SHORT-TERM FORECAST OF ELECTRICITY LOAD BEFORE AND DURING THE PANDEMIC COVID-19 IN MALAYSIA BY USING LEAST SQUARE SUPPORT VECTOR MACHINE (LSSVM)

SAIFUL ADLAN BIN JAFFRY

BACHELOR OF ELECTRICAL ENGINEERING WITH HONOURS
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2020

SHORT-TERM FORECAST OF ELECTRICITY LOAD BEFORE AND DURING THE PANDEMIC COVID-19 IN MALAYSIA BY USING LEAST SQUARE SUPPORT VECTOR MACHINE (LSSVM)

SAIFUL ADLAN BIN JAFFRY

A report submitted
in partial fulfilment of the requirements for the degree of
Bachelor of Electrical Engineering with Honours

Faculty of Electrical Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2020

DECLARATION

I declare that this thesis entitled "SHORT-TERM FORECAST OF ELECTRICITY LOAD BEFORE AND DURING THE PANDEMIC COVID-19 IN MALAYSIA BY USING LEAST SQUARE SUPPORT VECTOR MACHINE(LSSVM)" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : ....

Supervisor Name : SAIFUL ADLAN BIN JAFFRY

Date : 6.7.2021

APPROVAL

I hereby declare that I have checked this report entitled "SHORT-TERM FORECAST OF ELECTRICITY LOAD BEFORE AND DURING THE PANDEMIC COVID-19 IN MALAYSIA BY USING LEAST SQUARE SUPPORT VECTOR MACHINE(LSSVM)" and in my opinion, this thesis it complies the partial fulfilment for awarding the award of the degree of Bachelor of Electrical Engineering with Honours

Signature : ………………………………………………………………

Supervisor Name : DR. INTAN AZMIRA BINTI WAN ABDUL RAZAK

Date : 6.7.2021

DEDICATIONS

To my beloved mother, Noorainie binti Awang Anak and beloved father, Jaffry bin Hussain

ACKNOWLEDGEMENTS

To my supervisor, Dr. Intan Azmira Binti Wan Abdul Razak.

ABSTRACT

When there is demand for electricity, it is a must to generate it. Therefore, it is a necessity for the electric power utilities that the load on their systems should be estimated in advance. This estimation of load in advance is commonly known as load forecasting, it is essential for power system planning. Load forecast is vitally important for the electric industry and not only for deregulated economy. Since the Corona Virus Disease 2019 (COVID-19) pandemic occurs, it gave a big impact towards the world's economy mainly in Malaysia. Load forecast is essential to predict the demand after the pandemic happens. But it is found that there are no research papers had been done yet for this event and simultaneously, there has no previous work on using LSSVM with optimize Genetic Algorithm (GA) in Malaysia. The previous work for short-term load forecast in Malaysia needs to be improvise by achieving the smallest error. Thus, this project testing is held on June 2019 and June 2020 and focused on short-term forecast where it is trained on hourly data. The load data was attained from Grid System Operator (GSO) website. The goal of the result is to achieve the smallest error by using LSSVM approach by using MATLAB software. The collected data has been categorised into raw and normalized data, and has been set up into input and output forecast. The input has 24 hours of the specific day-type for three consecutive weeks starting from the 1st week of the specific month while the output will be on the 4th week. The function of GA is basically a search-based algorithm that have been used to solve optimization problems in machine learning where could solve difficult and time-consuming problems. GA optimization is to find the most optimum value of gamma and sigma of LSSVM. For the overall results, the average MAPE by using LSSVM stand-alone for 2019 and 2020 is 6.39268% and 6.63344% simultaneously. While the average MAPE for LSSVM-GA for 2019 and 2020 is 6.03614% and 4.73508% simultaneously. This shows that by using GA optimization, better accuracy is resulted.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

STLF	:	Short-term Load Forecast
MTLF	:	Medium-term Load Forecast
LTLF	:	Long-term Load Forecast
COVID-19	:	Corona Virus Disease 2019
MCO	:	Movement Control Order
SVM	:	Support Vector Machine
LSSVM	:	Least Square Support Vector Machine
GA	:	Genetic Algorithm
TNB	:	Tenaga Nasional Berhad
MAPE	:	Mean Average Percentage Error
MAE	:	Mean Average Error
GSO	:	Grid System Operator
ARMA	:	Autoregressive Moving Average
ARIMA	:	Autoregressive Integrated Moving Average
ARM\AX	:	Autoregressive Moving Average with Exogenous Variables
ARIMAX	:	Auto Regressive Integrated Moving Average with Exogenous Variables
SSEs	:	Sum Square Errors
QP	:	Quadratic Programming

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter reveals the introduction of the study following by the project background and the problem statement. According to the project background, an explanation of the electricity load in Malaysia has been described. The project's problem statement, objectives and scopes are discussed in this chapter. All the details for each section of the project have been discussed in chapter one.

1.2 Motivation

The aim of electricity load forecasting is to meet and foresee demand in the future that may rise or fall. Precise and exact energy demand predictive models rely on a variety of engineering applications. Precise forecasts of load allows to plan power supplies' capability and activities to deliver electricity effectively to consumers [1]. The power sector includes electricity generation, storage, delivery and sales to the public and industry. In 1882, when electricity was generated for electric lighting, the commercial distribution of electric power began. By the middle of the 20th century, the electricity was seen as a "natural monopoly" only if a limited number of firms were present on the market. Vertically integrated corporations provide all phases from generation to retail in certain regions, and only governmental supervision regulated the rate of return and cost structure. In order to have a more competitive energy market, several regions have split up the generation and distribution of electric power since the 1990s. Electricity is a product capable of being purchased, sold and exchanged economically. One of the ways for electricity retail is by using load forecast.

Load forecasting allows an electric utility to make critical decisions, including decisions on the purchase and generation of electricity, load switching and the construction of infrastructure. Load forecasts are a tool used by electricity or energy providers to estimate the power necessary to satisfy demand and supply balances. Accurate electrical load forecasting is a crucial issue for resource of electrical power generation utilities. There are three types of load forecast which is short-term load forecast (STLF), medium-term load forecast (MTLF) and long-term load forecast (LTLF). STLF covers a period from minutes, hours and a month. MTLF covers from a month up to a year and LTLF covers one year and above. Inaccuracy of load forecast have negative impact on the economics of power generation utilities. In Malaysia, Tenaga Nasional Berhad (TNB) is Malaysia's international electricity corporation and is Peninsular Malaysia's only electricity company and also Southeast Asia's biggest publicly traded power company. TNB also monopoly the electricity transmission and distribution in Peninsular Malaysia and Sabah and have significant electricity generation capacity through its subsidiaries. But since the Corona Virus Disease 2019 (COVID-19) pandemic in Malaysia happens, there has been a massive effect towards the country. The Ministry of Health Malaysia has taken a rapid response to prevent this virus to be spread continuously and encouraged the Federal Government to implement the 2020 Movement Control Order (MCO). One of the orders under the 2020 MCO is the workers need to work at their own residence and only attend the office a few times in a week which also effected the electrical usage of other companies. Hence, has been a difference of electrical usage before and after the pandemic hits Malaysia. Accurate load forecast is needed to predict the electrical demand.

1.3 Problem Statement

For the past years, many research about load forecast has been done in Malaysia's market. With this, a lot of knowledge and innovation can be discovered. Since the COVID-19 pandemic strikes Malaysia, a lot of things have been affected majorly on the country's economic. But it is known that there has no research has been done for short-term load forecast during COVID-19 pandemic in Malaysia. Thus, it is unknown on how big the impact of the pandemic towards electricity load demand in Malaysia.

After searching for research papers about short-term load forecast in Malaysia, it can be concluded that the previous work for short-term load forecast in Malaysia needs to be improvise by achieving the smallest error and there is no previous work on using LSSVM with GA in Malaysia. GA uses optimization technique which mimic the process of natural evolution. It is an effective and efficient technique for machine learning applications and are used widely in business and disciplines for science and engineering. It is also a stochastic learning machine and uses probabilistic transition rules.

1.4 Objectives

The purposes of this project are:

i. To develop accurate and efficient load forecast models using Least Square Support Vector Machine (LSSVM) to predict loads in Malaysia before and during COVID pandemic.
ii. To improve load prediction accuracy by optimizing LSSVM parameters using Genetic Algorithm.

1.5 Scope

This project mainly focuses on:

i. The implementation of LSSVM technique as the main forecast engine.
ii. The optimization process is held by Genetic Algorithm (GA) to optimize the parameters of LSSVM.
iii. The input used for day-ahead forecast is historical load from Grid System Operator (GSO).
iv. The output data forecast for day-ahead 24-hour load.
v. The training period for day-ahead forecast is from March 2019 until May 2019 and March 2020 until May 2020 prior to the testing period.
vi. The testing period is June 2019 and June 2020.
vii. The objective function of the model is Mean Absolute Percentage Error (MAPE) and MAE; where minimum error should be achieved to obtain good predictive accuracy.
viii. The model developed is tested in Malaysia electricity market.
ix. Day-ahead forecast using 5 day-type model and nine training samples and four testing samples was used for each day-type.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter is mainly on the review of literature from the journal, article and other resources. A brief understanding of the load forecast, Support Vector Machine, Least Square Support Vector Machine and Genetic Algorithm had been done in the next sub-chapters.

2.2 Load Forecast

Load predictions are a tool that power companies use to forecast the load necessary to satisfy request. It is designed as an integral part of the power system operation. Load forecasting can be divided as very short-term load forecast, short-term load forecast, medium-term load forecast and long-term load forecast [1]. Load prediction is a prediction method for future load requests. Precise load prediction models are important to the operation and planning of a utility corporation. Many models of forecasting have limited forecasts when dealing with the sensitivity of temperature, special events and dynamic calendar consequences like weekends and holidays [2]. This impacts the economic effects of the electricity industry and for the energy sales business and the users' final tariff [3]. In common, various types of simulation models such as conventional time series models are used for prediction purposes [4]. For the past few years, there are many different techniques have been constructed for electricity demand forecasting. Numerous reports of the use of artificial material have carried out about load forecasting intelligence (AI) techniques where AI techniques worked better than the traditional approaches in forecast for short-term load [5]. Among of the AI technique that is found in the literature are expert systems, neural network, fuzzy-neural models and fuzzy inference [5]. There are some studies included multi-phase optimization for the selection of parameter such as Evolutionary Algorithm was used to optimize the neural network parameters by three level of optimization [6].

There is various method for the similar day approach are used for short-term load forecasting regression models such as neural network [13], time series, fuzzy logic [16] and many more. Similar day approach is a method which focused on the search of historical data for the day of the prediction in one, two or 3 years. Regression techniques are typically used for electrical load prediction to model the load usage relationship and other variables, such as temperature, time of day, and customer class. On the other hand, time series method has been applied to a variety of different field in this few and the most common method are the ARMA, ARIMA, ARM\AX, and ARIMAX. Fuzzy logic is a widespread Boolean program used in computer circuit modeling. The feedback has a certain qualitative spectrum under fuzzy logic. Fuzzy logic is one of a variety of ways to map inputs to outputs. Fuzzy logic gives the benefits of the lack of a mathematical prototype input mapping to outputs and the lack of reliable or noiseless inputs [7].

2.3 Support Vector Machine (SVM)

SVM is a supervised research model that uses pattern recognition and analyzed data for the purpose of categorization and approximation [6], as reviewed by [8]. Suppose an empirical data is described (1):

$$[(x_1, y_1), \ldots, (x_m, y_m)] \in X \times \Re; \mathrm{X} = \Re^d \qquad (1)$$

The space of the input patterns was represented as X, while in (2) shows the linear function f:

$$f(x) = \langle w,\ x \rangle + b \text{ where } w \in X, b \in \Re \qquad (2)$$

The purpose of Support Vector Regression is to solve quadratic programs that limit inequalities[6]. That being said, SVM has a broad machine challenge, which describes the problem of optimization (3):

$$\min \frac{1}{2}\|\omega\|^2 + C\sum_{k=1}^{N}\left(\xi_k + \xi_k^{\phi}\right)$$

Where ξ is slack variable

$$\text{subject to}\begin{cases} y_k - \langle \omega, \phi(x_k)\rangle - b \le \varepsilon + \xi_k \\ \langle \omega, \varphi(x_k)\rangle + b - y_k \le \varepsilon + \xi_k^{\phi} \\ \xi_k, \xi_k^{\phi} \ge 0 \end{cases} \quad (3)$$

while the ε-insensitive loss function is represented as (4):

$$|\mathrm{y} - \mathrm{f}(x,\omega)|_{\varepsilon} = \begin{cases} 0, if\,|y - f(x,\omega)| \le \varepsilon \\ |y - f(x,\omega)| - \varepsilon, \text{otherwise} \end{cases} \quad (4)$$

Cost of error with constant C > 0 is the benchmark for the balance between maximization of margins and minimization of training errors. Finally, the following SVM would be used as in (5) for the non-linear feature evaluation:

$$f(x) = \sum_{k=1}^{N}\left(\alpha_k + \alpha_k^{\phi}\right)\langle \varphi(x), \varphi(x_k)\rangle + b \quad (5)$$

SVM can minimize local minima issues, overfitting, and can work brilliantly with broad input areas [6]. SVM can guarantee optimum solutions overall since the problems of optimization described by it are a usually convex quadratic programme (QP). One argument regarding the SVM is to be dealt with is that the regression model's coefficients of variables are sensitive towards the SVM’s structural parameters[9].

2.4 Concept of Least Square Support Vector Machine (LSSVM)

LSSVM was suggested to reduce the burden of computing From the SVM that refers to equality rather than limits on inequalities[10]. Instead of a quadratic programming (QP) problem, LSSVM solves a linear equation method that enhances computational speed. Karush-Kuhn-Tucker (KKT) or commonly known as the linear system that is straightforward compare to QP system. Although LSSVM still uses the fundamental of SVM, it can reduce the sum square errors (SSEs) of training set and minimize margin error at the same time[6]. The following (6) is the LSSVM formula that uses the least square loss function:

$$\min J(\omega, e) = \frac{1}{2}\|\omega\|^2 + \frac{1}{2}\gamma \sum_{k=1}^{N} e_k^2 \tag{6}$$

$$\text{subject to } y_k = \begin{cases} \langle \omega, \varphi(x_k) \rangle + b + e_k, \\ k = 1, \ldots, N \leq \varepsilon + \xi_k \end{cases}$$

$e_k \in R$ are error variables and $\gamma \geq 0$ is a clustering constant that establishes the balance between minimizing the fitting error and flexibility of the estimated function. As implemented in Lagrangian (7):

$$\mathrm{L} = \frac{1}{2}\|\omega\|^2 + \frac{1}{2}\gamma \sum_{k=1}^{N} e_k^2 - \sum_{k=1}^{N} \langle \varphi, \varphi(x_k) \rangle + b + e_k - y_k \tag{7}$$

The conditions for optimality are shown in (8):

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \rightarrow \omega = \sum_{k=1}^{N} \alpha_k \, \varphi(x_k) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \omega = \sum_{k=1}^{N} \alpha_k = 0 \\ \frac{\partial L}{\partial e_k} = 0 \rightarrow \alpha_k = \gamma e_k, (k = 1, \ldots, N) \\ \frac{\partial L}{\partial \alpha_k} = 0 \rightarrow \omega = \langle \omega, \varphi(x_k) \rangle + b + e_k - y_k \end{cases} \tag{8}$$

The following equation was verified after the exception of ω and e (9):

$$\begin{bmatrix} 0 & 1_v^T \\ 1_v & \Omega + \frac{1}{\gamma} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{9}$$

where $y = [y_1; \ldots; y_N]$, $1_v = [1; \ldots; 1]$, $\alpha = [\alpha_1, \ldots \alpha_N]$. The LSSVM model for regression becomes (10):

$$f(x) = \sum_{k=1}^{N} \alpha_k K(x, x_k) + b \qquad (10)$$

2.5 Concept of Genetic Algorithm (GA)

The Darwin's theory of evolution and Mendel's genetic theory is the bases for GA, that is a stochastic approach which imitates natural evolution organisms[11]. Genetic algorithms are still widely used for STLF recently. It also improves on concepts of approaches for neural networks [12]. GA approaches are models that emulate biological advances and the concept of survival for a formal random exchange of information[13]. After some iterative simulations, the optimum solution can be identified. GA contains three major operations which are the selection, crossover and mutation. The method of optimisation is initiated by random chromosome population accompanied by fitness evaluation [6]. Afterwards, the selection of the most suitable individuals or parents for reproduction. The most suitable chromosomes are the ones that have better fitness value because they have higher potential to produce children during subsequent generation. The best chromosomes is capable for information exchange by crossover and mutation. The purpose of it is to produce offspring chromosome during the reproduction phase[14].

After the size of population is gained, the most fittest parent will undergo crossover phase with other parent. This is when parts of two genotypes are exhanged and the normal rate for the crossover is from 0.6 untill 1.0[6]. Mutation will take place after crossover is performed. Mutation is an uncommon process where could bring big and small changes towards the offspring when randomly select any parent of the chromosome. [14]. Mutations are performed randomly when a "1" bit has been transformed into a "0" bit and vice versa[6]. The rate of the mutation is commonly 0.001[15] or not more than 0.1[14]. There are four main specifications for GA performance which is population size, number of generations, and the rate of crossover and mutation [14].