



Research Proposal

Improved ELM based on generalized inverse matrix

Name : Haoyi Sun

Student Number : 1904101

Module : Research Methods

31/03/2020

Contents

Figures.....	3
Tables.....	3
1. Introduction.....	4
1.1. The background of the researcher.....	4
1.2. Title and source of the selected project.....	4
1.3. The purpose and significance of the research.....	4
2. Background to the Proposed Research	5
2.1. Background.....	5
2.2. Proposed Research	5
3. Aims & Objectives.....	8
3.1. Aims.....	8
3.2. Objectives	9
4. Research Methods.....	9
4.1 Philosophy.....	9
4.2 Strategy	10
4.3 Approaches/ Experiment.....	11
4.4 Risks & Limitations	15
4.5 Data Evaluation.....	15
5. Development Methods	16
6. Resources & Planning.....	16
6.1. Arrangements and progress.....	16
6.2. Gantt Chart.....	16
7. References.....	17
8. Bibliography	18

Figures

Figure 1 ELM Structure Chart.....	5
Figure 2 Learning targets of ELM framework.....	8
Figure 3 Research philosophy in the ‘research onion’.....	9
Figure 4 Deductive method structure.....	10
Figure 5 L1 norm of matrix M & Error Linear diagram.....	14
Figure 6 Project Gantt diagram.....	15

Tables

Table 1 Experiment with control variables (1&2)	12
Table 2 Experiment with control variables (3&4)	13

1. Introduction

In the first part of the proposal, the basic information of the project will be briefly introduced. Including the background of the researcher, title and source of the selected project, the purpose and significance of the topic.

1.1. The background of the researcher

The researcher's re-reading is a dual master's degree, which are MSc Computational Mathematics and MSc Applied Computing.

1.2. Title and source of the selected project

Title: Improvement of ELM Algorithm Based on Generalized Inverse Matrix

Subject Source: Statistical Learning.

1.3. The purpose and significance of the research

(1) Research purpose

By improving the error of the hidden layer output matrix in the process of Extreme learning machine (ELM), the purpose of improving the algorithm error is achieved. ELM is a learning algorithm and one of single-hidden layer feedforward neural networks (SLFNs). Therefore, the ELM algorithm consisting of only the input layer, a single hidden layer, and an output layer is a one-time completion algorithm, which has the ability to efficiently process big data. For the algorithm with high efficiency, the accuracy of the algorithm becomes the primary issue. Therefore, this paper has carried out research on reducing the error and improving the accuracy of ELM algorithm.[1]

(2) Research significance

The ELM algorithm was selected for improvement in this project to have better characteristics when processing big data, that is, to ensure high efficiency and precision. As a researcher, you will learn about statistical learning and machine learning. If the experimental results meet expectations, you can complete this project and make efforts and contributions to the development of the ELM algorithm.

2. Background to the Proposed Research

2.1. Background

With the application and development of machine learning methods, the speed, accuracy, stability, and simplicity of machine learning algorithms have begun to form an evaluation system. Researchers are paying more and more attention to the theoretical knowledge of the algorithm itself, which involves a lot of statistical learning and mathematics. Therefore, improving the accuracy of an algorithm has become a hot topic. This project chooses the ELM algorithm for improvement in order to have better characteristics in processing large data, that is, to ensure high efficiency and high accuracy.

2.2. Proposed Research

(1) What is the ELM?

ELM is a multi-layer perceptron model proposed by Huang in 2006 in order to efficiently process large amounts of replicated data. ELM is a learning algorithm and one of single-hidden layer feedforward neural networks (SLFNs). ELM is a one-time learning algorithm.

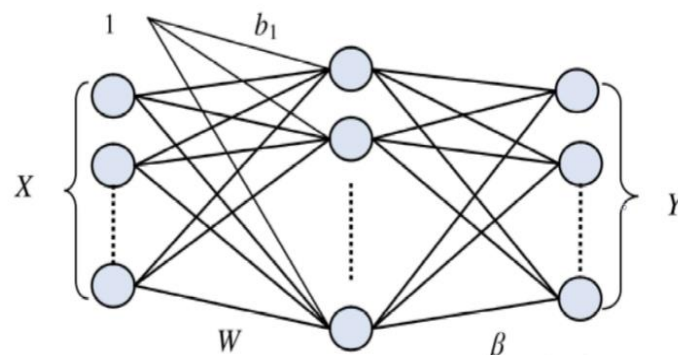


Figure 1 ELM Structure Chart

The ELM principle is that under the premise that the activation function is infinitely differentiable, the weights (ω) and offsets (b) of the input layer can be randomly determined, so the relative hidden layer output matrix (H) can be randomly generated. In the training set, given the input X and output T , given by

$$\begin{aligned} H\beta &= T & (\beta \text{ is the weight of the hidden layer}) \\ \hat{\beta} &= H^+ T & (H^+ \text{ is the Moore-Penrose} \\ & & \text{generalized inverse matrix of } H) \\ Y &= H \hat{\beta} & (\text{Estimated output } Y) \\ \varepsilon &= T - Y & (\varepsilon \text{ is error}) \end{aligned}$$

In the training set, each time the ELM algorithm is completed that a set of parameters is generated $\omega, b, H, \beta, X, T, Y, \varepsilon$, which is one-to-one correspondence. Therefore, after repeating N experiments, the optimal set of ELMs in N experiments can be selected by the index of error. [1]

(2) The development process of ELM

Regarding the improvement in principle:

a) As a SLFN, all parameters are adjustable. Lin et al. (2014) also found that certain activation functions, such as Gaussian functions, can reduce the generalization performance of ELM. However, this problem can be bypassed by training multiple ELM models or introducing regularization techniques (Huang, Zhou, et al., 2012; Lin et al. (2014). [3]

b) Wang, Shao et al. (2013) studied the generalization ability of ELM using the initial local error model (LGEM). The results show that the generalization error consists of training error, random sensitivity measure, and a constant. Based on the derivation results, an algorithm for automatic selection (selecting the number of hidden neurons) of the ELM model is proposed, and satisfactory results are obtained on the real data set.

c) Chen, Peng, Zhou, Li, and Pan (2014) established generalized boundaries for ELM variants for the "learning ordering" task. In their paper, the coverage number is introduced to measure the ca rate of the hypothesis space.

d) Rahimi and Recht (2007, 2008a) conducted a theoretical analysis of random Fourier features (one of the ELM random feature maps), and the results show that the inner product of the transformed data can consistently approximate some commonly used kernel functions in feature space Inner product. After that, they gave a generalization bound for the test error of SLFNs with random hidden layers (Rahimi & Recht, 2008b). It is proved that the empirical risk of random SLFNs converges to the minimum real risk that can be obtained by a family of functions with a convergence rate of $O(1/N + 1/D)$.

Various improvements and extensions on the classification of classic classification and regression problems:

a) Wang, Cao, Yuan (2011) proved that for some activation functions, such as the RBF function, there always exists an input weight, such as the mapping matrix H with full column rank or full row rank. On this basis, an effective input weight selection algorithm is proposed to replace

the random feature mapping in ELM, which improves the stability of the output weights. After that, the algorithm is extended to ELM with sigmoid activation function

b) (Chen, Zhu, & Wang, 2013), Yuan, Wang, and Cao (2011), according to the conditions of H : column full rank, row full rank, neither column or row full rank, and solved in different ways Output weight. In this way, the output weight can be calculated more stably than traditional ELM. [5]

Research to improve the tightness of ELM has also become a hot spot. The core idea is to train ELM in a dynamic way to grow ELM.

a) During incremental ELM (I-ELM) training, hidden neurons are pruned or replaced (Huang & Chen, 2007, 2008; Huang, Chen, et al. 2006). Newly added hidden neurons can be selected from candidates. The pool is selected and only the appropriate neurons are added to the network. Therefore, by removing unrelated neurons, the resulting network can be made more compact.

b) In order to reduce the network scale of classic ELM, a fast incremental ELM is proposed, called bidirectional ELM (B-ELM). In the trimming ELM (P-ELM), the traditional ELM is used to establish the initial network to remove hidden neurons that contribute less to the training performance. In adaptive ELM (A-ELM), the size of the hidden layer can be increased, decreased, or maintained at any step in the training process.

c) The two-phase ELM algorithm proposed by Deng et al. (2011) combines ELM and leave-one-out cross-validation with a step-by-step construction process, which can automatically determine the size of the network and improve the compactness of the ELM construction model. Subsequently, Du et al. (2013) extended the two-phase ELM to a local regularization framework, which also achieved higher compactness.

d) The frugal parsimonious ELM proposed by Wang, Er, and Han (2014b) uses the recursive orthogonal least squares method to perform forward selection and reverse elimination of hidden neurons in the ELM to obtain a frugal structure.

e) Luo, Vong, and Wong (2014) proposed a sparse Bayesian method for learning compact ELM models. The method of Luo et al. (2014) is not focused on explicitly adding or removing hidden neurons in most existing sparse ELMs, but using hypothetical prior distribution to automatically adjust most output weights to zero, thereby improving Increased sparsity and high generalization performance. [4]

f) Yu and Deng (2012) proposed to reduce the size of the ELM network by gradient descent of θ (the parameter of the hidden layer). Note that in Yu and Deng (2012), the activation function is assumed to be continuous and differentiable for the parameter θ . In this algorithm, the output weight β is represented as a function of θ . Therefore, without explicitly solving β , the parameter

θ can be updated in the direction in which the overall squared error is reduced the most. The author proves through experiments that the algorithm only needs 1/16 the size of the model, so the test time of about 1/16 is compared with the traditional ELM.

g) (2014) proposed a sparse ELM (S-ELM) to replace the inequality constraints of the traditional ELM model, which greatly reduced the storage space and test time. Similar to support vector machines, S-ELM with inequality constraints can also lead to quadratic programming problems. However, since no bias term is involved, S-ELM is more effective than svm in training. In addition, Bai (2014) proposed a fast iterative method for training S-ELM, which makes it more effective than traditional ELM on large data sets.

h) In Gastaldo Zunino, Wales, and Decherchi (2013), ELM combined with random projection (RP) shrinks the number of hidden neurons without affecting the generalization performance.

.....[2]

Throughout the development process of ELM algorithm development, it is mainly based on the number of hidden layer nodes and the output matrix H . There are I-ELM through branch reduction or replacement, and hidden P-ELM with less contribution to training performance. There are gradient descent methods to reduce the size of the ELM network and thus reduce the complexity. This project found that the error of the ELM algorithm originates from the error generated by the generalized inverse matrix of the output matrix H , and is also an improvement on the matrix H .

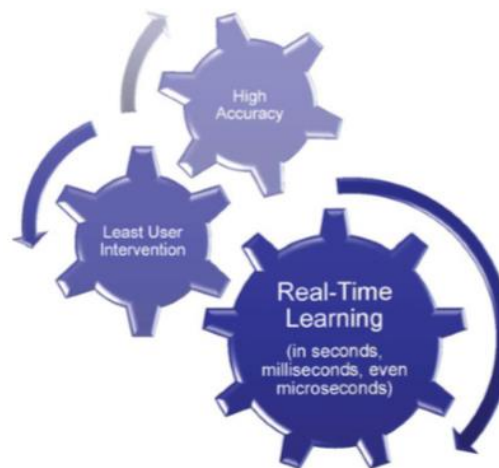


Figure 2 Learning targets of ELM framework

3. Aims & Objectives

3.1. Aims

To improve the accuracy of Extreme learning machine (ELM), which is a learning algorithm and one of single-hidden layer feedforward neural networks (SLFNs).

3.2. Objectives

In order to further achieve the aim, the aim is broken down into 4 steps

(1) To research SLFNs and the development process of Extreme Learning Machine;

To understand an algorithm, the necessary step is to understand the source and category of the algorithm. A broad understanding of the source of the algorithm and related knowledge can bypass the class, which is a solid first step to improve the algorithm.

(2) To comprehend the theory and find innovative points;

After mastering the principle of the ELM algorithm and understanding other machine learning algorithms, in order to improve the algorithm, we must get along with innovation points, which is also something that other researchers have not done. The necessary condition for independent experimental research is to find innovative points.

(3) To implement ELM with the innovative point (based on generalized inverse matrix and Lasso Regression) using code in MATLAB;

In order to further test whether the innovation point has value and can support the core of the paper, researchers began to design experiments to test. Using MATLAB software with excellent matrix calculation capabilities for algorithm programming, the data analysis involved will be a new challenge.

(4) To adjust parameters and improve the accuracy of ELM.

The risk of the last parameter adjustment step is that the experimental results may not meet expectations. However, from the perspective of theoretical calculation formulas, it is reasonable to reduce errors. Therefore, the control variable method is used for data evaluation experiments.

4. Research Methods

4.1 Philosophy

Positivism Research Philosophy--- This project is positivist for philosophy. Because this project is about improving the ELM algorithm, that is, discovering and using the characteristics of the algorithm to make objective improvements. Which objectively collect and interpret data, and analyse the data experiments. The application of the controlled variable method for experiments is to make the research results observable and quantifiable. [6]

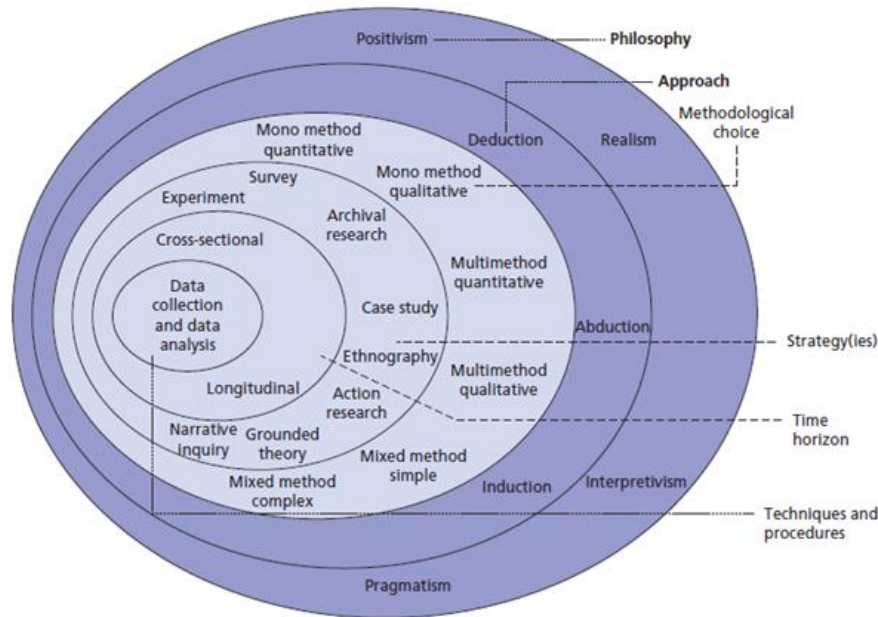


Figure 3 Research philosophy in the ‘research onion’

4.2 Strategy

The strategy of this project will be applied to logical deduction and quantitative data analysis. The improvement point of the algorithm is based on finding that the error of the algorithm comes from the error of the generalized inverse matrix of the output layer of the hidden layer, and then logically deduced to find a certain characteristic of the generalized inverse matrix to improve it, so as to achieve the goal of improving the algorithm error. The experiments are quantitative analysis, so using the control variable method is convenient to observe and adjust the experimental results.

Quantitative data analysis with the application of statistical software consists of the following stages [7]:

- (1) Preparing and checking the data. Input of data into computer.
- (2) Selecting the most appropriate tables and diagrams to use according to your research objectives.
- (3) Selecting the most appropriate statistics to describe your data.
- (4) Selecting the most appropriate statistics to examine relationships and trends in your data.

4.3 Approaches/ Experiment

(1) Deductive Approach

A deductive approach is concerned with “developing a hypothesis (or hypotheses) based on existing theory, and then designing a research strategy to test the hypothesis”. [101] This project is based on this order --- new theoretical findings, reasonable logical deduction, and preliminary experimental verification.

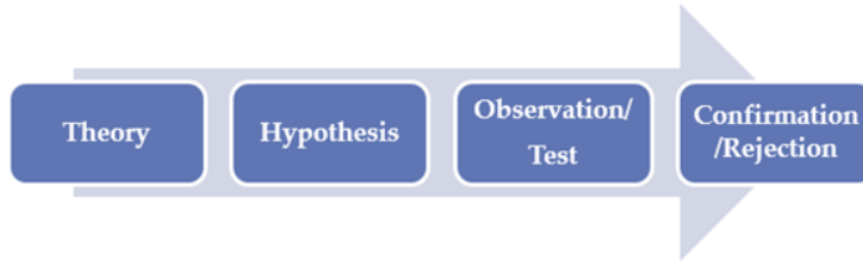


Figure 4 Deductive method structure

(2) Innovation point of ELM

According to the calculation process of ELM algorithm, we can get

$$\begin{aligned}\varepsilon &= T - H \hat{\beta} = T - HH^\dagger T = (E - HH^\dagger)T \\ M &= E - HH^\dagger \\ H &= g(\omega X + b)\end{aligned}$$

It is known from the formula that the error of the ELM algorithm comes from the output matrix (H) and the generalized inverse matrix ($H^\dagger$), and a conjecture is proposed—the error of the ELM algorithm comes from the generalized inverse matrix of the hidden layer output matrix, which also is the error of the pseudo inverse matrix.

Step 1: Definition matrix $M = E - HH^\dagger$, finding a certain property of matrix (M), and reduce the matrix error by improving this property. It is known after the experiment that its L1 Norm is linearly related to the error (ε);

Step 2: By improving the L1 norm of matrix (M), the error of the generalized inverse matrix ($H^\dagger$) is reduced, and finally the purpose of reducing the error (ε) of the ELM algorithm is achieved. Because the parameters in the elm algorithm correspond one-to-one, a set of parameters of the optimized ELM algorithm is obtained.

(3). Experiment

In order to find the relevant properties linearly related to the error (ε) in the matrix $M = E - HH^\dagger$, the L1 norm, F Norm, 1 Norm (column and norm), 2 Norm (spectrum norm),

and kernel norm of the matrix M are obtained. A preliminary comparison of these norms is then performed.

In the training set, the control variable method is used for the following comparison experiments, the purpose is to find and prove that a certain property of the matrix M is linearly related to the algorithm error.

The following four sets of comparative experiments are designed. The parameters that can be adjusted are the interval of input X , the interval of noise (σ), the regression function in the training set, the activation function of the hidden layer, the norm of the M matrix, and the number of loops of the algorithm. The training set is a uniformly distributed binary regression data set (a total of 500 groups), and the number of hidden layer nodes is set to 50. [8]

Table 1 Experiment with control variables (1&2)

Distribution of the training set	Interval of X	Interval of noise (σ)	Regression function in the training set	Activation function of the hidden layer	Number of hidden layer nodes	Correlation coefficient between error (ε) and L1 norm, F norm, 1 norm, and kernel norm (cycle 500/1000 times)
The first group---500 groups (Evenly Distributed) X, σ, T	[-10,10]	[-0.2,0.2]	$T = \frac{\sin x}{x} + e$	Hard limit function $G(w, b, x) = \begin{cases} 1, w \cdot x + b \leq 0 \\ 0, otherwise \end{cases}$	50	(-0.891, 0.535, -0.614, 0.063) / (-0.952, 0.642, -0.785, 0.142)
				Sigmoid function $G(w, b, x) = \frac{1}{1 + \exp(-(wx + b))}$	50	(0.790, 0, -0.885, 0) / (0.900, 0, -0.914, 0)
				Sine function $G(w, b, x) = \sin(wx + b)$	50	(0.892, 0, -0.567, 0) / (0.902, 0, 0.568, 0)
			$T = \sin(\frac{1}{x}) + e$	hardlim	50	(-0.865, 0.508, -0.611, 0.508) / (-0.937, 0.663, -0.792, 0.663)
				Sig	50	(-0.474, 0, 0.039, 0) / (-0.783, 0, -0.264, 0)
				Sin	50	(-0.560, 0, 0.025, 0) / (-0.566, 0, 0.286, 0)
The second group---500 groups (Evenly Distributed) X, σ, T	[-10,10]	[-2,2]	$T = \frac{\sin x}{x} + e$	hardlim	50	(-0.917, 0.799, -0.609, 0.158) / (-0.942, 0.837, -0.742, -0.160)
				sig	50	(0.593, 0, 0.880, 0) / (0.845, 0, 0.876, 0)
				sin	50	(0.873, 0, 0.890, 0) / (0.853, 0, 0.942, 0)
			$T = \sin(\frac{1}{x}) + e$	hardlim	50	(-0.540, 0.704, -0.228, 0.704) / (-0.597, 0.858, -0.498, 0.858)
				sig	50	(0.783, 0, -0.528, 0) / (0.807, 0, -0.723, 0)
				sin	50	(0.893, 0, 0.798, 0) / (0.920, 0, 0.903, 0)

Table 2 Experiment with control variables (3&4)

Distribution of the training set	Interval of X	Interval of noise (σ)	Regression function in the training set	Activation function of the hidden layer	Number of hidden layer nodes	Correlation coefficient between error (ε) and L1 norm, F norm, 1 norm, and kernel norm (cycle 500/1000 times)
The third group---500 groups (Evenly Distributed) X, σ, T	[-2,2]	[-0.2,0.2]	$T = \frac{\sin x}{x} + e$	hardlim	50	(-0.959, 0.848, 0.910, 0.065) / (-0.974, 0.915, 0.953, 0.106)
				sig	50	(0.306, 0, -0.377, 0) / (0.448, 0, -0.152, 0)
				sin	50	(0.777, 0, 0.762, 0.133) / (0.844, 0, 0.774, 0.193)
			$T = \sin(\frac{1}{x}) + e$	hardlim	50	(-0.623, 0.887, -0.677, 0.887) / (-0.532, 0.906, -0.794, 0.906)
				sig	50	(-0.519, 0, 0.095, 0) / (-0.571, 0, 0.222, 0)
				sin	50	(-0.585, 0, -0.897, 0.230) / (-0.681, 0, -0.926, 0.253)
The fourth group---500 groups (Evenly Distributed) X, σ, T	[-2,2]	[-2,2]	$T = \frac{\sin x}{x} + e$	hardlim	50	(-0.941, 0.818, 0.899, 0.881) / (-0.925, 0.967, -0.219, 0.967)
				sig	50	(0.848, 0, 0.852, 0) / (0.916, 0, 0.871, 0)
				sin	50	(0.948, 0, -0.302, 0.773) / (0.962, 0, -0.610, 0.787)
			$T = \sin(\frac{1}{x}) + e$	hardlim	50	(-0.956, 0.972, -0.836, 0.972) / (-0.953, 0.983, -0.820, 0.983)
				sig	50	(-0.679, 0, -0.001, 0) / (-0.835, 0, 0.216, 0)
				sin	50	(-0.345, 0, -0.272, -0.352) / (-0.304, 0, -0.166, -0.457)

From the data in the above 48 groups of comparative experiments, we can know that in most of the comparative experiments, the absolute value of the correlation coefficient between the error and the L1 norm is the largest, so the L1 norm can be an indicator, so that we can improve the matrix M in the future. L1 norm to achieve the purpose of reducing errors.

4.4 Risks & Limitations

The first risk of this project is that the final experimental results may not reach the expected accuracy of the ELM algorithm. Because this is an innovation, there is a possibility of experimental failure.

The second risk is that the training data is subject to uniform distribution, which has limitations and lacks realistic data testing. It is possible that the final improved ELM algorithm model cannot be applied well in the processing of realistic big data.

4.5 Data Evaluation

In the above comparative experiment, the following two pictures were selected, which also proved that L1 has a linear correlation with the ELM error.

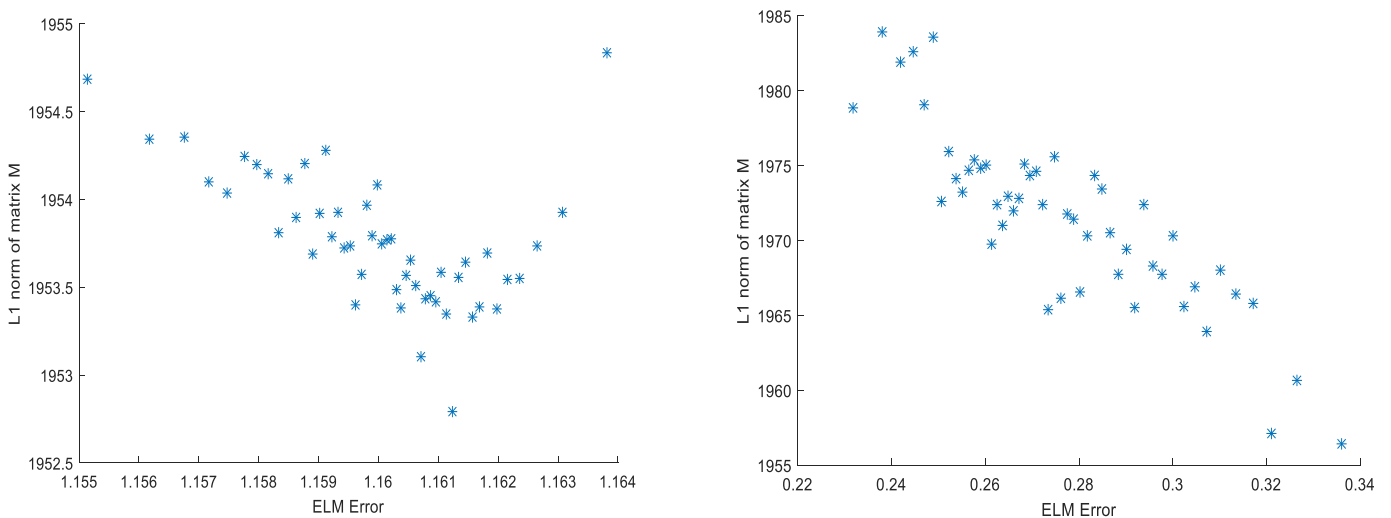


Figure 5 L1 norm of matrix M & Error Linear diagram

The previous comparative experiments have also verified the existence of the above innovations, so we can conduct further research. In the future, based on the results of the comparative experiment, we will study how to improve L1 norm of matrix M . Regression methods such as Lasso regression and ridge regression may be used to reduce the error of the matrix M , and finally the purpose of reducing the error of the ELM algorithm is achieved.

5. Development Methods

Literature survey method: Through searching the literature, comprehensively and correctly understand the ELM model content, model development and model application;

Statistical analysis method: By collecting stock data and applying mathematical software for analysis, the results are obtained, and the actual results are analysed and applied.

6. Resources & Planning

6.1. Arrangements and progress

Read the related books and literature of the ELM model systematically, and have a basic understanding of the basic theoretical knowledge of the ELM model, familiar with the principles of ELM and its parameter estimation. After reading a lot of relevant literature, we focused on the application areas of the model and related algorithms, laying the foundation for this topic.

Browse IEEE, UCI database data resources, CNKI foreign language journal database, ELM of outstanding Ph.D./Master thesis, etc., download relevant bibliographies, journals and literature at home and abroad in recent years, familiar with the current international development level of the research topic, domestic and foreign The work done by scientific researchers in this field, the current development trends and research results, summarize and organize the research and development of this topic.

6.2. Gantt Chart

Combining the content of each aspect of the project and fully estimating the difficulty of each aspect, the time course was planned and displayed in the form of Gantt chart.

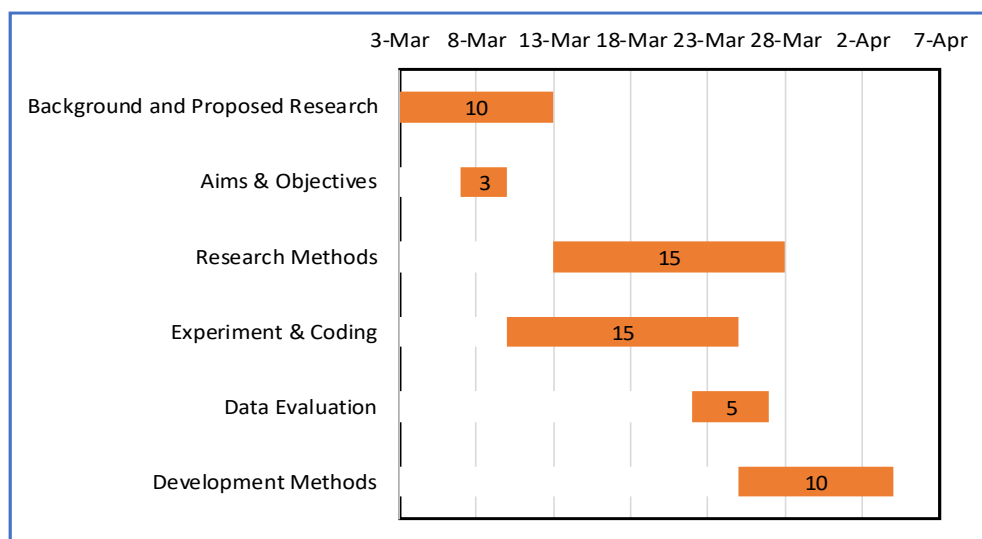


Figure 6 Project Gantt diagram

7. References

- [1] Guang-Bin Huang, Qin-Yu Zhu, Chee-Kheong Siew. Extreme learning machine: Theory and applications[J]. Neurocomputing, 2005, 70(1).
- [2] Huang G , Huang G B , Song S , et al. Trends in extreme learning machines: A review[J]. Neural Networks, 2015, 61:32-48.
- [3] Huang G , Song S , Wu C . Orthogonal Least Squares Algorithm for Training Cascade Neural Networks[J]. IEEE Transactions on Circuits & Systems I Regular Papers, 2012, 59(11):2629-2637.
- [4] Lin S J , Chang C , Hsu M F . Multiple extreme learning machines for a two-class imbalance corporate life cycle prediction[J]. Knowledge-Based Systems, 2013, 39(none):214---223.
- [5] Chen Z X , Zhu H Y , Wang Y G . A modified extreme learning machine with sigmoidal activation functions[J]. Neural Computing & Applications, 2013, 22(3-4):541-550.
- [6] Saunders, M., Lewis, P. & Thornhill, A. (2012) “Research Methods for Business Students” 6th edition, Pearson Education Limited.
- [7] Wilson, J. (2010) “Essentials of Business Research: A Guide to Doing Your Research Project” SAGE Publications, p.7
- [8] J. Cao, Z. Lin and G. Huang, "Voting base online sequential extreme learning machine for multi-class classification," 2013 IEEE International Symposium on Circuits and Systems (ISCAS), Beijing, 2013, pp. 2327-2330.

8. Bibliography

- [1] <https://www.ntu.edu.sg/home/egbhuang/>
- [2] <https://research-methodology.net/>