load data to the aspire output in a pre-operational coaching session [8].

The Hopfield (1982) introduce a chain neural network which; perform as a related memory that can recall a case from fractional or deformed version. The networks are not layered with complete attached between the nodes. The results of the network are not exactly being the function of the inputs [11]. The Kohonen‟s self-organizing maps (1982) are inspired by the self-organizing behaviour of the human brain and no supervised is required. The Kohonen‟s learn by itself through unsupervised competitive learning [13].

### 2.2.8 Introduction of Self-Organizing Maps (SOMs)

The self-organizing maps are based on the Kohonen‟s networks which developed by Tuevo Kohonen (1982). Initially, the method approach is applied to the image and sound analysis. Then later the Kohonen‟s network represents the SOM. The SOMs is a „self-organizing‟ due to its no supervision required. It learned based on its own through unsupervised ambitious learning. The „Maps‟ means that it endeavour to map its weight to comply with a given input data [13], [14].