Faculty of Computer Science and Information Technology

FAILURE PREDICTION OF ENGINEERING PROBLEMS USING INTERACTIVE COMPUTING NOTEBOOK ENVIRONMENT

FLORINA LING ANAK CASTELO

Bachelor of Computer Science with Honours

(Computational Science and Mathematics)

FAILURE PREDICTION OF ENGINEERING PROBLEMS USING INTERACTIVE COMPUTING NOTEBOOK ENVIRONMENT

FLORINA LING ANAK CASTELO

This project is submitted in partial fulfillment of the requirements for the degree of

Bachelor of Computer Science and Information Technology

Faculty of Computer Science and Information Technology
UNIVERSITI MALAYSIA SARAWAK
2020

RAMALAN KEGAGALAN DARIPADA MASALAH KEJURUTERAAN MENGGUNAKAN PERSEKITARAN PENGKOMPUTERAN NOTEBOOK YANG INTERAKTIF

FLORINA LING ANAK CASTELO

Projek ini merupakan salah satu keperluan untuk Ijazah Sarjana Muda Sains Komputer dan Teknologi Maklumat

Fakulti Sains Komputer dan Teknologi Maklumat
UNIVERSITI MALAYSIA SARAWAK

2020

UNIVERSITI MALAYSIA SARAWAK

THESIS STATUS ENDORSEMENT FORM

TITLE: FAILURE PREDICTION OF ENGINEERING PROBLEMS USING INTERACTIVE JUPYTER NOTEBOOK

ACADEMIC SESSION: 2019/2020

FLORINA LING ANAK CASTELO

hereby agree that this Thesis* shall be kept at the Centre for Academic Information Services, Universiti Malaysia Sarawak, subject to the following terms and conditions:

1. The Thesis is solely owned by Universiti Malaysia Sarawak
2. The Centre for Academic Information Services is given full rights to produce copies for educational purposes only
3. The Centre for Academic Information Services is given full rights to do digitization in order to develop local content database
4. The Centre for Academic Information Services is given full rights to produce copies of this Thesis as part of its exchange item program between Higher Learning Institutions [or for the purpose of interlibrary loan between HLI]
5. ** Please tick (√)

[ ]	CONFIDENTIAL	(Contains classified information bounded by the OFFICIAL SECRETS ACT 1972)
[ ]	RESTRICTED	(Contains restricted information as dictated by the body or organization where the research was conducted)
[√]	UNRESTRICTED	

Validated by

florinaling   Phang

(AUTHOR'S SIGNATURE)   (SUPERVISOR'S SIGNATURE)

Permanent Address

No.,698,KampungTanjongPoting,
Singai,94000 Bau
SARAWAK.

Date: 06/08/2020   Date: 06/08/2020

Note * Thesis refers to PhD, Master, and Bachelor Degree
** For Confidential or Restricted materials, please attach relevant documents from relevant organizations / authorities

DECLARATION

I hereby declare that this thesis with the title of "Failure Prediction of Engineering Problems Using Interactive Computing Notebook Environment" is my research work except for some information that quoted and cited from other resources which have been acknowledged. I also declare that no portion of the work that referred to in this report has been submitted in support of application for another degree at University Malaysia Sarawak (UNIMAS).

Signature:

florinaling

Author Name: Florina Ling anak Castelo

Date: 06.08.2020

ABSTRACT

In recent years, research has proposed several machine learning (ML) approaches to predict remaining useful life (RUL) in engineering field which involved computer science skills. This paper proposed to predict turbofan engine remaining useful life (RUL) based on the engine historical degradation data provided by NASA C-MAPPS. CMAPPS is a tool stands for 'Commercial Modular Aero-Propulsion System Simulation' to simulate realistic large commercial turbofan engine data. Prediction model built based on regression problem using dimensionality reduction method and regression algorithms. Dimensionality reduction would extract only important features for more accurate prediction. Model performance is dramatically affected by the algorithm robustness which are the basis of this thesis. The efficiency of the model is evaluated using Pearson correlation coefficient. Results showed regression model could give a satisfactory prediction result based on the test data provided by CMAPPS. The effectiveness of the methodology for early prediction provides alert in machine degradation before it reaches failure. This efficient procedure could prevent severe failure occurrence and maintenance costs.

ABSTRAK

Sejak kebelakangan ini, kajian telah melaksanakan beberapa pendekatan pembelajaran pengkomputeran (ML) untuk meramalkan baki hayat berguna (RUL) dalam bidang kejuruteraan yang melibatkan kemahiran sains komputer. Projek ini mencadangkan untuk meramal baki hidupan enjin kipas turbo (RUL) berdasarkan data degradasi enjin yang disediakan oleh NASA C-MAPPS. CMAPPS adalah kaedah yang bermaksud 'Commercial Modular Aero-Propulsion System Simulation' untuk mensimulasikan data realistik kipas turbo komersial yang besar. Model ramalan dibina berdasarkan masalah regresi menggunakan kaedah pengurangan dimensi dan algoritma regresi. Pengurangan dimensi akan hanya mengekstrak ciri-ciri penting untuk ramalan yang lebih tepat. Prestasi model dipengaruhi secara mendadak oleh kekukuhan algoritma yang menjadi asas kepada tesis ini. Kecekapan model dinilai dengan menggunakan 'Pearson correlation coefficient'. Hasil kajian menunjukkan model regresi boleh menghasilkan ramalan yang memuaskan berdasarkan data ujian yang disediakan oleh CMAPPS. Keberkesanan metodologi untuk ramalan awal memberi amaran dalam degradasi mesin sebelum mencapai kegagalan. Prosedur yang cekap ini dapat mencegah terjadinya kegagalan enjin dan kos penyelenggaraan.

ACKNOWLEGMENT

This project work has been developed to meet the academic requirements of University Malaysia Sarawak (UNIMAS) for the completion of Bachelor of Degree (Hons) in Computer Science. First, I would like to send my utmost gratitude to my final year project supervisor Dr. Phang Piau for the invaluable, guidance, comments and suggestions throughout the course of the project. Next, I would like to thank our Final Year Project coordinator, Professor Dr Wang Yin Chai for providing guidelines and coordination throughout the conduction of Final Year Project. I would also want to specially thank for my parent's support for providing me a lot of emotional support for completing this final year project and put tremendous faith in me. Finally, thank you to all my friends, for the moral support I have received for the past four years of my studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Project Title

Failure Prediction of engineering problems using Interactive Computing Notebook Environment.

1.2 Introduction

The research was prepared to be focusing based on NASA Engine Turbofan Degradation Simulation Dataset. A complex machine should be undergone maintenance proactively to reduce maintenance cost, prevent extensive unscheduled repair time and better availability of the engine. Inappropriate maintenance time could contribute to the increased of deterioration rate and maintenance cost. Predictive maintenance (PdM) is on the rise as the most doable maintenance strategy to solve such problems. PdM is just a concept of fixes things before it needs fixing. It is not a tool nor a technique. PdM conducts two main tasks which is diagnosis and prognosis to predict machine failures. However, due to the time restriction, this paper will only focus on the machine prognosis. Machine learning is a robust tool for analysing critical industrial activity involving machine. From the historical data of the turbofan engine, machine learning is the most feasible techniques to distinguish failure patterns. Specialty of a machine learning is it could estimate how much time remaining until the machine fail. This could help in scheduling maintenance earlier before the machine degrades.

1.3 Problem Statement

All machines are vulnerable to wear and tear. High cost of maintenance is the most common major problem in engineering sectors. Not only that, engineering sectors also experienced unexpected failures. Unexpected failures occur due to the unfeasible scheduling time. This is because the inaccuracy in the measurements of maintenance time. Hence, it is important to apply in the application of remaining useful life. In engineering, remaining useful life (RUL) is the length of time a machine is likely to operate before it requires repair or replacement. Sensor devices are commonly used to monitor the condition of engineering assets. It supplies data that can be used to predict when the asset will require maintenance and prevent equipment failure. Therefore, failure prediction is essential for predictive maintenance due to its ability to prevent failure occurrences and maintenance costs.

1.4 Scope

The scope of this project focuses on predicting the Remaining Useful Life (RUL) of turbofan engine using regression approach and machine learning techniques.

1.5 Objectives

The objectives of this project are:

(1) To explore the selected engineering dataset by using data manipulation and visualization tools in Jupyter notebook.

(2) To analyse engineering dataset by using statistical and machine learning tools.

(3) To build prediction model for Remaining Useful Life prediction of test data.

1.6 Brief Methodology

Overall methodology approach for conducting this research is shown in Figure 1.1 below. Exploratory analysis will be implemented to understand the nature of the dataset. Modelling approach is by applying Machine Learning method. Type of Machine Learning performed for this research is Supervised Learning. Supervised Learning is very suitable for calculating estimated lifetime value. The methodology proposed is based on the degradation data recorded from the sensor. A data-driven model will be used to predict future time failure. The datasets are contaminated with noise thus some techniques will be conducted to filter out the noise. Selected sensor data used reflects in accuracy measurement of the engine condition, thus, good model will be deployed and integrate. Efficient filtering method for selecting the suitable features will enhance the robustness of the built model.

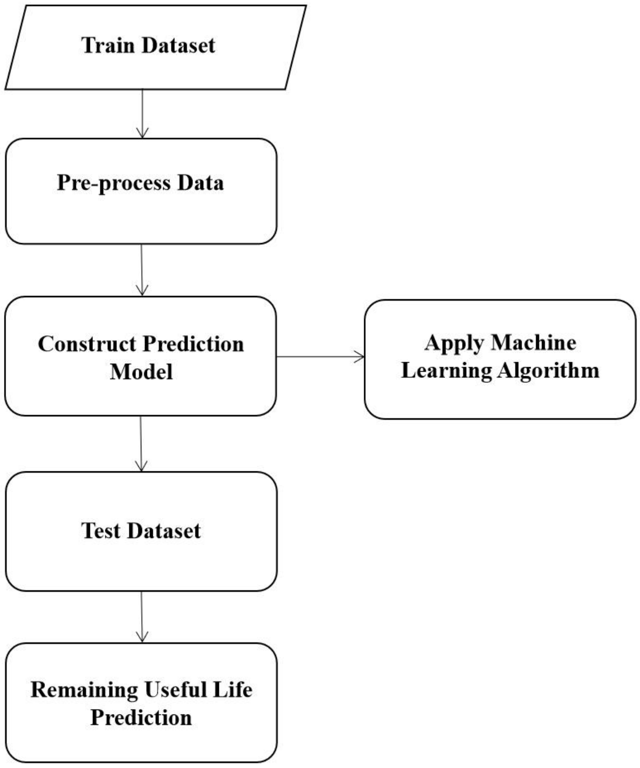

Figure 1.1: General flowchart for RUL prediction. (Mathew, Toby, Singh, Rao, & Kumar, 2018)

1.7 Significance of Project

Predictive maintenance is one of the promises solutions to address the issue with machine reliability leading to cost savings and a reduction of unplanned failures. Through data exploration and investigation, this project could predict data RUL using regression approach. The findings of this research are to predict engine remaining useful life using the engine historical degradation data. By conducting this project, we can schedule maintenance in advance for the engine before it reaches failure.

1.8 Project Schedule

Refer Appendix A.

1.9 Expected Outcome

Outcome is to be able to predict remaining useful life (RUL) of machine before the next failure occurs should be achieved in test dataset. The RUL will be in cycles as it is a multivariate time series.

1.10 Thesis Outline

Chapter 1: Introduction

Chapter 1 of the report propose project title and overall introduction of the project. This followed by problem statement, scope, aim, objectives, brief methodology, significance of project, project schedule and expected outcome behind this research project.

Chapter 2: Literature Review

Literature review is delivered in this chapter which introduce suitable machine learning approach utilize for machine failure prognosis and estimating remaining useful life (RUL). In this chapter, overview for previous researchers' state of art involved supervised learning algorithm, feature selection method and model learning method will be discussed.

Chapter 3: Methodology

Chapter 3 will be focusing on the details of project methodology, requirement analysis and design. This chapter provide detailed explanation of the proposed technique. Dataset applied to this project will be further explained and illustrated using diagrams and tables.

Chapter 4: Data Modelling

Chapter 4 provides a detailed explanation about implementation of the proposed project framework. Environment setup with software requirements, results of data preparation after feature selection and modelling the data were provided in this section.

Chapter 5: Results and Analysis

Chapter 5 indicated the results of regression analysis of base model and regularized model. A comparison made between actual and predicted results. The final accuracy results presented after pass through model evaluation using Pearson Correlation Coefficient and Mean Absolute Error.

Chapter 6: Conclusion

Chapter 6 explains about the contribution, limitation and overall conclusion of the project has been completed. The future works to improve the prediction model accuracy also discussed in this chapter.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter explained the related studies and researches related to predictive maintenance on machine learning techniques for predicting remaining useful life of turbofan engine. Predictive maintenance is the method of scheduling maintenance based on the prediction about the failure time of any equipment (Mathew et al., 2018). Predicting failure time or RUL conducted in this thesis by inspecting the historical data. Machine failure prediction are separated into two problems called regression and classification problem (Jardine, et al., 2006 according to Al-Busaidi, 2018, page 6). Thus, this section only focuses on supervised machine learning instead of unsupervised learning nor reinforcement learning.

2.2 Background

In this section, it is an overview of the project. Turbofan engine degradation datasets are acquired for the project research. The engine is used in any aircraft shown in Figure 2.1. Figure 2.2 shows where the airplane turbofan engine located and it operates with the aid of air by sucking air into the front engine through a fan. The engine is then compressing the air by mixing the fuel with it causes the fuel or air mixture to ignites and channel it out from the back of the engine thus producing thrust for the aircraft to move. Vast outspread of sensors and enormous amount of equipment monitoring data is relatively effortless to obtain which drive the data-driven RUL prediction method feasible. For aircraft engines, the RUL can be taken as the rest number of flight cycles from the current moment to the failure of the engine performance. Aviation company can enhance maintenance schedules to prevent severe damage by estimating the RUL of the turbofan engine. The precise and well-timed prediction of RUL is a vital requirement for guaranteeing the safe operation of aircraft. Maintenance of these engines are very costly yet significantly important as engine downtime can lead to catastrophes. Information are collected from the modern equipment that are fitted with sensors whereby large storage devices operates to capture and stored huge amount of data. All these events were simulated by NASA CMAPPS to obtain the turbofan engine simulated data applied in this paper. Machine learning algorithms are applied to train the prediction model to learn the simulated data and model accuracy is tested using the train simulated data for model validation.

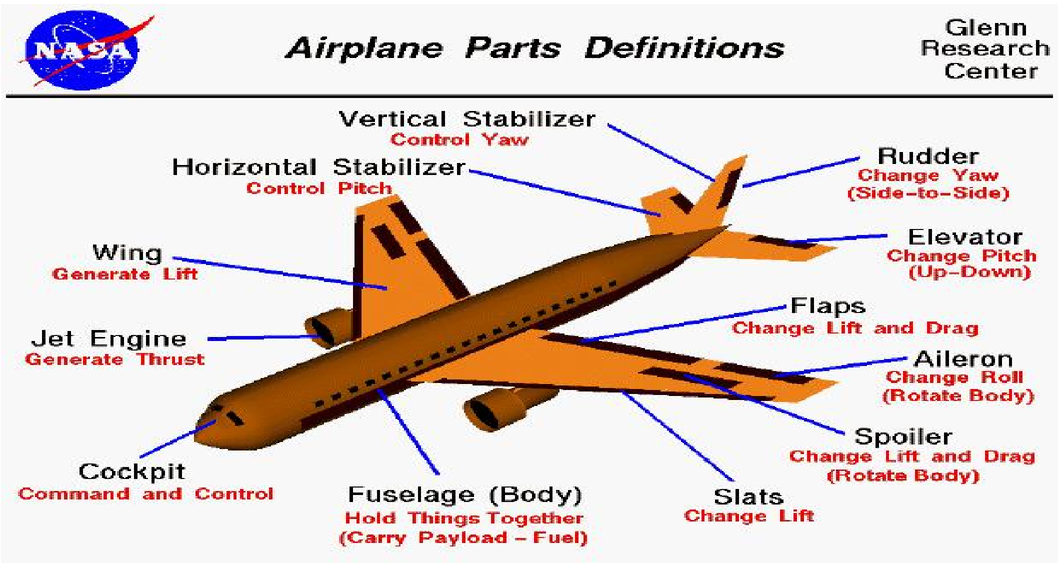

Figure 2.1: Overview of an airplane parts. (“Parts of Airplane,” n.d.)

Figure 2.2: Location of airplane turbofan engine. (“How Does A Turbofan Engine Work? | Boldmethod,” n.d.)

2.3 Related Research

This section presents the overview for state of art about existing predictive maintenance. There are two ways of doing predictive maintenance which are regression approach and classification approach (Mathew et al., 2018). For regression approach, it is to figure out relationship between variables mainly used in prediction. Classification approach is to classify data into its category or class. The techniques used for both approaches are feature selection method and suitable algorithm to build the prediction model for RUL prediction. Table 2.1 summarize the comparisons among the techniques applied.

2.3.1 Regression Approach

Regression approach is where it predicts the time left before the next failure called Remaining Useful Life (RUL) (Mathew et al., 2018). To solve regression problem, regression techniques vary according to the number of independent variables, type of relationship between independent and dependent variables. (Khelif et al., 2017) proposed using direct approach for RUL prediction based on the support vector regression (SVR)-RUL model. Turbofan engine degradation dataset known for its noisy data. Therefore, those noisy sensory data are smooth by applying local regression using weighted linear squares and second-degree polynomial model. Each sensor data treated independently and had its own smoothing parameter. Dealing with the denoised large volume of data is not efficient to build a model as it might produce overfitting. To solve this issue, only important features are selected from the data using wrapper selection method. There two other methods such as filters method and embedded methods but wrappers method is the most suitable for this literature as it is easy to be implemented. The sensor data are selected according to sensors different combinations score and performance.

Figure 2.3 shows the feature selection method using SVR-wrapper (Khelif et al., 2017). To build the model, SVR has two existing ways to execute a very efficient model mainly in modelling nonlinear relationships between target feature and several independent features. SVR has $v-SVR$ and $\in -SVR$. $v-SVR$ has parameter v which derived the distribution of support vectors number with respect to the total number of samples. As for $\in -SVR$, the parameter $\in$ act to manage the error in the model. According to the literature objective, $\in -SVR$ was implemented to search for a function that satisfy the $\in$ margin. To handle this