AUSTRALIAN NATIONAL UNIVERSITY

ANU College of Engineering and Computer Science

R E P O R T

on

Practical Experience
ENGN3100

by

Shu Liu

U5764321

2/7/2019

This Report is submitted in partial fulfilment of the Requirements for Practical Experience for the BE Degree in the ANU College of Engineering & Computer Science

TABLEOFCONTENTS

1. Summary of practical experience

Name of Employer:

Iluvatar CoreX Inc.

Employment period:

Stage1: 18/12/2017-10/2/2018

Stage2: 6/2018-7/2018

Stage3: 23/11/2018-15/2/2019

Position/job:

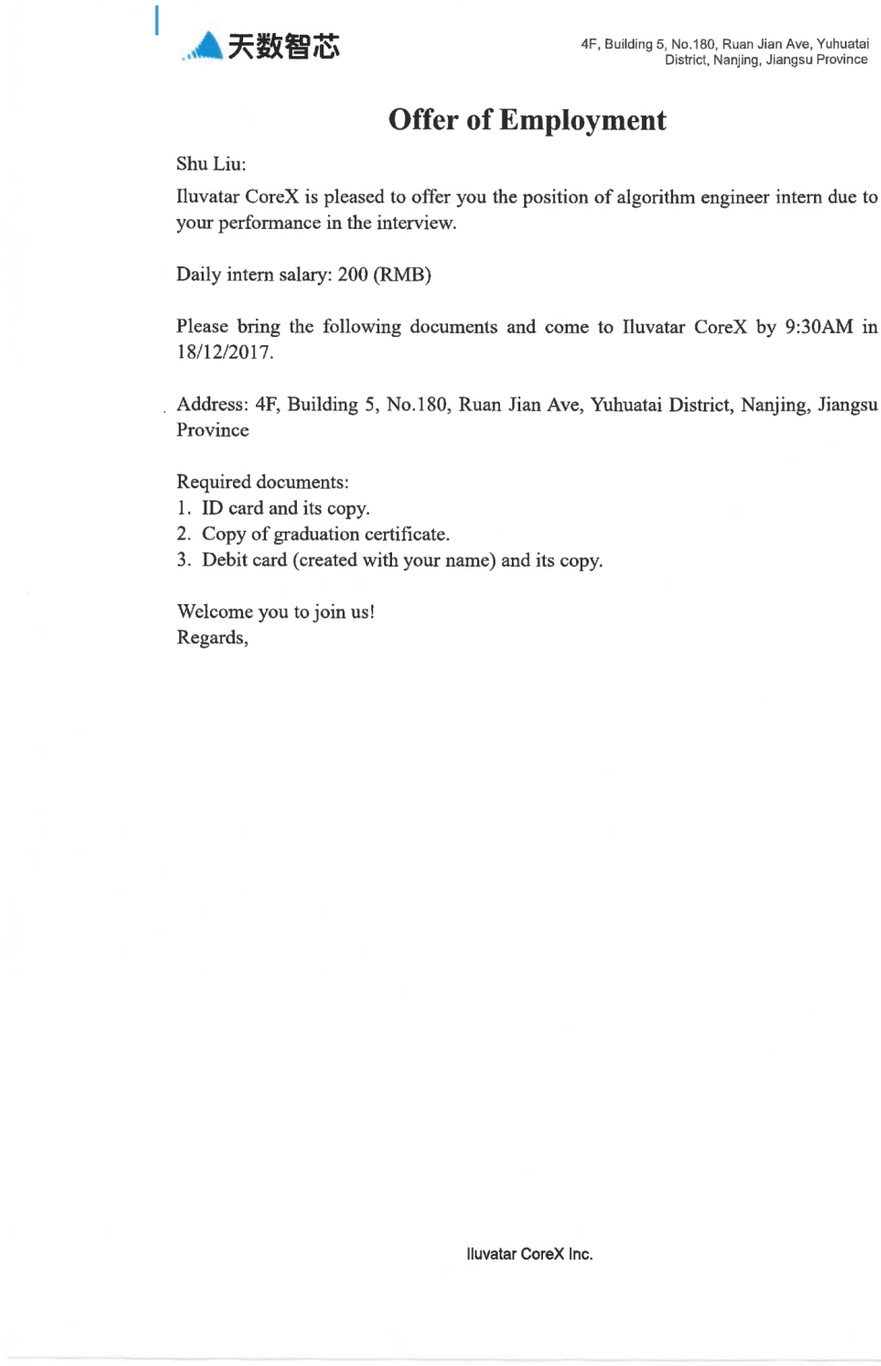
Algorithm engineer (Intern)

Total number of weeks of experience claimed:

18 weeks

2. Copy of Signed Letter of Offer of Employment

(The offer letter is attached in the following page)



天数智芯

4F, Building 5, No.180, Ruan Jian Ave, Yuhuatai District, Nanjing, Jiangsu Province

Offer of Employment

Shu Liu:

Iluvatar CoreX is pleased to offer you the position of algorithm engineer intern due to your performance in the interview.

Daily intern salary: 200 (RMB)

Please bring the following documents and come to Iluvatar CoreX by 9:30AM in 18/12/2017.

Address: 4F, Building 5, No.180, Ruan Jian Ave, Yuhuatai District, Nanjing, Jiangsu Province

Required documents:
1. ID card and its copy.
2. Copy of graduation certificate.
3. Debit card (created with your name) and its copy.

Welcome you to join us!
Regards,

Iluvatar CoreX Inc.

Figure 1 Offer of employment from Iluvatar CoreX

3. Signed Letter from Employer

天数智芯

4F, Building 5, No.180, Ruan Jian Ave, Yuhuatai
District, Nanjing, Jiangsu Province

Director – Research School of Engineering
College of Engineering and Computer Science
Australian National University
Canberra, ACT 0200

This is to certify that Mr. Shu Liu worked at Iluvatar CoreX (Nanjing) in three periods from 2017 to 2019 in a full-time capacity (18/12/2017-10/2/2018, 18/6/2018-13/7/2018 and 23/11/2018-15/2/2019). His average working days is 4.5 days per week.

During the employment his job involved working on outsourcing projects and developing internal product. The outsourcing projects includes power prediction of electricity generation and failure detection of wind turbine, while the internal product includes online self-learning platform and PHM software. He has worked with other engineers for the project solutions as well as tested and validate different algorithms during the developing period. He also has contributed to documentation works and shared innovative ideas with other engineers.

This report is a true and accurate account of work actually performed by Mr. Shu Liu.

Yours Sincerely,

Yanjie Lv
Senior algorithm engineer.

Iluvatar CoreX Inc.

4. Technical Report

4.1 The structure and operation of the company

An introduction of the company

Iluvatar CoreX (named 'the Company' in the following paragraph) is a company provides integrated high-performance computing hardware as well as related customizable software platforms. The goal of the Company is to design and produce a scalable, high-performance computing hardware for both edge device and cloud-based applications, while such product is to cater for the market demand of requiring specific computation hardware for the field of Artificial Intelligence technology.

The Company was founded in December 2015 by a group of software and semiconductor experts, who came from Silicon Valley and China, while this group has been working together for over 20 years in the field of Artificial Intelligence at companies like Oracle, AMD and ATI.

Company's full name, associates, parent or autonomous?

The full name of the company that I conducted my internship is Iluvatar CoreX (Nanjing) which is a start-up company. The company is associated with other 12 Chinese companies and is invested by Princeville Global (Hong Kong).

Currently, the company has three registered branches and two offices locate in Chongqing, Yuyao, Hangzhou, Beijing and Shanghai. Besides, it also has a research laboratory in the United States.

Company headquarter's full address, telephone, Internet, etc

Headquarter full address: 4F, Building 5, No.180, Ruan Jian Ave, Yuhuatai District, Nanjing, Jiangsu Province, China

Year founded: December 2015

Telephone: +86 52600691

Email: info@iluvatar.ai.

Website: http://www.sky-data.cn

Company size: around 400 employees

Managerial and administrative structure of the company.

Since the detailed managerial and administrative structure information of the company is not public, only a rough structure of the Nanjing branch is included in this report.

Figure 2 Organization structure of Iluvatar CoreX

Iluvatar CoreX is led by Yunpeng Li who is the chairman of the company and is also responsible for the sales and finance centre.
As for the research sections, it is divided into Shanghai And US division, which is in charge of Xidong Wang. Also, Xidong Wang takes the role of COO of the company.
As for the operating centers in China, they locate in Shanghai, Beijing and Yuyao, and are led by Haifen Song, Yanhuan Gao and Hua Zhang respectively. It is also noticed that the operating center is US has not been established so that the management information has not been updated.

Company's business/ products, production output, trading partners, markets.

Iluvatar CoreX focuses on the field of artificial intelligence (AI) computing. Based on its hardware accelerating technique'Soft Silicon', the company has designed a computing architecture which supports the independently developed all-in-one machine, such machine has the capability for establishing a large-scale distributed computing environment. Besides, the company also developed customized applicable software product of the field of manufacturing, medicine and trading.
The main product can be divided into two categories, intelligent data analysis platform and intelligent data applications.

Intelligent Data Analysis Platform:

SkyAXE(SkyData Al xEtreme Engine)

SkyAXE is an architecture of an all-in-one machine which integrates both hardware and software. It consists of storage layer, computing layer and application layer, which forms a stable and extendable system.

SkyDiscovery

SkyDiscovery is a computing platform for dealing AI, machine learning and neural network problem. It is able to accelerate the computing speed in all parts of hardware architecture includes CPU, memory, GPU and layer storage. With a clear user interface, SkyDiscovery can help to develop customized artificial intelligence and machine learning models with the mainstream scientific programming languages, such as Python and R.

Intelligent Data applications:

SkyFront

SkyFront is a customized prognostics and health management (PHM) system for preventing failure in industrial manufacturing. SkyFront monitors and collects the related critical data during manufacturing period; by considering the historical failure information, this PHM system is able to detect the emergency and provide notification to manager.

SkyHealth

SkyHealth is an automatic system for diagnosing with medical imaging. By evolving medical expect knowledge, the system has a good performance of detecting some specific illnesses, such as diabetes, cervical cancer and tuberculosis.

SkyEdu

SkyEdu is an education platform based on SkyDiscovery, and it is designed for students who are interested in machine learning. With a high-performance distributed computing system, students can customize their AI or machine learning model without worrying about the computing resource limitation. Besides, SkyEdu provides an algorithm library for guiding students to be familiar with the mainstream machine learning and deep learning framework.

Product Name	Description	Target customers
SkyAXE	Architecture	For the organizations or individuals who require fast computing resource
SkyDiscovery	General platform	
SkyFront	PHM	For manufacturing companies
SkyHealth	Medical diagnosis	For hospitals
SkyEdu	Education platform	For colleges or educating organizations

Table 1 A summary of the product of the company

Apart from the product above, the company also provides customized solutions to industrial customers' specific requirements, such as power forecasting, object tracking and failure detecting. These customized services are considered as outsourcing business so that they are not listed as the company's product.

Company's divisions (if there are any). Division's business/products, etc.

In the Nanjing branch, the divisions are divided based on their product and techniques. Except for the department of management, the rest divisions are Platform, Manufacture, Medicine and Sales.

Platform, Manufacture, Medicine work on their product, while Sales communicates with the customers and send the customers' requirements to the technical staff in other divisions.

Division name	**Product**	**Details**
Platform	SkyAXE & SkyDiscovery	This division co-works with other branch companies.
Manufacture (IoT) [2]	SkyFront & SkyEdu	This division is also responsible for taking outsourced task
Medicine	SkyHealth	/
Sales	/	The division bridges the customers and technical staff.

Table 2 function of all divisions

Name and address of the Head of the Division.

Division name	**Head name**	**Email address**
Platform	Fei Cai	fei.cai@iluvatar.ai
Manufacture	Xie yu	xie.yu@iluvatar.ai
Medicine	Yujie Shi	yujie.shi@iluvatar.ai
Sales	Xing Wu	xing.wu@iluvatar.ai

Table 3 Information of the Head of divisions

Company's financial base, is it private or public, is it listed on Stock Exchange?

Iluvatar CoreX is a limited liability company (Hong Kong and domestic joint venture). Based on the information provided by [1], the company is invested by 12 domestic companies and a Hong Kong company, while the registered capital is 98.809 million Yuan.

Currently, the company has not been listed on the Stock Exchange.

Total operation budget, division into business areas

The operation budget is confidential to public, so that the information is not included in this report.

Attitude of the company to its work-force, prevailing ethics in the company

As a technology company, Iluvatar CoreX has no restriction on individual work habit and provides a casual environment for its work-force. At the same time, the company has an achievement-oriented attitude to all employees. It means employees can decorate their working space and choose an arbitrary working time if they can obtain a good

[1] Tianyancha, a official website providing the operation details of companies.

[2] Manufacture division was called IoT team before 2019.

achievement before the deadline of the project.
Also, the company encourages the employee to develop themselves during working time. Tech-talk is held once a week in each division for sharing the knowledge. There are also some private activities such as reading technical paper or reviewing code each other.
Besides, among of salary of full-time employees is based on their achievement, which inspires the employees to do their best at work.

4.2 My position in the company

Title of my job/ jobs

My job in Iluvatar CoreX is algorithm and documentation engineer (intern) in 2017 and 2018, the job then changes to algorithm engineer (intern) in 2019.

My immediate supervisor, and my position within the structure

During the internship period in 2017 and 2018, my temporary supervisor was Liu Yang, who is a senior product manager and algorithm engineer. Since I was one of the three employees in the IoT team in that time, I was working as a full-time algorithm and documentation engineer.
In 2019, my supervisor became Yanjie Lv, who is a senior machine learning engineer and was working in the Medical division previously. During this intern period, I was working a full-time algorithm engineer in Manufacture division.

Responsibility and requirements in my job(s)

Since Manufacture division in Iluvatar is responsible for providing customized algorithm solutions against customers' requirements. In this case, my responsibilities and requirements include:

- Work in a team, and keep effective communication with other team members
- Apply the tools of agile software development
- Self-learn the necessary knowledge required in projects
- Keep recording the documents during the developing period
- Ensure a clear programming style
- Leaving the necessary comments in the codes
- Help other full-time employees
- Attend team activities and help to construct a friendly working environment

As for the knowledge, I have learned the techniques in the field of time-series data processing, image data processing, data mining, machine learning and deep learning. The programming languages used in this internship are Python and C.

Interaction with other employees

In 2017 and 2018, the IoT team was a small group, which had less than 10 employees. I was playing an active role in the group and working in all the projects. My supervisor Yang Liu provided great help to me, who introduced the basic knowledge of machine learning to me and gave me lots of chances to practice. My colleague Yujiao Shi (who received the scholarship and become an ANU PhD student in 2019) taught me how to use the deep learning framework and inspired me being active in work period. We formed a close relationship during the progress of customizing and completing projects. I would like to express my special thanks to them who helped me and trusted my work.
In 2019, IoT team is reorganized and named Manufacturedivision. Yanjie Lv and Ban Chen, who had belonged to Medical division, joined Manufacture division. I was encouraged to conduct some research on image processing by them. We also held tech-talk regular which is then become a common activity for all divisions.

Why was the job offered to me?

I was lucky as Iluvatar CoreX was lack of work-force in 2017, while my application was accepted with a simple interview. On the other hand, I can take this job as I had the skill of programming as well as the basic knowledge of signal processing and mathematics. Also, English is my advantage compared with other Chinese applicants.

After my first period of internship, my hard working attitude was praised by my supervisor who suggested me to continue the internship in the following holidays. Under such circumstance, I conducted three periods of internship in this company.

4.3 Technical description of my job

I had three periods of internship experience in this company. For each of the period, I co-worked with other senior engineers to design and complete solutions against specific customer requirements.

4.3.1 Power consumption forecasting

This project is processed during my first internship. It is about predicting the instantaneous power consumption based on the historic data for Taigu Power Company, while the recorded progress of the project is elaborated in Appendix 6.

Our division was called IoT team in this period and had only two employees, so that I took part in all parts in this project. In summary, my technical work involved data cleansing, data processing, and model construction. Also, I also attended a pre-bid meeting and help to write the technical report.

My programming skill was practiced during this period. Such experience also improved my understanding of the field of machine learning and data mining.

Data processing & analysis

The data we received is the amount of power consumption of 356 independent devices of an unknown company from 1/1/2016 to 30/6/2017. The data are recorded in every 5 minutes. We were required to quantitatively predict the daily power consumption of this company in July 2017, so that we planned to make a prediction on each of the devices.

Since data missing is a common problem occurred frequently, I initially applied Lagrange interpolation for filling the missing data. However, such interpolation caused the failure of following predicting, which forced us to abandon such incompleted data.

Besides, the public holiday has a significant influence on the power consumption (such as Figure 3), which means the date should be considered one of the critical features.

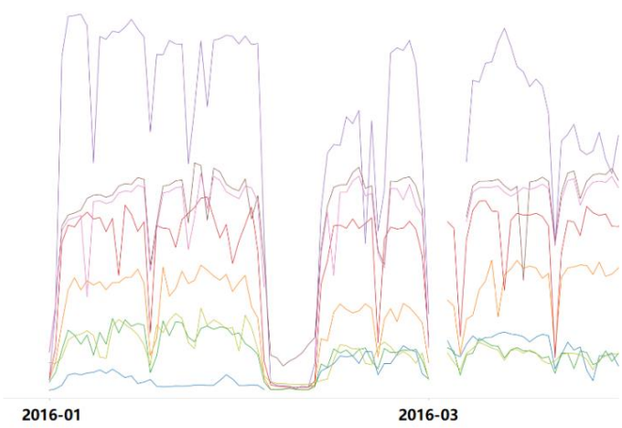


Figure 3 Power consumption of 8 devices between 01/2016 and 04/2016

(Since 6/2/2016-14/2/2016 is China public holiday, most of devices are in downtime.)

Since this project was based on the source data, my colleague and I spent several days for processing and cleansing the data.

Data clustering

Season decompression was applied to the time-series data for digging some regular pattern. Since the result of season decompression indicated the data does not have obvious long-term temporal tendency. It led to the conclusion that the power consumption is not long-term dependent. Under such circumstance, we planned to cluster the data of different devices based on their peak and nadir; then we can predict on each of these clusters.

In this case, by considering the power usage of one device as a vector A, we defined one element as a peak if its value is higher than 90% values in$A$. Similarly, nadir element was defined as its value is smaller than 90% values in A. By assigning the peak element to be '1' and the nadir element to be '-1', it contributed to a vector B having the same length with the original vector, while such vector B indicates attribute of element in$A$. In other words, B is an auxiliary vector for A, which tells whether an element is a peak (1), nadir (-1) or common (0).

A	25.9	45.3	45.9	46.2	37.8	35.4	28.1	11.8	12.1	11.2	11.3	15.6	18.4	20.5
B	0	1	1	1	0	0	0	-1	-1	-1	-1	-1	0	0

Figure 4 Example of vector A and its auxiliary vector B

By calculating the Euclidean distance among each pair of the auxiliary vector of all devices' power usage vector, we can acquire an error matrix which describes the similarity of power usage of these devices. It means that if two devices have the same auxiliary vector, they can be considered as belonging to the same cluster.

type_0_list	list	278	['PCMII1400107922', 'PCMII1400107307', 'PCMII1400107213', 'PCMII140010 ...
type_10_list	list	2	['PCMII1400131725', 'PCMII1400131726']
type_1_list	list	31	['PCMII1400107183', 'PCMII1400107317', 'PCMII1400107927', 'PCMII140011 ...
type_2_list	list	6	['PCMII1400107226', 'PCMII1400107312', 'PCMII1400107798', 'PCMII140010 ...
type_3_list	list	7	['PCMII1400107231', 'PCMII1400107786', 'PCMII1400107233', 'PCMII140010 ...
type_4_list	list	26	['PCMII1400107166', 'PCMII1400107328', 'PCMII1400107229', 'PCMII140010 ...
type_5_list	list	1	['PCMII1400107781']
type_6_list	list	1	['PCMII1400131739']
type_7_list	list	1	['PCMII1400131724']
type_8_list	list	1	['PCMII1400107293']
type_9_list	list	2	['PCMII1400107179', 'PCMII1400107178']

Figure 5 Clustering result with DBSCAN

In this case, we clustered all devices' power consumption with algorithm DBSCAN. As Figure5 shows, most of the devices belong to 'type0', 'type1' and 'type4'. After checking the magnitude of the data in these three clusters, we found these three clusters occupy over 98% of the power consumption. Therefore we abandoned the data in other clusters and focused on constructing a model for predicting the data in 'type0', 'type1' and 'type4'.

Prediction model construction

For predicting the power consumption with historical data, we proposed multiple models for extracting the temporal feature from the data. In general, our models are based on statistical method and neural network method. And I was responsible for the statistical methods.

I conducted this section with ARMA (Auto regressive moving average) which is a common method for predicting the future value with time-series data.

i. Prediction with ARMA + Wavelet

With the signal processing knowledge I learned in a university course, Fourier transformation can be used for acquiring data in the frequency domain, and such transformation would help to expose the frequency features [1]. Meanwhile, since the power consumption is a time-series data, applying wavelet transformation for substituting Fourier transformation seems a natural idea. In this case, the methodology of the prediction model is about taking wavelet transformation function for generating the data in the frequency domain and then taking ARMA to predict the value of data in the next period.

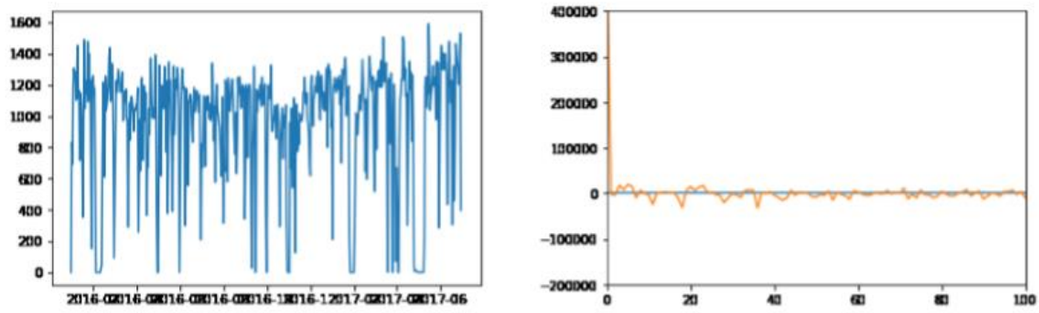


Figure 6 Prediction result on the device PCMII1400107184: monthly error is 1.8%
(Left: Ground truth, Right: Relative error)

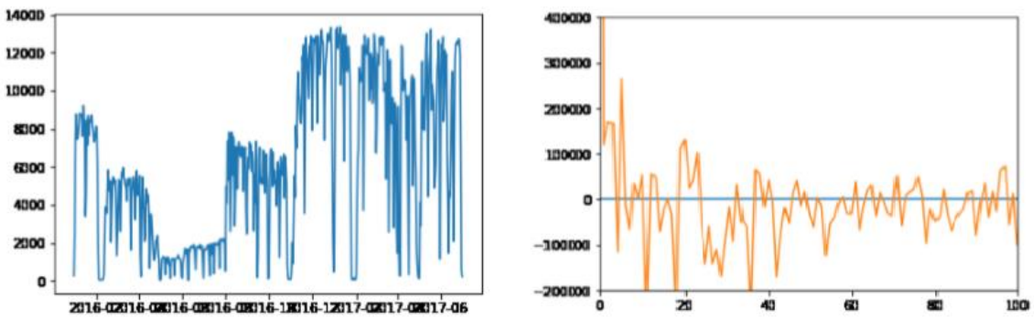


Figure 7 Prediction result on the device PCMII1400108238: monthly error is 91.77%
(Left: Ground truth, Right: Relative error)

The two devices shown in Figure 6 and Figure 7 both belong to 'type0', and it means the methodology does not have a consistent performance on all devices. In details, ARMA + wavelet has a good performance on the periodic data (such as Figure 6). While if the peak of data changes sharply (such as Figure 7), its frequency feature will be unclear which greatly influence the performance of the wavelet transformation. Therefore, this method is considered as having a high accuracy but bad robustness.

ii. Prediction with ARIMA

The operation of replacing each value in sequential data with the difference between the value and the previous value is denoted by 'differencing'. ARIMA (auto regressive integrated moving average) is a general version of ARMA model, which includes a repeatable differencing process [2]. It is capable for making both long-term and short-term prediction and has been applied in power prediction competition [3].

Index	item	error
124	PCMII1400107257	0.0060053
76	PCMII1400107345	0.00828256
39	PCMII1400107181	0.0108953
26	PCMII1400107284	0.0125763
100	PCMII1400108019	0.0211021
13	PCMII1400107132	0.0241008
68	PCMII1400107163	0.0261622
126	PCMII1400107435	0.0276643
118	PCMII1400107122	0.0279489
84	PCMII1400108236	0.0315199
65	PCMII1400107291	0.0334277
20	PCMII1400107335	0.0355686
55	PCMII1400107145	0.0358116
52	PCMII1400107447	0.0382345

Index	item	error
83	PCMII1400107341	1.83859
3	PCMII1400107420	0.885061
85	PCMII1400107324	0.840604
110	PCMII1400107994	0.829632
105	PCMII1400107418	0.771345
71	PCMII1400107154	0.765635
48	PCMII1400107315	0.755693
98	PCMII1400107777	0.691313
82	PCMII1400107318	0.548322
121	PCMII1400107347	0.487219
51	PCMII1400107780	0.467949
8	PCMII1400107280	0.451711
64	PCMII1400110522	0.422144
28	PCMII1400107929	0.389001

Figure 8 The relative error of the prediction on the different devices in 'type0'

The performance of ARIMA reflects from the relative error in Figure 8. Similar to ARMA + wavelet, ARIMA is not capable of acquiring a high accuracy on all devices in 'type0'. One of the problems causing the failure of ARIMA is the missing of data, as the temporal feature is destroyed by the missing data. Meanwhile, ARMA + wavelet can handle such

incomplete data as it applies the feature from the frequency domain.
In this case, I concluded that ARMA + wavelet and ARIMA have different advantages for making the prediction on the power consumption data.

iii. Prediction with LSTM

Although I did not have the knowledge of the neural network when I was conducting this project, I still worked on testing the performance of LSTM (long-short term memory network). While the architecture of the network contains one LSTM cell and three fully connection layers (FC layer).

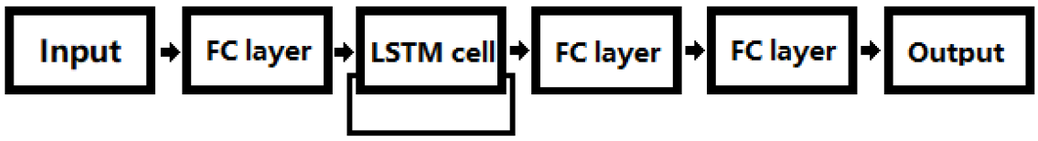


Figure 9 The architecture of the network model for power prediction

```
Epoch: 0, train_cost: 2824070.2565739327
Saving model ...
Epoch: 0, error percent in total: 0.5029832875644825
Epoch: 1, train_cost: 2248881.4055830794
Saving model ...
Epoch: 1, error percent in total: 0.08165630430708336
Epoch: 2, train_cost: 1429318.877667683
Saving model ...
Epoch: 2, error percent in total: 0.22969056419006448
Epoch: 3, train_cost: 1305792.1104230182
Saving model ...
Epoch: 3, error percent in total: 0.16941676234666314
Epoch: 4, train_cost: 1297000.1639672257
Saving model ...
Epoch: 4, error percent in total: 0.34824658487002985
Epoch: 5, train_cost: 1369785.7747713414
Epoch: 6, train_cost: 1226124.9332602897
Saving model ...
Epoch: 6, error percent in total: 0.14260498345661674
Epoch: 7, train_cost: 1326150.007050305
Epoch: 8, train_cost: 1177240.1005144818
Saving model ...
Epoch: 8, error percent in total: 0.18279153330720452
```

Figure 10 Relative error during the training of LSTM model

The training progress of LSTM model is shown in Figure 10, the relative error did not have a tendency of decreasing and was bouncing around 0.25. In this case, we concluded that such data is hard to be optimized by this LSTM model due to the lack of temporal dependence.

Overall, these three models were not robust enough for handling the data. In this case, the ensemble of multiple models would be a good choice. In addition, the information about public holiday has not been used during my internship.

Result of the work

Due to time limitation, I ended my internship before finishing this project; therefore, I did not see the final version of the model. Meanwhile, I still make a significant contribution to this project that let me have the second choice for working in this division.

4.3.2 Fault diagnosis of wind turbine

This project is processed during my second and third internship period, which requires us to detect the fault of the wind turbine and classify its fault mode with time-series data (in Figure11). The time-series data consists of over 20 columns which represent different features such as generated power, wind speed, motor speed and wind direction (some features show in Figure12).

pi10mr	WT00012	42516.63	1	3.37	1000.7	14.13	0	0	37	0	19	31	31
pi10mr	WT00012	42516.64	0	-902	-902	-902	-902	-902	-902	-902	-902	-902	-902
pi10mr	WT00012	42516.65	1	3.67	1001.11	14.08	0	0	38	0	19	31.86	31.11
pi10mr	WT00012	42516.65	1	2.67	961.37	13.55	0	0	38	0	19.03	31.73	31.13
pi10mr	WT00012	42516.66	1	1.8	714.27	10.06	0	0	38	0	19.94	29.64	31
pi10mr	WT00012	42516.67	1	1.31	530.6	7.29	0	0	38.36	0	20	28.42	30.01
pi10mr	WT00012	42516.67	1	2	681.55	9.73	0	0	35.45	0	17.27	28	29
pi10mr	WT00012	42516.68	0	-902	-902	-902	-902	-902	-902	-902	-902	-902	-902
pi10mr	WT00012	42516.69	0	-902	-902	-902	-902	-902	-902	-902	-902	-902	-902

Figure 11 The sample of time-series data

Index	Name	Description	Type	Length	Missing data
1	DATA_SOURCE	Source of data: wind10mh & wind10mr	integer		dex10mh
2	WT_NUMBER	ID of turbine	VARCHAR2	20	
3	DATA_TIME	Time	TIMESTAMP		Date format, yyyy-mm-dd hh:mi:ss.
4	WT_Status	1 Running 2 Standby 3 Malfunction 4 Maintain	integer		**-902**
5	WIND_SPEED	Wind speed	FLOAT		Average
6	CONVERTER_MOTOR_SPEED	Motor speed	FLOAT		Average
7	ROTOR_SPEED	Rotor speed	FLOAT		Average
8	Wind_Direction	Wind direction	FLOAT		Average
9	TurYawDir	Turbine yaw direction	FLOAT		Instantaneous value
10	GBXOILTMP	Temperature of oil in gearbox	FLOAT		Average
11	GBXSHFTMP	Temperature of bearing in gearbox	FLOAT		Average
12	EXLTMP	Temperature of environment	FLOAT		Average

Figure 12 The detailed description of the data (not all features included)

During my experience of this project, our division has been reorganized to be Manufacture division, while more employees joined this division. In summary, I was responsible for redesigning a new system for accurately detecting the fault with a baseline system under the supervision of Yang Liu.

During this period, I learned the principle of the wind turbine, while my understanding of the principles of programming was greatly improved. I was required to read over 1000 lines of Python code and create a system with over 2000 lines code, which is efficient programming practice for me. Last but not least, this project let me know the importance of system engineering in an industry-level project.

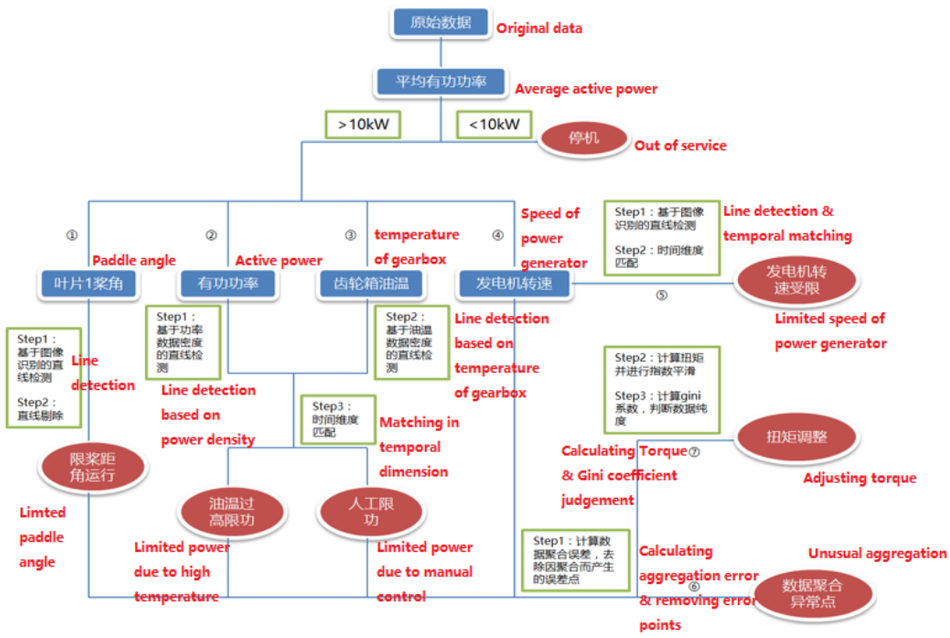


Figure 13The system diagram of the old fault detection system

Background study

My work is developed based on the system shown in Figure 13. Although such a system was capable of identifying seven different fault mode (illustrated by the red ellipses), the customers still wondered a more powerful system. Besides, these seven fault modes were too general, which cannot guide maintenancers to find the source of the fault. Under such circumstance, my work started from learning the principle of wind and I found that most fault modes could be observed with the plot of generated power verse wind speed (Figure 14, Figure 15, Figure 16).

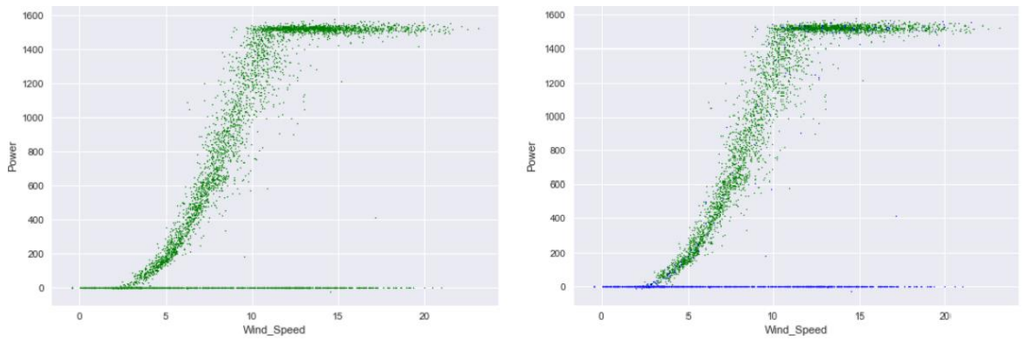


Figure 14The **downtime fault** detected on power-wind speed plot (blue points)

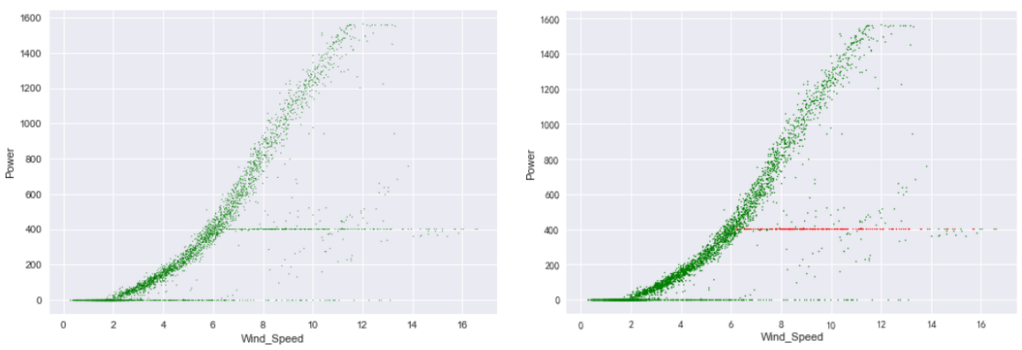


Figure 15The **power limitation fault** detected on power-wind speed plot (red points)

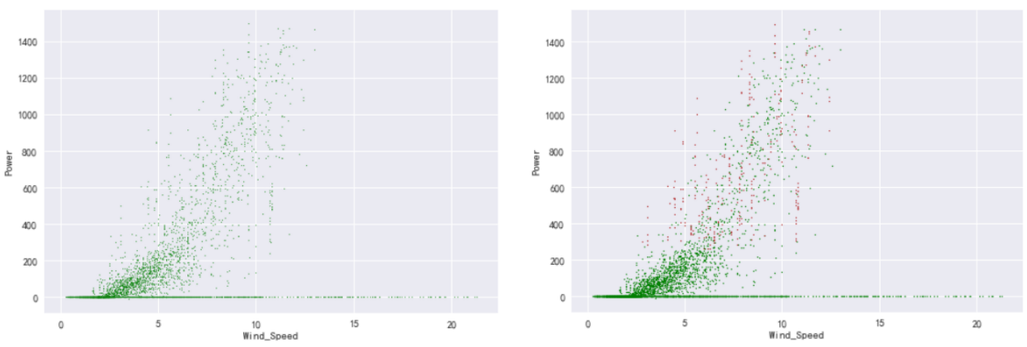


Figure 16The **scatter point fault** detected on power-wind speed plot (purple points)

Meanwhile, supervisor Yang Liu noticed that plot of power verse wind speed does not consist of any temporal feature as the dots in the plot are not sequential. He suggested that I can detect specific pattern on the power-wind speed plot for acquiring a proposed fault, and validate such result on the original time-series data. Such idea guided me in the following implementation work.

Model designing & implementation

After deciding the methodology, it cost me over two weeks for implementing the detection model. I firstly wrote some elementary functions, such as *line_detection,* which such elementary functions were then integrated for detecting specific fault mode. The new detection system is illustrated by the system diagram shown in Figure 17. Compared with the previous system, the new one had a more systematic structure and involved more features. Also, it can detect more fault modes, and have higher accuracy. At the same time, the code was uploaded on a system called Jira and was improved with the code review by other colleagues.

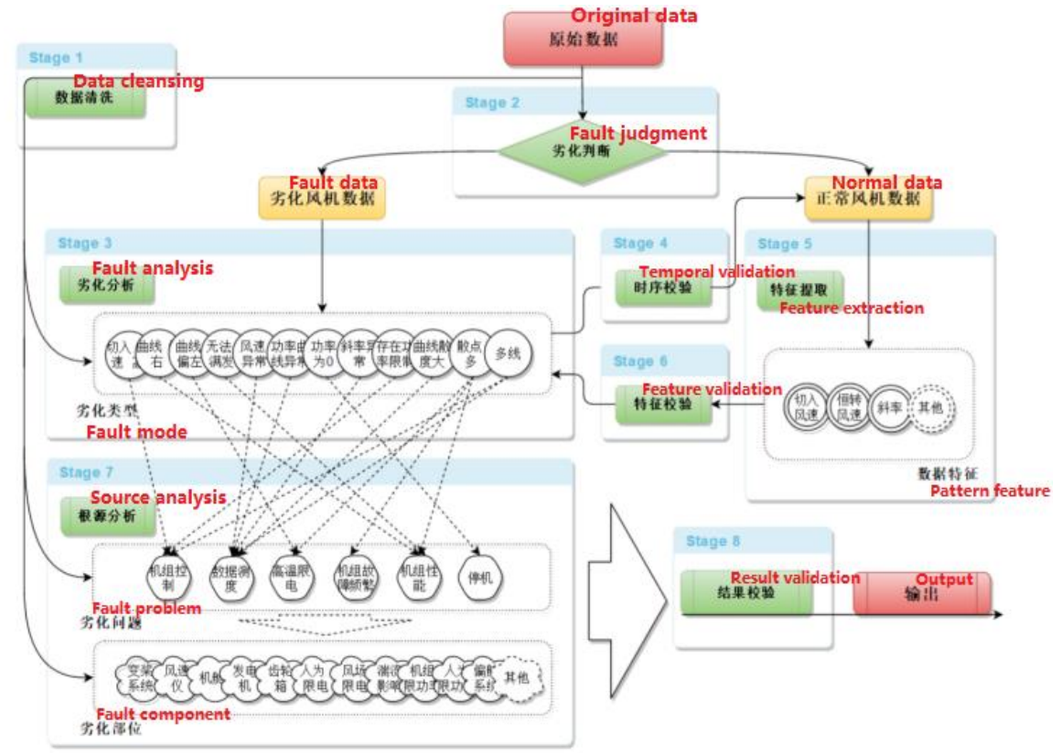


Figure 17 The system diagram of the new fault detection system

Result of the work

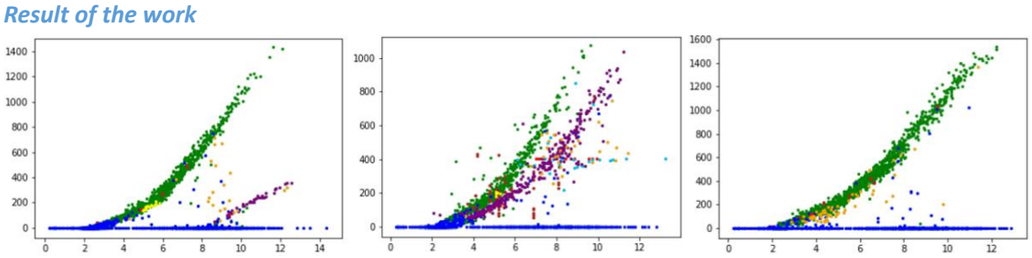


Figure 18The detected result. For example, blue is downtime; purple is multiple trajectory, etc.

Three examples of detected data are shown above, I colored the normal data with green while other colors represent different specific fault modes. Also, such fault points were validated in the time domain, while the result samples shown in Figure 19.

机组 Device index	开始时间 Start time	结束时间 End time	劣化类型 Fault mode	劣化根源 Source of fault
1	2013/6/14 12:51	2013/6/14 18:31	多线 Multiple trajectory	桨叶角参数调整 Paddle angle adjusting
1	2013/6/14 17:21	2013/6/15 9:31	多线 Multiple trajectory	限桨叶角运行 Paddle angle limited
1	2013/6/15 9:51	2013/6/15 15:21	多线 Multiple trajectory	桨叶角参数调整 Paddle angle adjusting
1	2013/6/15 15:31	2013/6/16 11:40	多线 Multiple trajectory	限桨叶角运行 Paddle angle limited
10	2013/6/8 22:38	2013/6/9 0:08	限功 Power limitation	齿轮箱油温过高自动降容 Gearbox oil temperature is too high
10	2013/6/14 13:37	2013/6/14 18:06	限功 Power limitation	齿轮箱油温过高自动降容 Gearbox oil temperature is too high
10	2013/6/24 23:14	2013/6/25 0:44	多线 Multiple trajectory	发电机转速异常 Unusual motor mode
10	2013/6/25 1:04	2013/6/25 2:24	风速异常 Unusual wind speed	风速仪损坏 Wind speed gauge damaged

Figure 19The detailed fault samples in time domain

Although my supervisor has come up with many suggestions against my code implementation and he spent lots of time to rewrite my implementation, he still praised me for the system design. I also helped to write the handover report even I was not in the company. In the end, the customer of this project paid around $100,000 for our work.

5. Development of Stage 1 Competency Standards for Professional Engineer

5.1 Competency Unit 1 Knowledge and skill base

Comprehensive, theory-based understanding of the underpinning natural and physical sciences and the engineering fundamentals applicable to the engineering discipline.

During my internship experience, understanding the underpinning sciences is the critical part which provides the prior knowledge for a systemic project. The physics was especially valuable when I was studying the principle of the wind turbine, and also, I had a deep understanding of such knowledge of physics.

On the other hand, the colleagues in this division had the mathematics and computer science background but were lack of engineering knowledge; in this case, I can share my understanding of engineering for processing practical projects.

Conceptual understanding of the mathematics, numerical analysis, statistics, and computer and information sciences which underpin the engineering discipline.

Algorithm engineer is a job that requires a good understanding of mathematics, statistics and information science. When I conduct this internship, I applied the knowledge of mathematics, statics and information to understand and reproduce algorithms. Also, computer science was the basic knowledge for me, as all the works were processed by programming.

In-depth understanding of specialist bodies of knowledge within the engineering discipline.

The work of wind turbine was based on a deep understanding of the mechanism of wind turbine. For example, the knowledge of dynamic system and mechanics that I learnt helps me to find that the wind speed and power would be the critical factors for detecting fault. While the following experiments proved that my understanding was correct.

Discernment of knowledge development and research directions within the engineering discipline.

All the projects that I conducted involved some advanced algorithms and architecture of system, which was a valuable choice for me to be familiar with such fields. Besides, due to the requirements of projects, it was necessary to read the latest paper, which inspired me to conduct some researches in my coursework.

Knowledge of contextual factors impacting the engineering discipline.

When I entered the company, the division had only two employees. Later, when more employees had been involved, I noticed the organization and project management become a problem, which was also a course for me to understand the importance of system engineering.

As for the techniques, I used to share the engineering knowledge with other staff as I was the only one having an engineering background. Under such circumstance, I can receive others' views the against engineering field, which helped me to obtain some knowledge of contextual factors impacting the engineering discipline.

Understanding of the scope, principles, norms, accountabilities and bounds of contemporary engineering practice in the specific discipline.

During my internship experience, although I needed to do research and the read paper, which looks like a researching job, I still required to process the projects in an engineering manner. For instances, I have to submit an experience report, an algorithm instruction document and a project handover document for ending a project, though what I had done is just writing code.

5.2 Competency Unit 2 Engineering application ability

Application of established engineering methods to complex engineering problem solving.

As for the wind turbine fault detection project, I applied the knowledge that I learnt in the system engineering course to form a system diagram (Figure 17) for illustrating the architecture of the algorithm.

Besides, after I introduced the idea of the spiral system engineering model [4], such a model was noticed by the head of the division and now become the basic development method in the division.

Fluent application of engineering techniques, tools and resources.

In the company, the manager used to require the employees to use Jira to manage the process of the project. I need to update the daily and weekly workload on the Jira system.

Besides, Confluence is applied to manage the resource and code used in projects.

Application of systematic engineering synthesis and design processes

As for the project about the wind turbine, I firstly reviewed the work that a colleague had done and learnt the advantage of the previous system. It helped me to quickly understand the critical part of the problem, which further became the baseline when I developed the new system.

When I designed the new system, I learnt some necessary sections from the old system and discussed with my supervisor. I also held a tech-talk about my system and ask colleagues to discuss and make decisions.

Application of systematic approaches to the conduct and management of engineering projects.

In the project of predict power consumption, I followed the process that I learnt in a university course to conduct the project. My supervisor and I communicated with the customers and design the initial plan. Although we did not construct system architecture, the other processes such as Problem Scoping and Idea Generation still helpful during the processing of the project.

Besides, I made up for the experience of designing a system architecture in the wind turbine project.

5.3 Competency Unit 3 Professional and personal attributes

Ethical conduct and professional accountability

As for the field of artificial intelligence or big data, the bound of engineering practice is quite sensitive, as the data we were taken for training would violate others' rights. One of the examples is images, although I did not directly work in an image processing project, I heard colleagues used to talk about the violation of personal right when some business entities want to collect public pictures. In this case, as an engineer, what we can do is to manually remove the personal information from images. Although such work is time-consuming, it is still a necessary responsibility of engineers.

Effective oral and written communication in professional and lay domains

I would like to have a technical discussion with colleagues and talked about the benefit of some specific algorithms or mechanisms, while my colleagues used praised me for my knowledge and communication skills. And also, my supervisor asked me to conduct some technical interviews for practicing my communication.

As for professional writing, I wrote the handover document of the wind turbine fault detection project. I was also responsible for writing several sections in an algorithm instruction document.

Creative, innovative and pro-active demeanor.

I was pro-active when came across problems. The wind turbine project was challenging for the division as its data consists of too many columns (shown in Figure 12). I spent several days to investigate the similar tasks which finished by other researchers and it is the breakthrough of the project.

Besides, I played an innovative role in the project of wind turbine fault detection. I did not follow the idea of the old system; instead, I rewrote the basic function and sought the relationship between the features and fault modes. With some innovative

hypotheses, I found some general clues which can effectively reflect faults.

Professional use and management of information.

As I mentioned above, I have conducted some background research for the wind turbine project. Most of the resource papers come from scholar.google.com, meanwhile, I also visited the technical consultants with my supervisor. Although the consultants did not provide a solution, they still came up with some inspired ideas.

Also, I summarized the critical ideas of some papers that I read. And my notes became a useful reference when I designed the solution.

Orderly management of self, and professional conduct

I kept a regular lifestyle when I worked in the company. I used to go to the office two hours earlier than others, and I would like to use these two hours to read some papers unrelated to the project. Although this habit did not directly help the process of project; on the other hand, I stored lots of knowledge, which somehow helped me to understand some hard algorithm.

I also kept following several specific fields such as optical flow and vision tracking. As I knew more knowledge of optical flow and vision tracking than others in the company, I used to play the role of expert consultant in these fields.

Effective team membership and team leadership

I was a friendly and hard-working member in the division. During the experience of the internship, I worked with professional engineers and practiced the skill of collaborating with others.

I was the youngest intern in the division, so that I did not have the choice to lead others in the company. But I still learnt from the experienced supervisor who leaded the group and inspired others.

6. Conclusions

In conclusion, my internship experience as an algorithm engineer in Iluvatar CoreX includes three periods from 2017 to 2019, which is totally 18 weeks. My work is mainly about completing the outsourced projects proposed by the customers. In details, I was responsible for conducting background research, requirement analysis, architecture design, model implementation and related documentation work. As for employee development, I had the experience of attending and holding tech-talk. Besides, I also learnt how to conduct research and how to collaborate with professional engineers.

I have completed two projects during internship experience, while the first project is about industry power consumption prediction. Since I had no experience in statistic machine learning and deep learning, my supervisor gave me lots of help. I also tried my best to apply the knowledge I knew to handle the problems. In this project, the main innovative contribution I have made is considering the data in the frequency domain and applying an ARMA + wavelet model for time-series prediction.

The second project that started in 2018 is about wind turbine fault detection. With the knowledge that I learnt from university, I was able to work independently for processing this project. More techniques were involved in this project, which forces me to conduct research to learn the latest methods. In this case, the experience provided me with the chance of researching and conducting an industry-level project. With the help of my supervisor, I designed a new detection system and, as a result, our customer paid for the system.

7. Reflection of my Work Experience

First of all, such internship program is a valuable experience for me as it gave me not only a chance of practice but also an engineer's view. Iluvatar CoreX provides a platform for me to look the field of artificial intelligence from industry-based view, which should be helpful in my following career.

As for personal development, internship let me learnt the importance of self-learning and self-managing. Different from life in university, the hard-working colleagues formed an environment which made me feel under pressure and encouraged me to keep learning and following others' work. At the same time, I have to express thanks to my supervisor who let me manage my own workload. It motivated me to manage my workload carefully and give me the chance to learn the content that I was interested.

At the same time, the intensive work of programming let me accumulate a sufficient programming practice and enhances my skill of programming. On one hand, it encourages me to take more programming work in university; on the other hand, it gives me the option that I could also become a programmer instead of a pure engineer in a company.

Another necessary lesson that I learnt in the company is the group collaboration and project management, which is an extended course of system engineering for me. Just like the course project on campus, the system engineering seems unnecessary when Manufacture division had few employees, but when more employees involved, the management was in chaos. Under such circumstance, my knowledge of system engineering started to be noticed by the head of the division, and I realized that the importance of system engineering in group works.

Communication with colleagues is also a meaningful experience for me. My communication skill was enhanced during the internship as I needed to communicate with people having different backgrounds. Pure programmers would like to talk about the code implementation, while algorithm engineers focused on the mathematical theory behind the algorithm. In this case, I learnt that choosing a proper topic can lead to efficient communication with different people. And I believe I can keep a friendly relationship with others in my following career.

In general, I feel grateful for having the chance to work in such great environment. I think I am benefited a lot from the internship as well.

8. Acknowledgement

I would like to express my gratitude to all colleagues in the Manufacture division, especially my supervisor Mr. Yang Liu and Mrs. Yanjie Lv who provided technical help during my internship period. I also feel grateful for the healthy lifestyle suggestions provided by the lovely colleagues in Platform division.

9. References

[1] Boashash B, Azemi G, Khan N A. Principles of time–frequency feature extraction for change detection in non-stationary signals: Applications to newborn EEG abnormality detection[J]. Pattern Recognition, 2015, 48(3): 616-627.

[2] Contreras J, Espinola R, Nogales F J, et al. ARIMA models to predict next-day electricity prices[J]. IEEE transactions on power systems, 2003, 18(3): 1014-1020.

[3] Radha S, Thenmozhi M. Forecasting short-term interest rates using ARMA, ARMA-GARCH and ARMA-EGARCH models[C]//Indian Institute of Capital Markets 9th Capital Markets Conference Paper. 2006.

[4] Spiral Technology, Inc. The Spiral System Engineering Process

Appendix 1. Stage 1 Competency Matrix – mandatory

Unit	UNIT Descriptor	Competency Claimed [Y/N]	Section of Line or paragraph numbers where covered in report
1	KNOWLEDGE AND SKILL BASE		
1.1	Comprehensive, theory-based understanding of the underpinning natural and physical sciences and the engineering fundamentals applicable to the engineering disciplines	Y	4.3.2 Background study
1.2	Conceptual understanding of the, mathematics, numerical analysis, statistics, and computer and information sciences which underpin the engineering discipline.	Y	4.3.1 Prediction model & construction 4.3.2 Model design & implementation
1.3	In-depth understanding of specialist bodies of knowledge within the engineering discipline.	Y	4.3.2 Background study
1.4	Discernment of knowledge development and research directions within the engineering discipline.	Y	4.3.1 Data processing & analysis 4.3.2 Background study
1.5	Knowledge of contextual factors impacting the engineering discipline.	Y	4.3.2 Model design & implementation
1.6	Understanding of the scope, principles, norms, accountabilities and bounds of contemporary engineering practice in the specific discipline.	Y	4.3.2 Background study
2	ENGINEERING APPLICATION ABILITY		
2.1	Application of established engineering methods to complex engineering problem solving.	Y	4.3.2 Model design & implementation
2.2	Fluent application of engineering techniques, tools and resources	Y	4.3.2 Model design & implementation
2.3	Application of systematic engineering synthesis and design processes.	Y	4.3.2 Background study
2.4	Application of systematic approaches to the conduct and management of engineering projects.	Y	4.3.1 Data processing & analysis
3	PROFESSIONAL AND PERSONAL ATTRIBUTES		
3.1	Ethical conduct and professional accountability	Y	4.3.2 Background study
3.2	Effective oral and written communication in professional and lay domains	Y	4.3.2 Result of the work
3.3	Creative, innovative and pro-active demeanour.	Y	4.3.2 Background study
3.4	Professional use and management of information.	Y	4.3.2 Background study
3.5	Orderly management of self, and professional conduct.	Y	4.3.1 Prediction model & construction 4.3.2 Model design & implementation
3.6	Effective team membership and team leadership.	Y	4.3.1 Prediction model & construction

Appendix 2. Letter of Declaration from Employer

天数智芯

4F, Building 5, No.180, Ruan Jian Ave, Yuhuatai District, Nanjing, Jiangsu Province

15/2/2019

Director – Research School of Engineering
College of Engineering and Computer Science
Australian National University
Canberra, ACT 0200

The aim of engineering work experience is

I. to expose the student to the workplace and workplace issues (such as human and industrial relations, job organisation, maintenance, safety and environmental issues),

II. to provide direct insight into professional engineering practice.

This is to certify that while working at Iluvatar CoreX (Nanjing) under the supervision of accredited engineer yanjie Lv, the following minimum set of Stage 1 competencies will be manifested in the work performed and in a manner that can be documented by student Mr Shu Liu.

3.1	Ethical conduct and professional accountability.	✓
3.2	Effective oral and written communication in professional and lay domains.	✓
3.3	Creative, innovative and pro-active demeanour. **(NOT ASSESSED)**	
3.4	Professional use and management of information.	✓
3.5	Orderly management of self, and professional conduct.	✓
3.6	Effective team membership and team leadership.	✓

Yours Sincerely,

HR manager. 2019.2.15

Iluvatar CoreX Inc.

Appendix 3. Photocopy of Passport Entry and Leaving (overseas work experience only)

National immigration administration

国家移民管理局

出入境记录查询结果（电子文件）

Entry-Exit record query results (Electronic copy)

编号：E190704370249　　查询日期：2019年07月04日

Name: Shu Liu　Gender: Male　Data of birth: 20/01/1997

查询人姓名：刘书 ，性别：男 ，出生日期：1997年01月20日 ，

ID number: 320103199701202036　enquire entry-exit records through NIA system

身份证号码：320103199701202036 ，通过国家移民管理局出入境记录查询系统查询，

who has the following entry-exit records between 04/07/2014 and 04/07/2019

其本人在2014 年07 月04 日至2019 年07 月04 日期间有下列出入境记录：

Index　Entry/Exit　Date　Name of ID　Code of ID　Port　Flight number

序号	出境/入境	出入境日期	证件名称	证件号码	出入境口岸	航班号
1	入境	2019-06-20	普通护照	E39381130	南京机场	MU792
2	出境Leave	2019-02-18	普通护照	E39381130	南京机场	MU727
3	入境Entry	2018-11-22	普通护照	E39381130	南京机场	MU728
4	出境Leave	2018-07-16	普通护照	E39381130	南京机场	MU727
5	入境Entry	2018-06-16	普通护照	E39381130	南京机场	MU728
6	出境Leave	2018-02-16	普通护照	E39381130	南京机场	MU727
7	入境Entry	2017-11-21	普通护照	E39381130	南京机场	MU728
8	出境	2017-07-19	普通护照	E39381130	南京机场	KA811
9	入境	2017-06-21	普通护照	E39381130	南京机场	KA810
10	出境	2017-02-13	普通护照	E39381130	南京机场	MU727
11	入境	2016-12-19	普通护照	E39381130	广州白云机场	CZ3054
12	出境	2016-12-13	普通护照	E39381130	广州白云机场	CZ3063
13	入境	2016-11-24	普通护照	E39381130	南京机场	MU728
14	出境	2016-07-15	普通护照	E39381130	南京机场	MU727
15	入境	2016-06-23	普通护照	E39381130	南京机场	MU728
16	出境	2016-02-12	普通护照	E39381130	南京机场	MU727
17	入境	2015-11-21	普通护照	E39381130	南京机场	MU728
18	出境	2015-07-17	普通护照	E39381130	上海浦东机场	QF130
19	入境	2015-06-20	普通护照	E39381130	上海浦东机场	QF129
20	出境	2015-02-13	普通护照	E39381130	南京机场	MU727
21	入境	2014-07-17	普通护照	G41717221	广州白云机场	CZ322
22	出境	2014-07-06	普通护照	G41717221	广州白云机场	CZ325

Passport　Nanjing Airport

1. 本电子文件原件为PDF格式的数据电文，左上角图标处载有电子签名认证数据。通过电子签名认证的为有效数据电文，否则为无效数据电文。
2. 本电子文件的使用者可通过国家移民管理局政务服务平台（https://s.nia.gov.cn）下载用于验证本电子文件所载电子签名认证证书有效性的颁发者证书。
3. 经国家移民管理局电子签名的电子文件与盖有公安机关出入境管理部门或出入境边防检查机关印章的同类纸质文件具有相同效力。
4. 本电子文件由国家移民管理局出入境记录查询系统自动生成，因数据采集、传输、处理等原因，可能存在差错或者缺漏。

第 1 页 / 共 2 页

制作日期：2019年07月04日08时57分

Appendix 4. Feedback Form

THE STUDENT'S OVERALL PERFORMANCE (Please cross X)

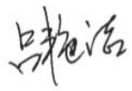
Unsatisfactory	Below Average	Average	Good	Exceptional/ Beyond Expectations
			X	

MAJOR STRENGTHS:

Shu Liu is a hard-working engineer in the group, who has good programming and report writing skill. He is active in a group work and is able to keep high work efficiency under pressure. He is also studious and keeps a good habit of reading paper, it is necessary for either engineer or researcher.

RECOMMENDATIONS TO MAKE THE STUDENT BETTER PREPARED FOR THE WORKPLACE

Shu Liu has good theoretical knowledge of some certain fields, but his practical experience seems not enough. It would be desired if he can have more practices such as implementing the idea of paper or completing an algorithm.



Appendix 6. Daily records of the progress of power forecasting project

day1
回顾python相关语法 Review Python
学习tensorflow语法，了解机器学习及数据挖掘相关知识 Learn TensorFlow, learn the knowledge of machine learning & data mining
了解当前电力预测项目，阅读文档，跟上目前进度SQL处理 Learn related knowledge of power forecasting, follow the code of SQL

day2
对数据基于公司给出的要求进行预处理 Pre-process the data based on the requirement of customer
由读数及电表规格计算实际用电量 Calculate the amonut of the consumed power based on raw data
处理偏差数据，确定失真读数的相关设备 Handle the biased data, identify the distorted devices

分析数据来源及不同设备权重 Analyse the source of data and their weight
基于峰时谷时的周期 进行聚类 Cluster the data based on the period of peak

day3
对聚类结果进行建模预测 Make prediction based on the clustered data
平稳性测试（ADF test）结果较差 ADF test has a bad performance
直接套用ARIMA模型 Apply ARIMA model
效果较差 It has bad performance as well

day4
研究之前失败模型的误差来源->考虑下一步的工作方向->尝试新的模型 Discuss the reason of failure -> decide the work in the next stage -> try new model
参加招标会议 Attend tender meeting Get the detail of the structure and product of Taigu -> globally monitar all devices in a certain region
了解太谷电力的产业结构及产品->对区域内企业设备电量的全局检测（目标：对企业的用电量进行预测->减少售电的经济损失） (Goal: predict power consumption-> control the waste of power)
了解其他具体要求->帮助政府对地区整体产业进行监控，帮助企业评估其能源支出->从而实现产业导向
查看其他竞标公司（日立电子）的情况： Get the detailed requirement -> help the government to monitar manufacture, help company save power
主要业务是电子产品，有独立开发的数据挖掘系统，有处理大数据的经验 Get the details of other tenderer (Hitachi)

day5
尝试新算法 Test new algorithm
小波预测+ARIMA Wavelet + ARIMA
对原始数据进行小波分解->自动设定ARIMA模型->对指定时间进行电量预测 decompose data with wavelet decomposition -> set ARIMA model -> predict the power in the test region
reference

day6
对小波预测的模型进行尝试修改 Modify wavelet model
改变小波分解的阶数，尝试其他的小波母波形 Change degree of wavelet decomposition, test other waveform
对两个数据集进行测试 Compare the performance on two dataset
比较带有积分的ARIMA算法及小波+ARIMA算法的优劣 Compare the performance between ARIMA (with integral) and ARIMA + wavelet
（对于有明显周期性的数据有较好的结果，对于缺失数据较为敏感） Wavelet has a good performance on peroidic data, it is sensitive to missing data

day7
参与完成竞标报告 Complete technical report
参加竞标会议 Attend pre-bid meeting
讨论下一步的研究方向 Discuss the work in next stage
获得更完整的源数据for test Get more data for testing the generalization of model
获得其他的维度的相关数据便于辅助预测 Acquire the data from other dimension for enhencing predicting