Publisher: State and Provincial Joint Engineering Lab. of Advanced Network Monitoring and Control (ANMC)

Cooperate:

Xi'an Technological University (CHINA)
West Virginia University (USA)
Huddersfield University of UK (UK)
Missouri Western State University (USA)
James Cook University of Australia
National University of Singapore (Singapore)

Approval:

Library of Congress of the United States

Shaanxi provincial Bureau of press, Publication, Radio and Television

Address:

4525 Downs Drive, St. Joseph, MO64507, USA

No. 2 XueFu Road, WeiYang District, Xi’an, 710021, China

Telephone: +1-816-2715618 (USA) +86-29-86173290 (CHINA)

Website: www.ijanmc.org

E-mail: ijanmc@ijanmc.org

xxwlcn@163.com

ISSN: 2470-8038

Print No. (China): 61-94101

Publication Date: March 28, 2023

Associate Editors

Table of Contents

Product Recommendation System Based on Deep Learning

Ping Lu
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 295075022@qq.com

Pingping Liu
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 1341369601@qq.com

***Abstract*—With the development of Internet big data and e-commerce, the widespread popularity of information, information acquisition and personalized recommendation technologies have attracted extensive attention. The core value of personalized recommendation is to provide more accurate content and services around users. The recommended scenarios are not uniform, and different dimensions need to be considered. For example, we are facing enterprises or individuals, different age groups, different levels of education, social life and other aspects. In this paper, the classic DNN (Deep Neural Networks) double tower recommendation algorithm in the recommendation algorithm is used as the ranking algorithm of the recommendation system. It is divided into user and item for embedding respectively. The network model is built using tensorflow. The data processed by the initial data through feature engineering is sent into the model for training, and the trained DNN double tower model is obtained. Recall adopts collaborative filtering algorithm, and applies tfidf, w2v, etc. to process feature engineering, so as to better improve the accuracy of the system and balance the EE problem of the recommendation system. The recommendation module of this system is divided into data cleaning as a whole. Feature engineering includes the establishment of user portraits, the analysis of multiple recall and sorting algorithms, the adoption of multiple recall mode, and the implementation of a classic recommendation system with in-depth learning. This makes the recommendation system better balance the interests of both the platform and users, and achieve a win-win situation.**

Keywords-Deep Learning; Recommendation System; Deep Neural Network

I. INTRODUCTION

At present, the Internet is developing rapidly in China. Information plays an increasingly important role in people's daily life. Many enterprises are also involved in the wave of the Internet and build relevant platforms with the help of the Internet, such as shopping, video, news, music, marriage, social networking and other types. However, the highly developed number of goods also brings a huge problem, that is, how to flexibly help users to accurately meet their needs. At the same time, under the influence of education and social environment, everyone has put forward new requirements for self-expression. Self personality has become an important indicator of consumer demand. In the network environment, the variety of complex goods and services and the difference of people's personalized needs together constitute a dazzling choice. Under the combined effect of these multiple factors, the recommendation system also came into being. If you want to overcome the above needs, personalized recommendation system is essential [1].

The product recommendation system based on deep learning has the following advantages and characteristics: efficiency, improve the efficiency of users to obtain the required goods; Transparency, the user's imperceptibility in the recommendation process will not affect the user's use; Accuracy: the recommended results are consistent with the products required by users; Novelty, recommend products in many ways, and break the inherent search deadlock.

At present, domestic and foreign recommendations mainly include three traditional recommendation methods: content based recommendation, collaborative filtering recommendation, and hybrid recommendation, and four common deep learning recommendation

DOI: 10.2478/ijanmc-2023-0041

models: convolutional neural network (CNN), deep neural network (DNN), recurrent neural network (RNN), and graph neural network (GNN).

At present, due to the continuous development of machine learning technology and the deepening of people's understanding of this technology, the recommendation system has made new progress. Take Alibaba and Toutiao today, two companies with significant influence in China, for example. The two key technologies included in their recommendation systems are machine learning and deep learning. These two technologies together constitute their recommendation systems. However, the traditional collaborative filtering and physical recommendation methods, which were more popular in the past, have faded out of people's vision. Such strategies are often difficult to solve the drawbacks of the long tail content's refined personalized recommendation and popularity penetration. At the same time, the benefits of the new recommendation methods are dwarfed by the traditional ones. In 2017, the National University of Singapore proposed NeutralCF, the famous double tower structure, which uses neural networks to optimize the original collaborative filtering algorithm. Youtobe and Google have successively modified the proposed model. The traditional collaborative filtering and LR mode recommendation algorithm are gradually moving towards the double tower model [2].

This paper designs a product recommendation system based on deep learning. The system conducts model training through the product data and user data in the data set. The main contents include: first, feature engineering is carried out, and basic product data, user data, and user behavior data are processed to build user portraits and item portraits; Then, the model is constructed and trained. The DNN double tower model is constructed using tensorflow, and the data set is imported for model training; The next stage is the commodity recall stage, which specifies the quantity of commodities for use in the subsequent sorting stage; The next step is to sort the goods. The recalled goods are sent to the model for sorting to get the final result; Finally, the result is output, and the result is returned through Python construction[3].

II. Introduction to Machine Learning and Deep Neural Network

A. Fundamentals of Machine Learning Theory

The development of artificial intelligence has produced a wide range of branches, and machine learning is one of the important branches that can not be ignored. If we want to realize artificial intelligence, we must not neglect the technology of machine learning. Under machine learning, the algorithm that allows the machine to modify and develop itself is adopted. At the same time, the computer obtains the corresponding rules in continuous data analysis, and the obtained rules can be used to speculate on new sample data. If we use formal language to describe its content. Its definition should be: for a certain type of task T and performance measurement P, if a computer program's performance measured by P on T is self improving with experience E, it can be called learning from experience E.

In machine learning, LR is one of the core algorithms [4]. Logistic Region is also recognized by all types of enterprises because of its simplicity, parallelization, and strong interpretability. The essence of logistic regression is to assume that the data obey this distribution, and then use maximum likelihood estimation to estimate the parameters. Logistic distribution is a continuous probability distribution. Its distribution function and density function are shown in (1) and (2):

$$F(x) = P(X \le x) = \frac{1}{1 + e^{-\frac{x-\mu}{\gamma}}} \tag{1}$$

$$f(x) = F'(X \le x) = \frac{e^{-\frac{x-\mu}{\gamma}}}{\gamma\left(1 + e^{-\frac{x-\mu}{\gamma}}\right)^2} \tag{2}$$

Logistic regression essentially uses the predictive value of the linear regression model to approximate the logarithmic probability of the real mark of the classification task. Its benefits lie in that: it directly models the probability of

classification without realizing the hypothetical data distribution, which prevents the disadvantages caused by inaccurate distribution (different from the generative model); Not only the category can be predicted, but also the probability of the prediction can be obtained[5]. When encountering some kind of tasks that use probability to assist decision-making, it can often play a huge role; The logarithmic probability function is a convex function with arbitrary order derivability, and many kinds of numerical optimization algorithms can calculate the optimal solution. Both gradient descent and Newton iteration can be used to solve the logic regression, and the gradient descent method is briefly introduced:

$$J(w) = -\frac{1}{n}\left(\sum_{i=1}^{n}\left(\begin{matrix} y_i \ln p(x_i) + \\ (1-y_i)\ln(1-p(x_i)) \end{matrix}\right)\right) \quad (3)$$

$$g_i = \frac{\partial J(w)}{\partial w_i} = (p(x_i) - y_i)\, x_i \quad (4)$$

$$w_i^{k+1} = w_i^k - \alpha\, g_i \quad (5)$$

Gradient descent is to find the descent direction through the first derivative of J (w) to w and update the parameters in an iterative manner, where k is the number of iterations.

B. Theoretical Basis of Artificial Intelligence

In 1943, the neuron model was successfully drawn by psychologist Warren McCulloch and mathematician Walter Pits. It can be seen that there are many inputs on the left side of the figure below, similar to the dendrites of neurons [6]. After a nuclear processing, that is, the weighted sum part in the figure below, and then through the activation function, an output is finally obtained. When several single neurons are combined, the output of some neurons will be used as the input of others, thus forming a neural network. If these neurons are regarded as a whole, the structure has input and output. The layer receiving data input is called the input layer, and the layer outputting the results is called the output layer[7]. The middle layer composed of middle neurons is called the hidden layer or hidden layer. The number of neural network layers is mainly determined by the number of hidden layers. As shown in Figure 2.3, it is a single-layer neural network. This neural network has 4 neurons in the input layer, 5 neurons in the hidden layer, and 2 neurons in the output layer [8].

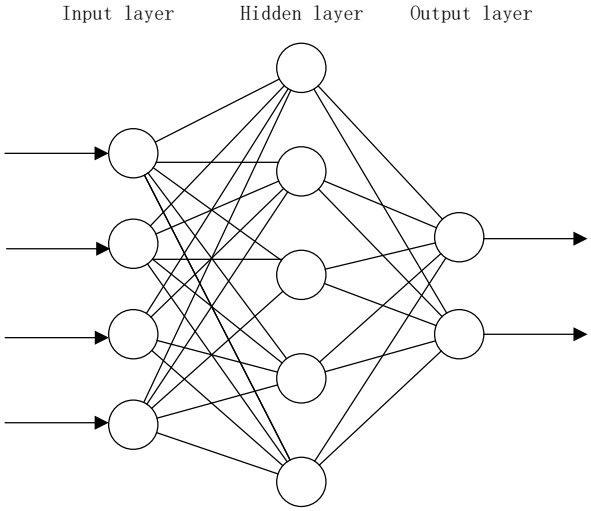


Figure 1. Neural network structure diagram.

The corresponding neural network needs to be designed according to the specific problem to be solved. Generally speaking, the number of neurons in the input layer and output layer is fixed, while the number of layers in the middle hidden layer can be adjusted freely. If compared with machine learning, deep learning is more convenient in these aspects:

1) Stronger fitting ability of deep learning model

Among feature crossing methods, dot product and other methods are relatively simple, so that when the sample data is more obscure and complex, the phenomenon of under fitting often occurs. Deep learning can effectively solve this problem by improving the fitting ability of the model. Take NeuralCF (Neural Network Collaborative Filtering) model as an example, in this model, the point accumulation layer is exchanged with the multi-layer neural network. Theoretically, multi-layer neural network can fit any function, so as long as the number of neural network layers is increased, the problem of insufficient fitting will be solved [9].

2) The structure of deep learning model is more flexible

The structures of deep learning models are different. Most of them can stack network layers with different functions. The simplest is a serial structure, some like a mesh structure, some like a pyramid structure, and so on. The typical cases are Alibaba's DIN and DIE, which adopt a sequence model with attention mechanism and interest evolution simulation in the model structure to achieve a better effect of simulating user behavior.

C. Fundamentals of Deep Neural Networks

The neural network layer inside DNN can be divided into three different types: input layer, hidden layer and output layer. In the absence of special circumstances, the first layer is called the input layer, the last layer is called the output layer, and the number of layers in the middle is called the hidden layer.

Each layer is closely connected. It can be said that a neuron randomly selected from layer i must be connected to any neuron in layer i+1. The small local model is the same as the perceptron, that is, a linear relation input data is used by the neurons in the first layer (not hidden), then the output is provided to the neurons in the next layer, and the output is repeated until the final successful output. The output has the possibility of yes or no (expressed in probability) prediction. Each layer often has one or more neurons, and each neuron calculates a small function, that is, the activation function. The activation function simulates signal transmission to the next connected neuron. The value of the afferent neuron will be compared with the threshold value. When the neuron value is large, the output can pass. The weights formed by the connection between two neurons of the continuous layer are related. The influence of input on the next neuron and the final output is defined by weight. In the neural network, the initial and weight values are randomly generated. Of course, in order to successfully predict the accurate output value, certain weights will be updated in time. Decompose the network mainly by defining some logical building blocks, such as neurons, layers, weights, inputs, outputs, activation functions, and so on. At the end is the learning mechanism (optimizer), which is of great significance in the process of gradual weight optimization of the network (random initialization). The weight value can be optimized to help predict the correct result [10].

III. RESEARCH ON RECOMMENDATION ALGORITHM AND MODEL

A. Recall Algorithm

In the recommendation system, recall is the first step. This algorithm mainly requires to select a small amount of data from the huge item information database quickly and accurately, depending on the user's relevant information and the characteristics of the products they are interested in. After the recall, the next step is the sorting process. In the recall process, huge data volume and high speed requirements are significant features. Therefore, the strategies, models and features adopted by the algorithm should not be too complicated. Data acquisition at the recall stage needs to be carried out quickly. Multi way recall is generally a classic form, such as content-based collaborative filtering recall, user based collaborative filtering recall and popular recall, as shown in Figure 2.

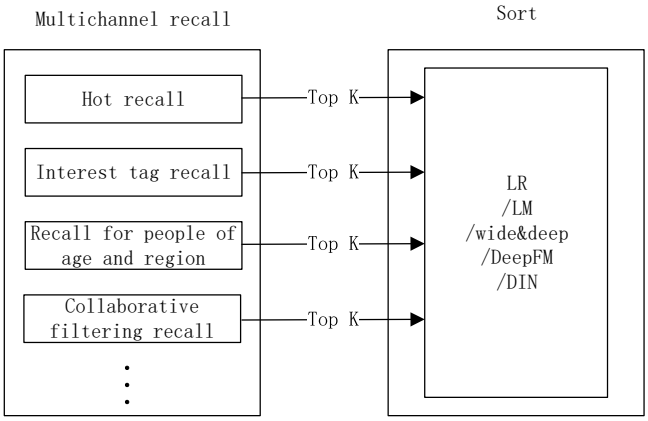


Figure 2. Recall ranking diagram.

To be specific, multi-channel recall means that users can select the products they are interested in from the commodity library in different forms at the same time, and the products selected from multiple ways are summarized and sorted by the sorting model, and finally provided to users to achieve personalized recommendations. The name of multi-channel recall proves that the core of multi-channel recall is "multi-channel". Such

multi-channel form ensures the maximum degree of diversification of product content, and also meets the needs of users to the maximum extent[11].

This system will also adopt the collaborative filtering algorithm based on items, and apply tfidf, w2v, etc. to process feature engineering. The collaborative filtering based on items mainly depends on the weight of the items, selects the items that users like, and then calculates the similarity between other goods in the item library and the products that users like, and finally pushes them to the similar products that users once liked. In addition, their methods for calculating the similarity weight of items have different focuses, mainly including the following categories:

- Click, purchase or download co cash as reference data to realize weight calculation. It can be said that for a user who has purchased Class A, Class B and Class C products at the same time. For article A, there are two co occurrence pairs (A, B) and (A, C). Another user has purchased Class A and Class B products, so there is another co occurrence pair (A, B). The meaning of co occurrence is similarity to some extent. In this way, for article A, the similarity weight value of article B is 1+1=2, and the similarity weight value of article C is 1. There may be many different calculation methods for co occurrence. But the guiding significance of this thought is significant.
- The article weight is calculated based on Word2Vec method. The sequence of items is the key point of this method. When the user purchases, clicks or downloads, the products that the user browses will generate the corresponding sequence. In this case, it is considered that this order is relevant, and items with similar distances from each other are also highly similar. The modeling of this method depends on the similarity between the order of goods. Among them, the method of modeling the item sequence according to word2vec is often adopted. According to the difference of weight vectors of different items, the Euclidean distance between item weight vectors is used to divide the similarity of items.
- Item based collaborative filtering combined with TF-IDF method. This method refers to taking the user's browsing and purchase records as a word, while specific items appear as letters in the word. The tf idf value of each item is not missed, and the value reflects the user's preference for items. Finally, you can rely on the item collaborative filtering method to find products that are highly similar to those users like.

B. Sorting Algorithm

When some high-quality goods are selected from the recall stage, the sorting stage is entered. The requirements of this stage are high, and users are required to accurately recommend the products they like according to various materials. For sorting algorithms, fast tracking is an important requirement that cannot be ignored. Users need quick feedback, but they also need to implement accurate recommendations. In the final sorting stage, ctr estimation is performed on all goods, and then topK data is delivered upwards [12].

The design of this paper in the sorting stage adopts the double tower model, and the model architecture is shown in the figure below.

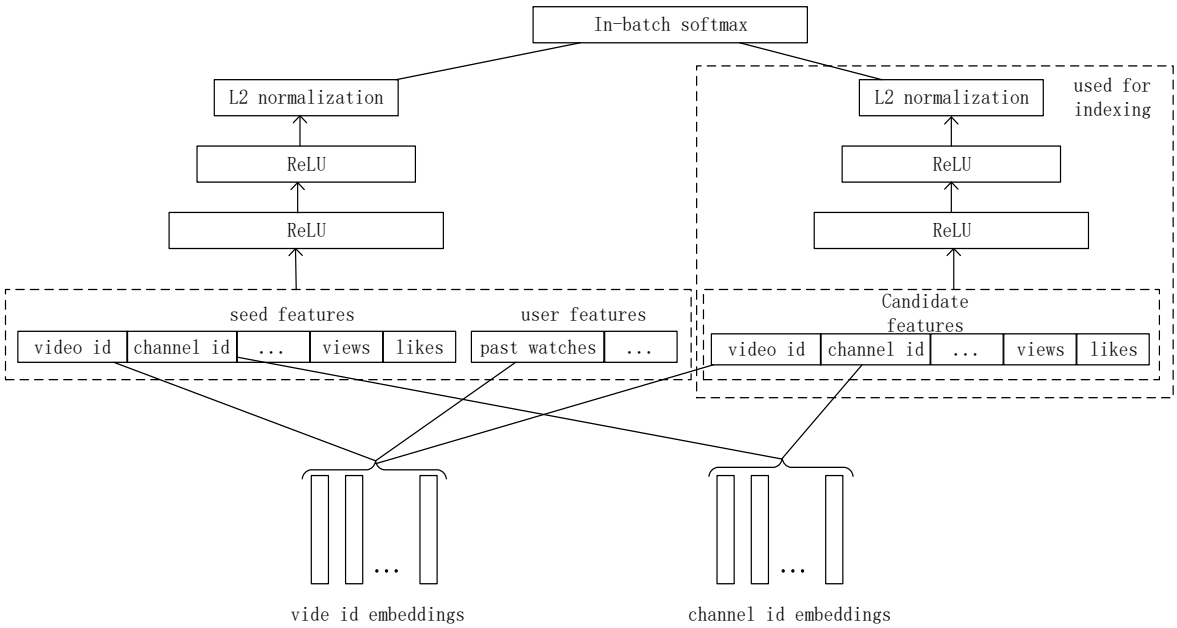


Figure 3. DNN double tower model architecture diagram.

The double tower model in this paper constructs the embedding on the request side and the embedding on the commodity side. The output of the two towers, and the final model is the output of their respective embedding vectors, which is the result of the inner product of the two embeddings. The inner product formula of the vector is as follows:

$$s\left(x, y\right) = u\left(x,\theta\right), v\left(y,\theta\right) \tag{6}$$

Based on the output of the softmax function and user preference weight, the loss function is in the form of a weighted logarithmic likelihood function, as follows:

$$L_T\left(\theta\right) := -\frac{1}{T}\sum_{i\in[T]} r_i \cdot \log\left(\mathcal{P}\left(y_i \mid x_i;\theta\right)\right) \tag{7}$$

When there are a large number of commodities in the commodity library, that is, (M is very large), it is very difficult to accurately calculate the softmax function above. In order to overcome this problem, it is obviously an appropriate method to sample a full set of commodities. The traditional form is to sample the negative samples required for training from a fixed set, however, the more efficient method for the traditional form is to sample a batch of data in the real-time stream and obtain the negative sample for training, which is also the negative sample in this batch. However, this method may cause errors, making it possible for some popular samples to become negative samples. Therefore, it is necessary to perform logQ correction on the inner product calculated by the two embedding vectors above, that is, to perform the embedding inner product correction.

$$s^c\left(x_i, y_j\right) = s\left(x_i, y_j\right) - \log\left(p_j\right) \tag{8}$$

Where, p_j represents the sampling probability of commodity j randomly selected into batch. Then use SGD to update the parameters. The specific model training algorithm is as follows:

- Obtain a batch sample from the real-time data stream;
- The sampling probability pi is obtained based on the sampling probability estimation algorithm to be mentioned below;
- Calculate the modified loss function described above.

IV. Product Recommendation System Design

The functional requirements of the system are the tasks that must be established before the system development, which specify the functions

that the system must achieve and ultimately need to meet the requirements. Through the previous research and analysis, the modules of this system are determined as user interface module, search function module, recommendation function module, and data processing module [13].

During the development of this program, pycharm is used as the IDE to write Python code. At the same time, Tensorflow gpu and Keras are used as the development tools for deep learning algorithms.

A. Functional Requirement

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

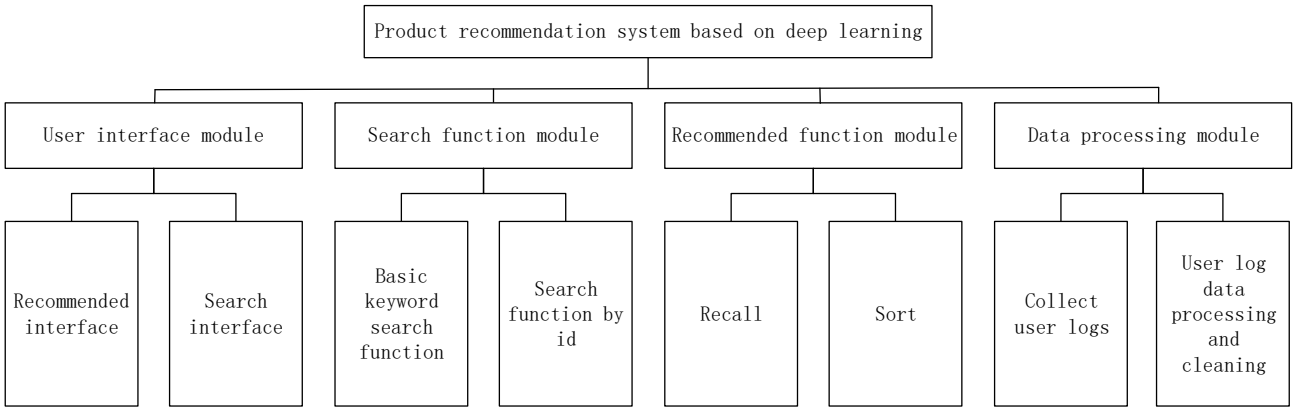


Figure 4. Function module diagram.

This topic is to design and complete a product recommendation system with complete functions. The program can recommend products to users in real time, and provide simple product search function and simple display function. The functional module is shown in Figure.

- User interface module: log in and register, search, recommend, browse favorites and purchase;
- Search function module: basic search function, keyword search;
- Recommendation function module: recall, sorting, popular recommendation, personalized recommendation;
- Data processing module: provide log service, transmit the user's information to the background in real time, and clean the data.

B. Technical Scheme for Recall Sequencing

1) Recall part (machine learning)

The recall part is a multi-channel recall. Currently, collaborative filtering is used, including item based, user based, and model based recalls. Matrix decomposition and failure are used.

2) Sorting part (deep learning)

First of all, the recall part will filter all commodity IDs, and the IDs arriving at the sorting module are only topN. The sorting module is responsible for sorting the topN commodities. (The value of N is controlled by the user). It is recommended to use tensorflow to build some parts and use the python environment of conda.

C. System Architecture Design

The recommended architecture adopts big data lambda architecture, and the front and rear end adopt golang web framework gin for connection, and each service uses grpc microservice framework for connection. This topic uses MVC mode, and the system architecture is shown in Figure 5.

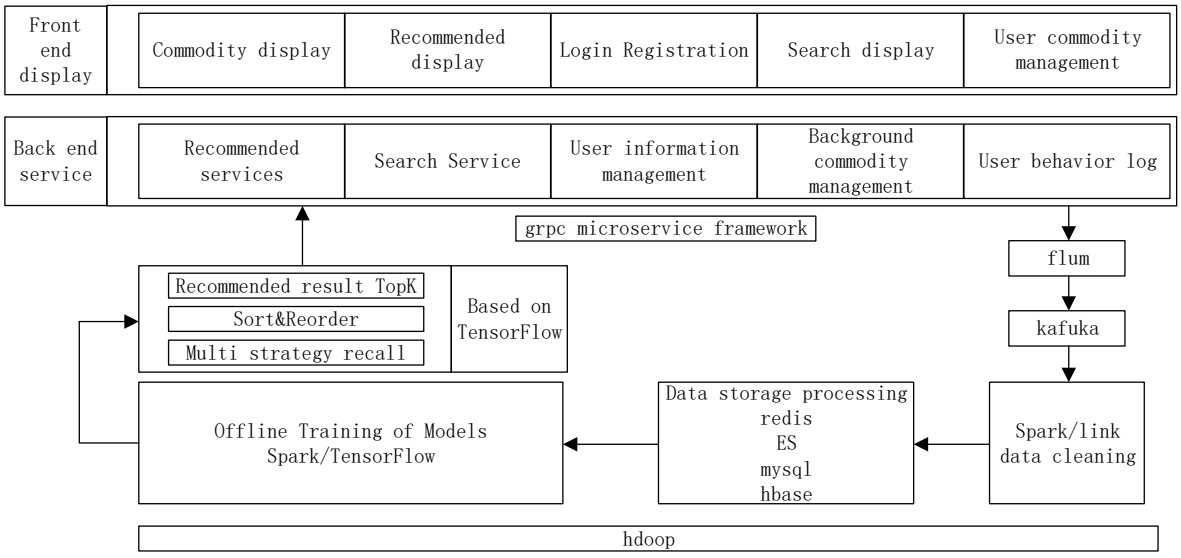


Figure 5. System architecture diagram.

D. Implementation of Commodity Recommendation System

First, it introduces the program development environment, and then analyzes the technical feasibility of program development. This program is developed on the PC side using the MAC environment and Python language. At the same time, it involves in-depth learning and network training, so it needs to use high-performance GPU for computing acceleration [14].

The specific development environment is as follows: the development platform is PC; The software environment operating system is MAC; The development language is Python; The IDE is pycharm vscod.

Next, we will introduce various tools and frameworks used in the development of this program. During the development of this program, pycham is used as the IDE to write Python code. At the same time, Tensorflow gpu and Keras are used as the development tools because of the deep learning algorithm [15].

The program development mainly uses the following development tools and frameworks:

1) Tensorflow gpu

Tensorflow is a symbolic mathematical system based on data flow programming, which is widely used in the programming implementation of various machine learning algorithms. Its predecessor is Google's neural network algorithm library DistBelief. Its gpu version can improve the training speed.

2) Keras

Keras is an open source artificial neural network library written by Python. It can be used as the high-level application program interface of Tensorflow to design, debug, evaluate, apply and visualize in-depth learning models.

3) CUDA

NVIDIA has introduced a general GPU parallel computing architecture. Because the deep neural network will involve a lot of matrix operations when training and using, the use of CUDA to operate on the GPU can greatly improve the efficiency of the model.

In recent years, the development of deep learning has driven the development of the field of artificial intelligence. Research on recommendation systems through deep learning has emerged in an endless stream, and there are many open source resources. Therefore, it is technically feasible to develop a commodity recommendation system [16].

V. CONCLUSIONS

This paper first introduces the development background and research significance of this topic. By analyzing the status quo of recommendation technology and deep neural network, and referring to the development status at home and abroad, this paper explains the significance of the development of this system. Then it explains the technology

used in the system development. This system uses Python language and Tensorflow architecture. In the requirement analysis part, the feasibility is analyzed in many aspects, and then the functional requirements and non functional requirements of the system are introduced, and the three functions of the system are analyzed. It describes how different functions are progressing, clearly and concisely shows the system construction method, how different modules collaborate, and shows the completion of the system and the core code.

In a word, the design of this system can meet the needs of the development of the times, so that people can enjoy the application of the recommendation system in the commodity industry. It is an efficient and novel system.

REFERENCES

[1] Elkahky A M, He X. A multi-view deep learning approach for cross domain user modeling in recommendation systems//Proceedings of the 24th International Conference on World Wide Web. Florence, Italy, 2015:278-288.

[2] Huang P S, He X, Gao J, et al. Learning deep structured semantic odels for web search using clickthrough data//Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. Boston, USA, 2016:29-34.

[3] Cheng H T, Koc L, Harmsen J, et al. Wide & deep learning for recommender systems//Proceedings of the 1st Workshop on Deep Learning for Recommender Systems.Boston, USA, 2016:7-10.

[4] Guo H,Tang R, Ye Y, et al. DeepFM:A factorization-machine based neural network for CTR prediction//Proceedings of the 26th International Joint Conference on Artificial Intelligence.Melbourne,Australia, 2017:1725-1731.

[5] Chen C, Meng X, Xu Z, et al. Location-aware personalized news recommendation with deep semantic analysis. IEEE Access, 2017:173-182.

[6] Rendle S. Factorization machines//Proceedings of the 2010 IEEE 10th International Conference on Data Mining. Sydney, Australia, 2010:995-1000.

[7] LYU L, MEDO M, YEUNG C, et al. Recommender systems［J］. Physics Reports, 2012, 519(1):1－50.

[8] Shan Y, Hoens T R, Jiao J, et al. Deep crossing: Web-scale modeling without manually crafted combinatorial features. SIGKDD 2016: 255-262.

[9] Song W, Shi C, Xiao Z, et al. Autoint: Automatic feature interaction learning via self-attentive neural networks. CIKM 2019: 1161-1170.

[10] Chen Q, Zhao H, Li W, et al. Behavior sequence transformer for e-commerce recommendation in Alibaba. Proceedings of the 1st International Workshop on Deep Learning Practice for High-Dimensional Sparse Data. 2019: 1-4.

[11] Hao-Jun Michael Shi;Dheevatsa Mudigere;Maxim Naumov; Jiyan Yang Compositional Embeddings Using Complementary Partitions for Memory-Efficient Recommendation Systems [C], 2020.

[12] Antonio Ginart; Maxim Naumov; Dheevatsa Mudigere; Jiyan Yang; James Zou Mixed Dimension Embeddings with Application to Memory-Efficient Recommendation Systems. [J] IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2019.

[13] Hao-Jun Michael Shi; Dheevatsa Mudigere; Maxim Naumov; Jiyan Yang Compositional Embeddings Using Complementary Partitions for Memory-Efficient Recommendation Systems. [J] IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2019.

[14] Yongxin Yang and Timothy Hospedales. 2016. Deep multi-task representation learning: A tensor factorisation approach. arXiv preprint arXiv:1605.06391 (2016).

[15] RAMEZANI M, MORADI P, AKHLAGHIAN F. A pattern mining approach to enhance the accuracy of collaborative filtering in sparse data domains [J]. Physica A: Statistical Mechanics and its Applications, 2014, 408:72–84.

[16] ZHANG F, YUAN N J, LIAN D, et al. Collaborative Knowledge Base Embedding for Recommender Systems[C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 353–362.

Research on Visibility Estimation Model Based on DenseNet

Guang Li
Department of Launch and Power Engineering,
Rocket Officer Academy
Wei'fang, China
Email: 865413666@qq.com

Zhiqiang Chang
Department of Launch and Power Engineering,
Rocket Officer Academy
Wei'fang, China
Email: 865413666@qq.com

***Abstract*—In recent years, the road visibility detection method based on video has been paid more and more attention. It has overcome the deficiency of laser visibility meter to some extent. Deep learning has a good effect in image processing and analysis. This paper firstly analyzes the current situation of deep learning, and then compares DenseNet and ResNet to propose a visibility estimation model based on deep DenseNet. The model firstly integrates airport video data and visibility data. Secondly, the DenseNet algorithm is used to automatically extract the features of the airport data set. Finally, Softmax classifier is constructed to evaluate the visibility accuracy. They reduce the problem of disappearing gradient, enhance feature propagation, encourage functional reuse, and greatly reduce the number of parameters, well train the deep model, has a good visibility estimation effect. On this basis, this paper based on Canny operator lane dividing line extraction edge extraction and visibility analysis based on edge detection, and do the corresponding test. Finally, a video visibility analysis model based on Kalman filter is built based on the given data, and Gaussian process regression model is used to predict the fog change trend.**

Keywords-Densenet; Visibility Estimation; Softmax Classifier

I. INTRODUCTION

A. Background

Visibility is a common indicator in meteorology, highway traffic and aircraft flight. The unit is usually meters. The main factors affecting visibility are fog and haze. As we all know, visibility is very important to highway safety. When visibility is very low, highway managers usually close the road in order to drive safety. However, in the aviation field, runway visibility is used to reflect the size of fog and haze near the airport, which is defined as the maximum distance at one end of the runway along the direction of the runway to identify the runway or the target object close to the runway (runway side lights at night). Normally, flights are banned when visibility at the airport is only about 400 meters. When visibility at the airport is only about 600 to 800 meters, flights can take off and land normally. However, due to security concerns, the airport will take temporary measures to control the flight flow, widening the flight departure interval, which is easy to cause flight delays. Therefore, visibility prediction is the highway management department and airlines are very concerned about the problem.

In recent years, video-based road condition (runway) visibility detection method [1,2] has attracted people's attention, which to some extent

overcomes the deficiency of laser visibility meter. Video visibility detection method is a combination of atmospheric optical analysis, image processing and artificial intelligence technology, through the analysis and processing of video images, establish the relationship between the video image and the real scene, and then calculate the visibility value indirectly according to the changes of image features. Wang Yaxue et al [3] established estimation and prediction models for visibility under different fog conditions. Through the establishment of multiple classification of multiple regression model. Han Huihui [4] introduced the ideas of closed-loop control and transfer learning into deep convolutional neural networks and designed a model of intelligent fog classification and visibility estimation based on deep transfer learning. Liu Jianlei et al. [5] analyzed the characteristics of anisotropy, continuity and horizontality of the inflection point line. Then, an inflection line detection filter is constructed according to these characteristics to improve the accuracy and speed of the detection. With the application of machine learning in various fields, the strong feature learning ability of deep learning network model has been applied to the classification and prediction of meteorological information, which has played a great role in promoting the improvement of forecast accuracy. In foreign countries, the artificial neural network model was initially used to forecast the visibility of winter weather in Milan, Italy [6], and the effect was better than the traditional method. By extending the standard probabilistic neural network method, Bremnes et al. [7] used multiple models to obtain an average output to improve the visibility forecast. Marzban et al. [8] combined the output of the numerical model with the ground observation data, and mapped the data to 39 different neural networks. The results showed that the performance of neural networks was generally due to logistic regression and MOS model. In China, Zhou Yongjiang et al. [9] used meteorological parameters such as temperature, air pressure, atmospheric precipitable water volume and PM 2.5 data of the same period to establish a haze prediction model integrating time series network and regression network to detect and forecast haze weather, proving that the fusion network model with meteorological parameters is more adaptable and accurate than the single network model.

In order to solve the visibility prediction problem in foggy weather, this paper constructs the visibility estimation model mainly through the method based on deep learning. The following is the introduction of deep learning-related issues and the main work of this paper.

B. Research status of deep learning technology

The concept of deep learning stems from research on artificial neural networks. Deep learning is a feature learning method that transforms raw data into a higher level and more abstract representation through some simple but nonlinear models. A more abstract high-level representation (attribute category or feature) is formed by combining low-level features to discover distributed feature representation of data [9]. In the process of learning the internal rules and representation hierarchy of sample data, deep learning has the ability to analyze, learn and interpret the acquired text, image and sound data, making it achieve remarkable achievements in image recognition, speech recognition, natural language understanding, weather prediction, gene expression, content recommendation and other aspects. Nowadays, deep learning plays a crucial role in various tasks of image analysis by discovering or learning information features that describe the inherent laws or patterns of data.

C. The Comparison between DenseNet and ResNet

In the field of computer vision, convolutional neural network (CNN) has become the most mainstream method. A milestone event in the history of CNN is the emergence of ResNet model, which can train deeper CNN model and achieve higher accuracy. The core of the ResNet model is to train a deeper CNN network by establishing a "short-circuit connection" between the front layer and the back layer, which facilitates the back propagation of the gradient during the training process. The DenseNet(Densely connected convolutional networks) model has the same basic idea as ResNet, but it establishes the dense connection between all the front layers and the back layers. Another feature of DenseNet is feature reuse through the concatenation of features on channels. These features allow DenseNet to achieve better performance than ResNet with fewer parameters and computational costs,

II. Related Work

Based on the video data and visibility data of an airport, this paper establishes a deep learning model for visibility estimation based on video data. The main work is to carry out data fusion of the video data and visibility data of the airport. Secondly, the convolutional network DenseNet algorithm is used to carry out automatic feature extraction of the merged airport data set. Softmax classifier was built for visibility and accuracy evaluation.

Data fusion technology refers to the information processing technology that utilizes the computer to analyze and synthesize the observed information obtained in time order under certain criteria to complete the required decision-making and evaluation tasks. Data layer fusion: fusion is carried out directly on the collected original data layer, and data synthesis and analysis are carried out before the original measurements of various sensors are processed. Data layer fusion generally adopts centralized fusion system for fusion processing

A. Camera calibration

In this paper, the mapping relationship between image coordinates and road coordinates is established for the video images collected by the camera, and the image distance information is converted into road distance information. The mapping relationship between image coordinates and road coordinates is completed by camera self-calibration technology, and its working steps are as follows:

1) Establish the road condition camera imaging model, as shown in Figure 3, in which three coordinate systems are defined, among which the ground coordinate system XW-YW-ZW and the camera coordinate system XC-YC-ZC are used to represent the three-dimensional space; Image plane coordinate system Xf -- Yf represents the imaging plane. The world coordinate system is established, whose origin is the intersection point of the camera optical axis and the ground. The camera coordinate system is established, and the origin is the camera optical center position. Set the distance between the optical center of the camera and the origin of the world coordinate system as l, the pitch Angle of the camera as t, the declination Angle as p, and the rotation Angle as s, and the area between parallel lines on the ground plane corresponds to the highway pavement in the camera's view field.

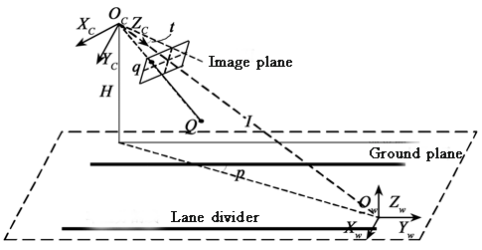


Figure 1. Highway camera model

2) Based on the defined camera spatial azimuth parameter [9], the coordinate transformation relationship between the ground coordinate system and the two-dimensional image coordinate system under the ideal penetration model can be established, as shown in Equations (1) and (2).

$$X_w = \frac{\left[l\sin p(x_f \sin s + y_f \cos s) + l\cos p\sin t(x_f \cos s + y_f \sin s)\right]}{x_f \cos t\sin s + y_f \cos t\cos s + f\sin t} \quad (1)$$

$$Y_w = \frac{\left[-l\cos p(x_f \sin s + y_f \cos s) + l\sin p\sin t(x_f \cos s - y_f \sin s)\right]}{x_f \cos t\sin s + y_f \cos t\cos s + f\sin t} \quad (2)$$

3) Parameter solution. The corresponding relationship between uncalibrated camera parameters and image feature parameters is established using highway lane lines as reference. The parallelogram based on the corner of the divider is used as the calibration block. According to the parallel correspondence between corner points, unknown camera parameters p, t, s, f, l in Equations (8) and (9) can be solved, as shown in Equations (1) and (2).

$$t = -\arcsin\left\{\frac{v_0^{\,2}(V_A - V_B + V_C - V_D)}{A \times B}\right\} \quad (3)$$

Among them:

$$A = \left[(V_0 - V_D)u_A - (V_0 - V_C)u_B + (V_0 - V_B)u_C - (V_0 - V_A)u_D\right] \quad (4)$$

$$B = \left[-(V_0 - V_B)u_A + (V_0 - V_A)u_B - (V_0 - V_D)u_C + (V_0 - V_C)u_D\right] \quad (5)$$

$$\left.\begin{array}{l} f = V_0 / \tan(t), l = \dfrac{H}{\sin t}, \tan s = -\dfrac{V_0 - V_1}{u_0 - u_1} \\ p = \arctan\left\{\dfrac{C * \sin(t)}{V_0(V_A - V_B + V_C - V_D)}\right\} \end{array}\right\} \quad (6)$$

Among them:

$$C = \left[(V_0 - V_D)u_A + (V_0 - V_C)u_B + (V_0 - V_B)u_C - (V_0 - V_A)u_D\right] \quad (7)$$

Where: (u,v) is the image coordinate system, (u0,v0) is the extinction point determined by the lines xaxd and xbxc, (u1,v1) is the extinction point determined by the lines xaxb and xdxc ,H camera column height.

B. Description of DenseNet network

Further, more accurate and effective training can be performed if the convolutional network contains shorter connections between the input layer and the output layer. In this paper, we accept this observation and introduce a dense convolutional network (DenseNet) that connects each layer to each other in a feedforward manner. A traditional convolutional network with L layers has L connections -- one connection between each layer and the one that follows -- and our network has L (L + 1) / 2 direct connections. For each layer, the feature maps of all previous layers are used as inputs, and its own feature maps are used as inputs for all subsequent layers. DenseNets have several compelling advantages: they alleviate the problem of disappearing gradients, enhance feature propagation, encourage functional reuse, and dramatically reduce the number of parameters.

The structure of the DenseNet network is shown in the figure 2. Each layer of the network is connected to all subsequent layers in a dense connection way.

C. Network backpropagation

In the training of DenseNet model, error information of training samples can be transmitted Back to the hidden layer through Back Propagation (BP) algorithm to realize constant updating and iteration of weight matrix between hidden layers until network convergence.

BP network is a feedforward neural network, which was proposed in 1985. In this kind of network, there are two kinds of signals flowing: one is the working signal, which is the input signal

applied and propagated forward until the actual output is generated at the output end of the signal, is a function of the input and weight; The second is the error signal. The difference between the actual network output and the theoretical output is the error, which starts from the output end and propagates back layer by layer. The output of the JTH unit at the output end in the NTH iteration is $y_j(n)$, then the error signal of the unit is $e_j(n) = d_j(n) - y_j(n)$, and the square error of unit j is defined as $\frac{1}{2} y_j^2(n)$. The instantaneous value of the total square error of the output is:

$$\varepsilon(n) = \frac{1}{2} \sum_j \mathrm{e}_j^2(n) \tag{8}$$

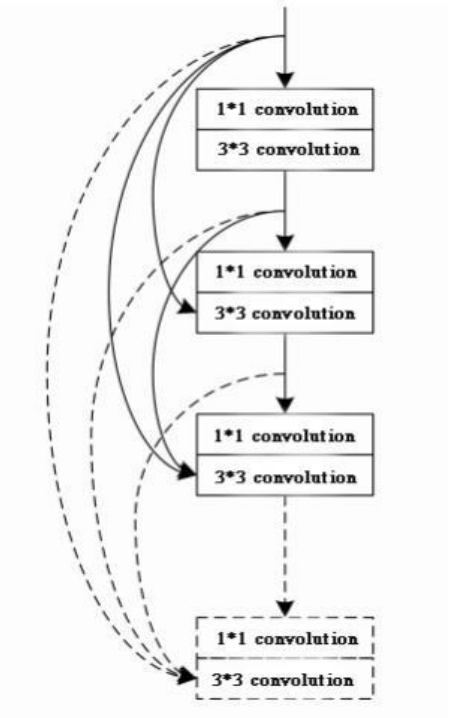


Figure 2. Schematic diagram of structural network

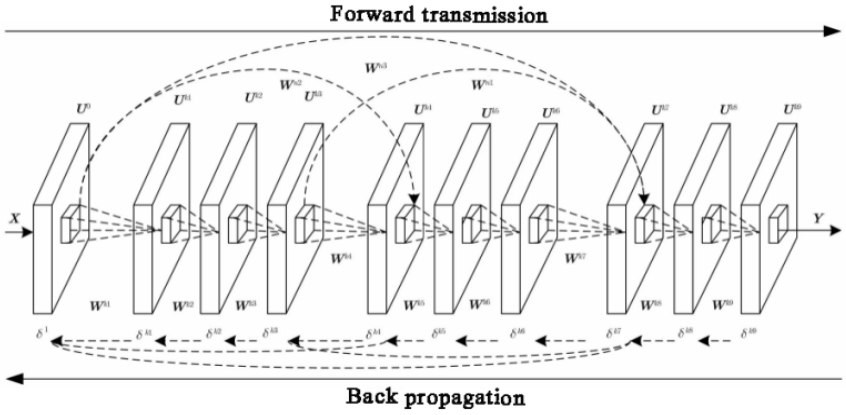


Figure 3. Training of DenseNet

Densenet can not only make efficient use of multi-dimensional feature information through the "feature re-calibration" strategy, but also slow down the attenuation of error terms of each hidden layer through the reverse transmission mechanism of the network itself, ensure the stability of gradient weight information, and enhance the learning and expression ability of the deep network, so as to further improve the network performance.

The data set used in this paper is the airport video and airport AMOS observations provided by the 2020 Modeling Competition. Firstly, the original airport video information and meteorological data were cleaned. Using python data analysis technology to carry out data fusion

of features; In addition, when using DenseNet for feature extraction, direct input of disordered discrete data will mislead the internal training mechanism of the network and confuse the importance of features. Discrete data is encoded by assigning values in turn according to the frequency of variable occurrence. For continuous data, feature values are mapped to [0,1] by minmax normalization to remove dimensional differences between features.

The classification prediction module mainly consists of three layers of network structure: global pooling layer, full connection layer and Softmax classifier. Through the combined operation of the three-layer network structure, the n×n×k dimension size matrix is converted into the probability corresponding to the occurrence of each category. Figure 4 below shows the integrated visibility analysis Organization.

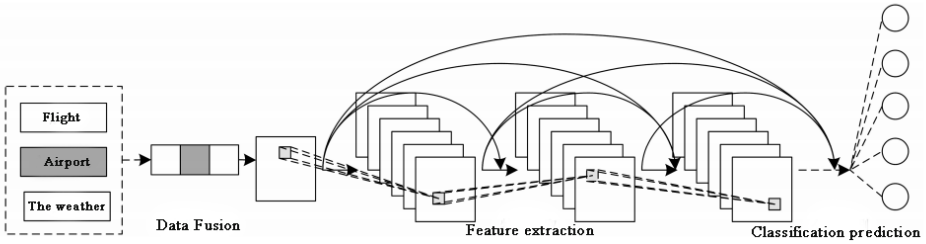


Figure 4. Airport visibility analysis structure

III. ANALYSIS OF EXPERIMENTAL RESULTS

The airport video used in the experiment in this paper has a total of 994379 samples of image data captured by python. The visibility of the human eye is 3000 meters, divided into 27 categories.

The network model is shown in the figure 5.

Classification accuracy: The depth network verifies the degree of model fitting mainly through the final accuracy of the test set, that is, the ratio of the sum of all correctly predicted flight data to the total number of test sets, expressed as:

$$\omega = \sum_{i=0}^{i=q} TP_i / N \tag{9}$$

Where, N is the total number of test sets, q is the number of visibility levels, the maximum is 27, TP_i represents the number of correct predictions in visibility levels.

Model: "dense_net"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	multiple	3200
batch_normalization (BatchNo	multiple	256
max_pooling2d (MaxPooling2D)	multiple	0
dense_block (DenseBlock)	multiple	209280
transition_layer (Transition	multiple	19296
dense_block_1 (DenseBlock)	multiple	226176
transition_layer_1 (Transiti	multiple	26096
dense_block_2 (DenseBlock)	multiple	234624
transition_layer_2 (Transiti	multiple	29880
dense_block_3 (DenseBlock)	multiple	238848
global_average_pooling2d (Gl	multiple	0
dense (Dense)	multiple	6723

Total params: 994,379
Trainable params: 984,171
Non-trainable params: 10,208

Figure 5. Network model

TABLE I. ACCURACY AND LOSS VALUES OF DENSENET

NUMBER OF NETWORK LAYERS	DENSENET(%)	LOSS
17	94.14	0.1459

Generally, the network judges whether the model finally converges by the loss value of the training set. The smaller the loss value is, the closer the predicted value of the model is to the real value. With the increase of the number of iterations, the model gradually converges and the loss value finally decreases to a fixed range. The loss value of DenseNet is reduced to about 0.15, and DenseNet can train the deep model well.

IV. FOLLOW-UP WORK

The video-based visibility detection method of road condition (runway) has attracted more and more attention. It overcomes the deficiency of laser visibility instrument to some extent. Video visibility detection method is a combination of atmospheric optical analysis, image processing and artificial intelligence technology, through the analysis and processing of video images, establish the relationship between the video image and the real scene, and then calculate the visibility value indirectly according to the changes of image features. In the follow-up work, we mainly focus on the edge extraction of lane divider based on Canny operator and the visibility analysis of single image based on edge detection, and do corresponding tests. Finally, we make the video visibility analysis model based on Kalman filter on the given data. In addition, a video provided by the mathematical modeling topic was used to draw the curve of the highway visibility changing with time in this period.

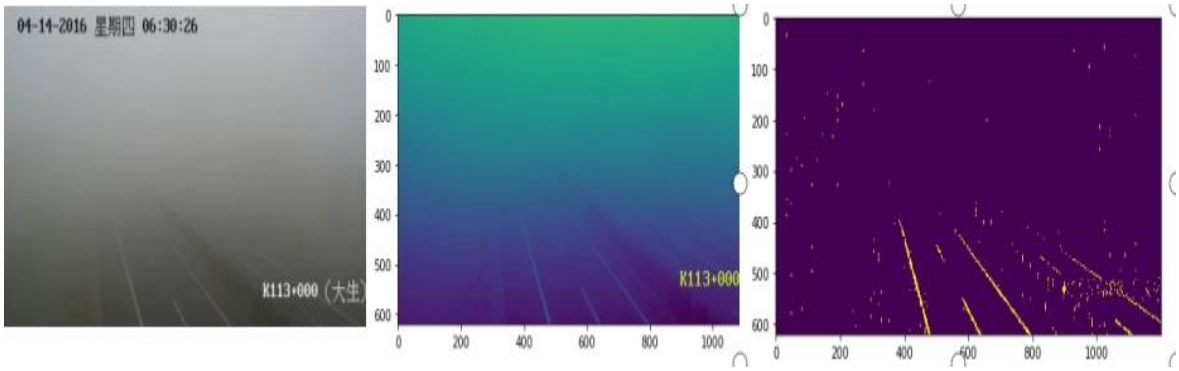


Figure 6. Canny operator edge extraction results

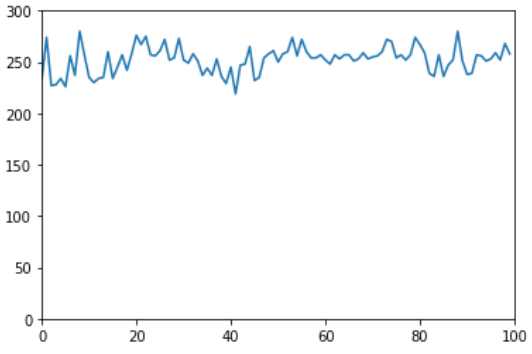


Figure 7. Highway visibility curve with time

Finally, a mathematical model is established to predict the change trend (worsening or weakening) and when the fog will disperse (reaching the specified visibility, such as MOR=150m). In this paper, Gaussian process regression model is proposed to predict the change trend of fog. Because Gauss process (GP) is a general supervised learning method designed to solve regression and probabilistic classification problems.

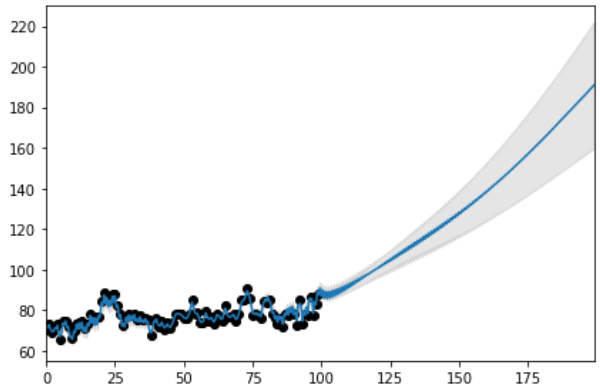


Figure 8. Gaussian process regression

V. CONCLUSION

In recent years, the video-based road condition (runway) visibility detection method has been paid more and more attention. It has overcome the deficiency of laser visibility meter to some extent. In fact, fogs form and dissipate in their own way, often related to near-ground meteorological factors. Based on the video data and visibility data of an airport, this paper establishes a deep learning model for visibility estimation based on video data. They reduce the problem of disappearing gradient, enhance feature propagation, encourage functional reuse, and greatly reduce the number of parameters. The loss value of DenseNet is reduced to about 0.15, and the deep model is well trained. With good visibility estimation effect, the visibility estimation model is used to forecast the dissipation of fog.

With the higher and higher level of numerical prediction, the accuracy of weather forecast is also getting higher and higher. Visibility estimation based on deep learning can be used to predict the dissipation of fog in the future, thus improving the transformation of weather to decision-making information and improving service quality.

REFERENCES

[1] S. K. Nayar, S. G. Narasimhan, Vision in bad weather, ICCV '99.

[2] 2.R. T. Tan, Visibility in bad weather from a single image, CVPR, 2008

[3] Wang Yaxue, Li Cheng, Liu Ting. Visibility Estimation and Prediction under Fog -- Based on the 2020 Graduate Mathematical Contest in Modeling E [J]. 2021.

[4] Han H H. Research on intelligent fog classification and visibility estimation method based on deep transfer learning.

[5] Liu Jianlei, Liu Xiaoliang. Fog visibility detection algorithm based on inflection line [J]. Journal of Computer Applications, 2015, 35(2):528-530.

[6] Pasini, A., Pelino, V. and Potesta, S. (2001) A Neural Network Model for Visibility Nowcasting from Surface Observations: Results and Sensitivity to Physical Input Variables. Journal of Geophysical Research, 106, 14951-14959. https://doi.org/10.1029/2001JD900134

[7] Bremnes, J.B. and Michaelides, S.C. (2007) Probabilistic Visibility Forecasting Using Neural Networks. Pure & Applied Geophysics, 164, 1365-1381.

[8] Marzban, C., Leyton, S. and Colman, B. (2006) Ceiling and Visibility Forecasts via Neural Networks. Weather & Forecasting, 22, 466-479. https://doi.org/10.1175/WAF994.1

[9] Zhou Yongjiang, Yao Yibin, Yan Xiao, et al. Prediction of haze by BP neural network based on GNSS meteorological parameters [J]. JGG, 2019, 39(11): 1148-1152.

[10] Li Bo, Dong Rong, Chen Qimei. Video Contrast Road Visibility Detection without Manual Marking [J]. Journal of Computer Aided Design & Computer Graphics,2009,21(11):1575-1582. (in Chinese)

[11] Xiao Hong, Cao Maojun, LI Panchi, WANG Haiying, Process Neural network Training Based on Piecewise Linear Interpolation [J]. Computer Engineering, 37(20) : 211-212, 215,2011.

[12] Christos Sakaridis · Dengxin Dai · Luc Van Gool.Semantic Foggy Scene Understanding with Synthetic Data, International J. Computer Vision, 2018, 3

[13] N. Hautiere, J-P Tarel, J Lavenant D. Aubert, Automatic fog detection and estimation of visibility distance through use of onboard camera, Machine Vision and Application, 2006, 17（1）: 8-20

[14] Gao Huang,Zhuang Liu,Laurens van der Maaten.Densely Connected Convolutional Networks.

A Compound Optimization Greedy Strategy with Reverse Correction Mechanism

Han Shen
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
e-mail: sunny_shine_zj@163.com

Zhongsheng Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
e-mail: wzhsh1681@163.com

***Abstract*—Greedy strategy is an algorithm thinking with local optimization as the core idea, but only when the problem has no after-effect, the global optimization can be achieved. Therefore, greedy strategy is not the first choice for researchers to solve the problem. Based on the greedy strategy, this paper adds the mechanism of reverse correction thinking, transfers the local optimal solution to the global optimal solution, and puts forward a compound optimal greedy strategy integrating reverse correction thinking. Based on the actual application scenario of blood robot operating costs, the overall "simple greedy strategy model" is constructed and tested based on the greedy strategy as the main modeling basis according to the application needs. On this basis, the interaction relationship between local optimal solutions is deeply analyzed, and the reverse correction mechanism is integrated to optimize the system through the two steps of reverse allocation and reverse merge repair. Gradually improve the model to get the optimized "reverse modified greedy strategy model", the algorithm can effectively reduce the operating cost. On this basis, in order to test the optimization effect, the effectiveness and stability of the reverse correction mechanism were verified by modifying some parameters of the application scene and randomly generating multiple arrays for re-test, etc., and new parameters were selected to re-run the application scene, and satisfactory verification results were obtained. Compared with other modeling ideas of the same topic, this model weakens the expression of the overall function and emphasizes the change relationship and action mechanism between data, and obtains better operation results. Greedy strategy is very conducive to the analysis of the relationship between requirements, constraints and variables. According to the actual application needs, combined with the mathematical analysis method, the reverse correction mechanism is added to the greedy strategy modeling. In the demand sequence test of 100 groups of simulation, the maximum saving rate can be close to 1.6%, while the lowest saving rate is less than 0.6%, and the average saving rate is 0.9677%. It can save tens of thousands of operating costs for application scenarios.**

Keywords-Greedy Strategy; Lowest Operating Cost; Reverse Correction; Fallback Mechanism; Compound Algorithm

I. INTRODUCTION

Greedy strategy refers to a solving method that starts from the initial state of the problem and obtains the optimal solution or better solution through several times of greedy selection, also known as greedy strategy, greedy algorithm or greedy algorithm. The most important feature of greedy strategy is the local optimal solution, which is an algorithm strategy that depends on the choices made rather than the choices not made. From a global perspective, the local optimal solution does not mean the global optimal solution. The greedy strategy used in each local optimal solution may fight with each other, and the latter may affect the former, leading to the final global result is not ideal. The after-effect is one of the biggest influencing factors of greedy strategy. Therefore, greedy strategy has great limitations in solving problems such as multi-objective programming and dynamic programming, so it is not the first choice for researchers to solve problems in the past.

From the perspective of mathematical thinking, it is a logical thinking perspective to find the global optimal solution by means of the local optimal solution. However, in the actual application of greedy algorithm, local contradictions and mutual influences are likely to occur, resulting in poor global performance or even "collapse". However, this does not mean that greedy strategy is nothing. It may also work well.

Based on the application scenario of "the lowest operating cost of vascular robot", this paper incorporates a "reverse correction" mechanism into the greedy algorithm. Greedy strategy can meet the operational requirements of the problem set. On this basis, according to the requirement of "lowest operating cost", two key points that can be corrected are further analyzed from the mathematical perspective, and two main correction modules of reverse allocation and reverse merger are established by using "reverse thinking". In the original model, the modified part is searched in reverse, and the greedy strategy of the "local best" principle is still adopted for optimization, and a better optimization result is obtained. At the same time, on the basis of the original application, some application parameters were modified for comprehensive test. In addition, the method of randomly generating array was adopted to generate multiple groups of different data for further test. It was found that the "compound improved greedy strategy integrating reverse correction mechanism" was stable and effective. It can have better operation results and clearer action mechanism.

II. APPLICAIOTN SCENARIOS

A vascular robot used by a hospital consists of a Container Boat and four operators. A Container Boat (similar to a submarine) that is powered and swims in the blood; The vessel is equipped with a human-like operator hand around the Container Boat, with a biological brain and a mechanical arm. The biological brain controls the mechanical arm to work. The operator can remove, install and replace from the Container Boats. It is necessary to combine the actual demand with the cost, operation, maintenance, damage and other requirements of the vascular robot, and calculate the number and cost of parts purchase, maintenance and training of this type of vascular robot on a weekly basis. On the basis of meeting the treatment needs, the operating cost can be minimized. The requirements for weeks 1 to 104 are shown in Table 1.

TABLE I. NUMBER OF VASCULAR ROBOTS USED IN WEEKS 1-104

Week 1-8	11	5	4	7	16	6	5	7
Week 9-16	13	6	5	7	12	5	4	6
Week 17-24	9	5	5	11	29	21	17	20
Week 25-32	27	13	9	10	16	6	5	7
Week 33-40	11	5	5	6	12	7	7	10
Week 41-48	15	10	9	11	15	10	10	16
Week 49-56	26	21	23	36	50	45	45	49
Week 57-64	57	43	40	44	52	43	42	45
Week 65-72	52	41	39	41	48	35	34	35
Week 73-80	42	34	36	43	55	48	54	65
Week 81-88	80	70	74	85	101	89	88	90
Week 89-96	100	87	88	89	104	89	89	90
Week 97-104	106	96	94	99	109	99	96	102

The specific requirements for the composition, purchase, operation and maintenance of the vascular robot are as follows:

a) A vascular robot is composed of a Container Boat and four operators. Each vascular robot works for a week.

b) There were 13 usable Container Boats and 50 skilled operators before the start.

c) The newly purchased Container Boat needs to undergo a week of inspection and commissioning before it can be put into use. The commissioning of a Container Boat requires no additional Container Boat "guidance" and no additional cost.

d) Container boats do not need maintenance after use, can be used continuously, but not used container boats need maintenance. The maintenance fee for container boats is 10 yuan per week.

e) The newly purchased operators (new workers) need to learn the good operators (skilled workers) under the "guidance" (a certain proportion, not one to one, each skilled operator can "guide" the number of new operators is not more than 20), for one week of biological learning (training) can be put into use. For the purpose of differentiation, the newly purchased "novice" operators who need training are called trainees, and the "skilled" operators who "guide" trainees to

train are called coaches. The "trainer" must be a working operator, that is, the one who has just been removed from the patient's blood Container Boats must enter the maintenance process and cannot serve as the "trainer". The "coach" cannot work in the patient's blood Container Boats during the week, but does not need maintenance. He can work directly the next week and continue to serve as the "guidance" work. The unit price of training is not divided into "coach" and "trainee", and the unified price is 10 yuan per week.

f) The operator needs a week of maintenance to start work again, if there is no arrangement, it needs to be maintained all the time. Therefore, there are two reasons for the maintenance of the operating hand, that is, just removed from the patient's blood Container Boats and idle do not work. In order to facilitate the distinction, the idle do not work operator is called maintenance, and the maintenance of the operating hand just removed from the blood Container Boats is called maintenance. Maintenance and maintenance of the operator is priced at 5 yuan per week.

g) Purchased Container Boats and operators will arrive at the beginning of each week and immediately arrange for inspection, debugging and biology learning (training).

h) Every week, 10 percent of vascular robots are destroyed (rounded off), and for each vascular robot, a Container Boats and four operators are destroyed. The number of damaged vascular robots is calculated by multiplying the number of vascular robots that actually entered the patient's blood Container Boats in that week by a specific proportion, while other vascular robots in that week are not included in the damaged base.

i) The unit price of container boats is 200 yuan per Container Boat when the purchase quantity is not more than 5; If the one-time purchase quantity of container boats is more than 5 but not more than 10, the unit price of the part of more than 5 boats is 180 yuan per boat; When the one-time purchase quantity of container boats is more than 10, the unit price of the part over 10 Container Boats is 160 yuan/piece.

j) Manipulators preferential policy for the one-time purchase of not more than 20 when the unit price of 100 yuan/a; When the operator purchases more than 20 but not more than 40 pieces at a time, the unit price of the part exceeding 20 pieces is 90 yuan/piece; When the operator purchases more than 40 units at a time, the unit price of that part exceeding 40 units is 80 yuan/unit.

III. Greedy Strate Modeling

This application scenario studies the problem of the lowest cost, and the problem setting conditions are very rich, and the direction of the modeling results is very clear -- the lowest cost to meet the requirements. According to the requirements of the problem, the composition of the cost mainly lies in the purchase, maintenance and training, so according to the local optimal idea of greedy strategy, on a weekly basis, in principle, buy less rather than more, buy near rather than far. "Buy when you need it" is the core initial design concept of this model.

In addition, it should be noted that either the Container Boats or the operator cannot be directly put into use in the week of purchase, which requires a week of debugging or training. The training of the trainee operator also requires the "guidance" of the trainer who can work normally, and the trainer in the training stage cannot participate in the normal work of the vascular robot. Therefore, the requirements of the following week need to be prepared one week in advance. In case of reverse data update, the global data will be updated in the way of restarting data for each forward update.

A. Analytical Variable Relationship

Although the manipulators and Container Boats of the vascular robot in question work at the same time, the actual data of the two do not interfere with each other, so the two can be considered separately in modeling and calculation. Obviously, compared with the analysis of the operator who needs a "coach" to "guide" the "student", the variable analysis of the container boat is relatively simple, which only needs "debugging" rather than "guidance". Therefore, the variable of the container boat is analyzed first.

With week as the minimum time unit, Container Boats variables include: total number of the week, the number of the week purchased, the number of the week debugging, the number of the week able to work, the number of the week needed to work, the free amount of the week, the maintenance amount of the week.

With week as the minimum time unit, operator variables include: the total number of this week, the number of this week's purchases, the number of this week's trainees, the number of this week's coaches, the number of this week's work, the number of this week's work, the number of this week's spare maintenance, the amount of this week's maintenance.

Since the Container Boats and operator do not interfere with each other but their internal relations are complicated, their internal variable relations are analyzed separately.

B. *The Variable Relationship of Container Boats*

Data changes and updates of container boats are carried out on a weekly basis. Take a certain week (not the first or last week) as an example, and its process diagram is shown in Figure 1.

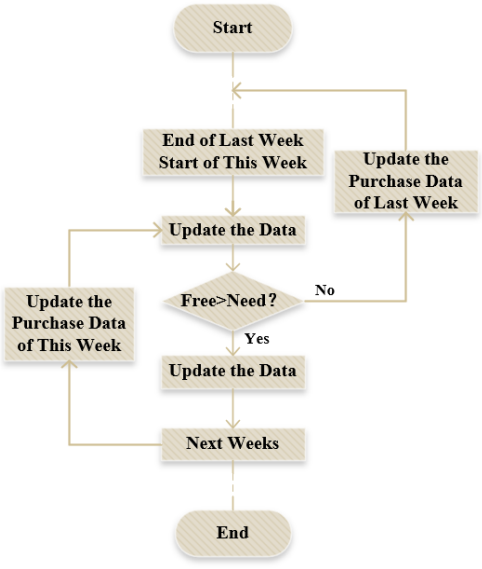


Figure. 1. Data update flow chart of Container Boats

a) The new container boat purchased this week is R_{buy-1}. Note that the purchase quantity is updated by the following weekly. According to hypothesis 1, debugging is needed in the week of purchase, so the number of debugging in this week $R_{study\text{-}1}$ is equal to the number of purchases in this week, as shown below:

$$R_{study\text{-}1} = R_{buy-1} \tag{1}$$

In practice, the number of debugging this week is always equal to the number of purchase this week, and this rule is always maintained in the subsequent data update. Therefore, the variable of debugging number can be removed when programming the model, but in order to conform to the process of the problem and the integrity of the analysis, the variable is kept during the analysis.

b) The total amount of container boats this week $R_{total-1}$ is variable, which includes the total container boats of last week $R_{total-0}$ and the newly purchased container boats of this week R_{buy-1}. At the same time, 10% of the loss in last week's work R_{use-0} should be deducted, namely:

$$R_{total-1} = R_{total-0} + R_{buy-1} - 10\% \times R_{use-0} \tag{2}$$

c) The Container Boats needed for this week R_{use-1} are obtained according to the needs of this week N_1. Each vessel robot carries one Container Boat, namely:

$$R_{use-1} = N_1 \tag{3}$$

d) Due to the commissioning of the newly purchased Container Boats, the Container Boats actually available for use this week R_{can-1} cannot include the newly purchased part, that is, only on the basis of the total amount of last week's Container Boats $R_{total-0}$, the work loss of last week should be subtracted. The Container Boats that can be used this week during the work of this week, include the actual work and need warranty due to spare time, namely, the following formula:

$$R_{can-1} = R_{total-1} - R_{buy-1} = R_{total-0} - 10\% \times R_{use-0} \tag{4}$$

e) If there are enough container boats available for use this week ($R_{can-1} \geq R_{use-1}$), then there will be free container boats R_{free-1} that need warranty.

Variable R_{keep-1} is established for the container boats that need warranty, and the relationship is as follows:

$$R_{free-1} = R_{can-1} - R_{use-1} \tag{5}$$

$$R_{keep-1} = R_{free-1} \tag{6}$$

Similar to the amount of debugging this week equals the amount of purchase this week, the number of free container boats and the number of container boats that need warranty is always the same, which can be deleted in the actual programming to build the model without affecting the whole. Two variables were retained in the analysis process, not only to conform to the process and analysis integrity of the problem set, but also because in the analysis of the operator's model, the free amount and the warranty amount were not equal, so as to compare and use them in the in-depth analysis of the subsequent data. This warranty variable can be removed during programming modeling, as long as it is matched in subsequent analysis.

f) If there are not enough working container boats for the week ($R_{can-1} < R_{use-1}$), you will need to purchase a supplement. Since the newly purchased Container Boats cannot be used directly, but needs to be debugger for a week, the demand gap of this week needs to be purchased last week, so that it can be used this week after training. The relationship is as follows:

$$R_{buy-0}(new) = R_{buy-0}(old) + R_{use-1} - R_{can-1} \tag{7}$$

Note the following:

If there was demand for purchases last week, then this week's demand should be added to last week's demand. But in fact, the change in last week's purchase quantity is only related to this week's demand gap, and next week's demand gap will only affect this week's purchase quantity but not last week's purchase quantity, so it will only change once at most. Therefore, direct assignment does not affect the result. But in both analysis models and programming modeling, it is important to keep in mind that the purchase amount itself should be an additive amount.

This week's demand gap caused the change in the purchase volume of last week, which indirectly affected the total number of container boats last week, the number of container boats debugging last week, the total number of container boats this week, the number of available services this week and other data. However, it did not affect the previous weeks, so relevant data can be updated. However, similar to the way that the purchase quantity is accumulated rather than assigned, it should be noted in both the analysis model and the actual programming modeling that for the case of reverse order affecting variables, each change should be sorted out from the beginning, so as to keep the data updated for the overall coordination and once the purchase quantity changes, the data should be updated from the first week.

g) After the completion of the above operation, the operation of the next week is the same as that of this week. Once the container boats that can be used cannot meet the needs, it is also necessary to increase this part of the gap on the basis of the previous week's purchase, and then update the relevant data.

h) End up buying policy on price adjustment strategy approximation as piecewise function, purchase quantity and purchase price can present certain functional relation, using $f(x)$ said gunner purchase policy function:

$$f(x) = \begin{cases} 200x, & x \le 5 \\ 100 + 180x, & 5 < x \le 10 \\ 300 + 160x, & x > 10 \end{cases} \tag{8}$$

C. The Variable Relationship of the Operator Hands

Compared with the updated data of Container Boats, the operator data needs more consideration of the roles of "trainee" and "coach". In addition to the difference between the operators they can provide and those who need to work, the number of "coaches" should also be considered. Take a week (not the first or last week) as an example, and its flow diagram is shown in Figure 2.

The specific affectation is as follows:

a) The new operator purchased this week is C_{buy-1}. Note that this purchase quantity is also updated by the following weekly updates. According to the assumption, the week of purchase is the training, so the number of trainees this week $C_{study-1}$ is equal to the number of purchases this week:

$$C_{study-1} = C_{buy-1} \tag{9}$$

In practice, the number of debugging this week is always equal to the number of purchase this week, and this rule is always maintained in the subsequent data update. Therefore, the variable of the number of students $C_{study-1}$ can be removed at the early stage of programming model building. However, in addition to conforming to the process of the problem set and the integrity of the analysis, we also found in the subsequent analysis that, the number of students is a variable that contributes to the optimization of the analysis and model, so it is retained during the analysis and modeling process.

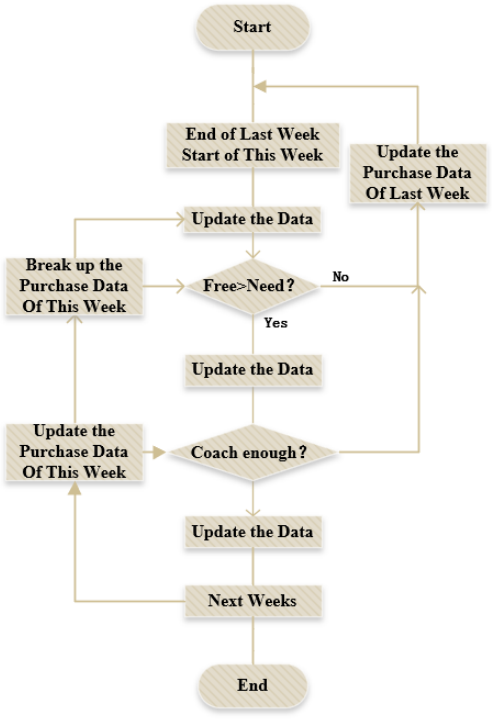


Figure. 2. Data update flow chart of the Operator Hands

It should also be noted that the number of trainers $C_{coach-1}$ should be configured for the trainee training of the operator. When the number of trainees is 0, no coach is needed for that week; When the number of trainees is not 0, coaches are needed in the week, and the number of coaches should be calculated according to the ratio. The matching assumption of the test environment is that one coach can guide 20 trainees at most. The conversion relationship is as follows:

b) The total number of operators this week is changed. Variable $C_{total\text{-}1}$ is established, which includes the total number of operators last week $C_{total-0}$ and the newly purchased operators this week C_{buy-1}. Meanwhile, the loss in use last week is subtracted, namely:

$$C_{total\text{-}1} = C_{total-0} + C_{buy-1} - 10\% \times C_{use-0} \tag{10}$$

c) The manipulators that need to be used this week C_{use-1} (to actually work in the patient's blood vessels) are obtained according to the demand of this week N. Each vascular robot carries 4 manipulators, namely:

$$C_{use-1} = 4N \tag{11}$$

d) The operators who participated in the actual work last week C_{use-0} need to maintain this week, so this part needs to set up a separate variable $C_{upkeep-1}$ to record for analysis and modeling. The lost part does not need to be maintained:

$$C_{upkeep-1} = 90\% \times C_{use-0} \tag{12}$$

Since the newly purchased operator needs training, the actual standby operator that can be used this week C_{can-1} does not include the newly purchased part, but also excludes the maintenance part of this week. In essence, it is the total number of last week minus the number of last week's work:

$$C_{can-1} = C_{total-1} - C_{buy-1} - C_{upkeep-1} = C_{total-0} - C_{use-0} \tag{13}$$

There are three parts of the task this week: the actual working operator C_{use-1}, the trainer operator $C_{coach-1}$ and the spare maintenance operator C_{free-1}.

e) First of all, consider whether the operator can meet the needs of the actual work. If the operator cannot provide the required hand, i.e. $C_{use-1} > C_{can-1}$, at this time, the demand gap of this part shall be fed back to last week, and the

purchase quantity of last week shall be accumulated:

$$C_{buy-0}(new) = C_{buy-0}(old) + C_{use-1} - C_{can-1} \quad (14)$$

At the same time, it should be noted that changes in the number of purchases may also lead to changes in the number of coaches and idle numbers of last week, or even new demand gaps (demand gaps will be generated due to the increase in the number of coaches and the shortage of idle numbers, which will be detailed in the subsequent analysis). Therefore, once the reverse demand update occurs, you need to update the data from scratch.

f) If the number of operators available this week is sufficient for regular work ($C_{use-1} \le C_{can-1}$), further consideration needs to be given to the number of trainers to train the trainees $C_{coach-1}$.

g) If the demand of the number of coaches cannot be met, i.e. ($C_{can-1} - C_{use-1} < C_{coach-1}$), then the demand gap of the number of coaches is generated. Similar to the previous step, the demand gap of this part needs to be reversely updated into the purchase data of last week:

$$C_{buy-0}(new) = C_{buy-0}(old) + C_{coach-1} - (C_{can-1} - C_{use-1}) \quad (15)$$

Similarly, after the data is updated backwards, all the data is updated from scratch.

h) If the demand for the number of trainers can also be met, i.e., ($C_{can-1} - C_{use-1} \ge C_{coach-1}$), then the number of free operators this week is:

$$C_{free-1} = C_{can-1} - C_{use-1} - C_{coach-1} \quad (16)$$

Note that the number of free operators in this part needs maintenance work, which should be distinguished from the operators in the maintenance part and also from the algorithm of the container boat.

i) After the above operations are completed, the operation of the next week is the same as that of this week. Once the available operators cannot meet the needs, it is also necessary to increase this part of the gap on the basis of the previous week's purchase, and then update the relevant data.

j) Purchasing policy of price adjustment strategy approximation as piecewise function, purchase quantity and purchase price can present certain functional relation, use of $g(x)$ operator to buy policy function:

$$g(x) = \begin{cases} 100x, & x \le 20 \\ 200 + 90x, & 20 < x \le 40 \\ 600 + 80x, & x > 40 \end{cases} \quad (17)$$

D. Programming

The python language and Jupyter platform are selected for the compilation. The logic is clear, which can be edited in blocks, and the data can be further analyzed on the basis of running.

1) Initialization

In order to facilitate global analysis and derivation, data storage and extraction, and subsequent analysis, one-dimensional array is used to store weekly data changes of each variable in python language environment (matrix storage can also be used). In addition, the package guide and preset data import are also completed in this step.

2) Write the model by parts

According to the separation and analysis of variables, the Container Boats and operator calculated separately and did not affect each other, so the two models were coded separately.

For the Container Boats part, under the preset premise and in accordance with the sequence of variable analysis, the value of this week's total amount, this week's available workload and this week's needed workload are updated respectively to judge the demand gap. If there is no demand gap, update the idle quantity and maintenance quantity and enter the next week calculation; If there is a shortfall, reverse add it to last week's purchases and update the data from the first week.

The operator part can be regarded as an extension of the Container Boat part. Modeling was carried out according to the sequence of variable analysis, and the total amount of this

week, the maintenance amount of this week, the available amount of this week, the workload of this week, the number of students of this week and the number of coaches of this week were updated respectively, and then judged step by step. If there is a gap between the available usage and the required workload, the reverse update purchase data is added to last week's purchase volume, and the data is updated from scratch; If the available amount can meet the workload, then judge whether it meets the demand of the number of trainers. If it does not meet the demand, reverse update the demand gap to the purchase amount of last week and update other data from scratch. If it meets the demand, update the free maintenance amount of this week and enter into the calculation of the next week.

After the monomer model of the two parts is established, the cycle structure corresponding to the number of weeks is loaded respectively and the function is packaged. The return value is each variable array updated respectively. See appendix for model code implementation.

3) Control the operation part

The individual parts of the operator and container boat are packaged as functions, and the running part simply calls the function, noting that the result of the function needs to be inherited with a preset number of arrays corresponding to the number of arrays each returns.

4) Calculate operation cost module

The calculation of operation cost adopts the summative method, setting the operator's purchase, operator's maintenance, operator's training, container boat's purchase, and container boat's maintenance. The cycle needs weekly, weekly sum, after the results are obtained, combined calculation is made according to the needs, and the required results are output.

5) Other results output

After the execution, you can call to read the finished data as needed.

E. The Results of the Model

The output results such as operator purchase array, container boat purchase array and operation and maintenance cost are summarized in Table 2.

TABLE II. THE OUTPUT OF THE SIMPLE GREEDY STRATEGY MODEL

The output of the simple greedy strategy model								
The purchase program of Container boat								
Week 1-8	0	0	0	5	0	0	0	1
Week 9-16	0	0	0	2	0	0	0	0
Week 17-24	0	0	2	19	0	0	0	7
Week 25-32	0	0	0	0	0	0	0	0
Week 33-40	0	0	0	0	0	0	0	4
Week 41-48	0	0	0	5	0	0	5	12
Week 49-56	0	2	15	18	0	4	18	13
Week 57-64	0	0	1	12	0	0	6	11
Week 65-72	0	0	2	11	0	0	0	10
Week 73-80	0	1	11	16	0	10	16	21
Week 81-88	0	9	18	24	0	6	11	19
Week 89-96	0	7	10	24	0	4	10	25
Week 97-104	1	8	14	20	1	7	16	0
Amount		484		Cost		91600		
The purchase program of operator hands								
Week 1-8	14	0	0	36	0	0	0	1
Week 9-16	0	0	0	8	0	0	0	0
Week 17-24	0	0	13	98	44	0	0	10
Week 25-32	0	0	0	0	0	0	0	0
Week 33-40	0	0	0	0	0	0	0	0
Week 41-48	0	0	0	0	0	0	18	68
Week 49-56	26	0	68	117	50	0	32	66
Week 57-64	0	0	0	36	10	0	9	57
Week 65-72	2	0	0	34	0	0	0	16
Week 73-80	10	0	47	89	37	19	89	126
Week 81-88	41	11	90	138	45	0	31	83
Week 89-96	24	0	35	99	36	0	23	104
Week 97-104	60	0	43	98	40	0	39	0
Amount		2290		Cost		311505		
Total Cost				403105				

IV. REVERSE CRRECTION MCHANISM

According to the greedy strategy thinking mode, according to the way of "need to buy", can perfectly solve the problem of "demand", but the "lowest operating cost" can be optimized, at this time need to reverse correction.

A. Backward Distribution

1) Analysis of source

As shown in Figure 3, it is the operator purchase demand matrix of the simple greedy algorithm of week 1-104. First, two consecutive

non-zero numbers in the result -- 18 and 68 are taken into consideration. It is noted that the free number of the next week corresponding to 68 is not 0, indicating that at least two more operators are purchased here, because the subsequent operator vacancies are directly added to the purchase volume of this week. This week created a gap in the number of coaches and continued to pass on purchases, but there is room for improvement. For the purposes of analysis, call purchase 18 last week and purchase 68 this week.

Simple - Operator Hands - Buy:

[0	14	0	0	36	0	0	0	1	0	0	0	8	0	0	0	0	0
0	13	98	44	0	0	10	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	18	68	26	0	68	117	50
0	32	66	0	0	0	36	10	0	9	57	2	0	0	34	0	0	0
16	10	0	47	89	37	19	89	126	41	11	90	138	45	0	31	83	24
0	35	99	36	0	23	104	60	0	43	98	40	0	39	0]			

Simple - Operator Hands - Free:

[0	5	0	24	12	0	1	39	32	0	1	28	21	0	5	32	26	4
6	17	0	2	5	41	36	0	20	81	88	56	68	106	100	74	79	99
93	63	57	72	57	22	18	36	28	0	0	13	0	2	4	2	0	3
6	4	0	4	8	53	30	3	0	18	0	2	3	34	16	0	8	45
30	0	1	5	0	2	3	0	0	4	1	0	0	4	2	12	0	3
5	11	0	3	5	21	0	3	6	9	0	3	5	11	0]			

Figure. 3. The purchase and idle situation of the Operator Hands in the Simple Mode

Combined with the background analysis box selected this part of the data source. Due to the demand gap next week, the purchase volume of this week is 68, that is, there are 68 students and 4 coaches this week. However, there is also a demand gap this week, which makes the purchase volume of last week is 18, but there is no purchase volume last week, so there is no demand gap last week, and even there are spare operators.

It can be assumed that if the purchase volume of 68 students this week is transferred to last week, and there are enough spare numbers to provide coaches last week, then the demand of this week will be reduced by 4, and the planned purchase volume of last week will be reduced from 13 to 9. Although the purchase in advance increases the maintenance cost of 68 operators for a free week (5 yuan per operator per week) by 340 yuan in total, the purchase quantity of 4 operators reduced (100 yuan per operator) is saved by 400 yuan, so the comprehensive calculation saves some money.

On this basis, if 4 of the 68 newly purchased operators are purchased and trained last week in advance, and the remaining 64 are purchased this week, then the 4 operators purchased in advance can be used for training this week, which can save the maintenance cost of 64 operators for a week totaling 320 yuan.

Based on the above analysis, the optimization method based on this example can save 380 yuan locally. The essence is to advance the purchase of some of the operators originally needed this week to the last week. The amount of this demand gap is "self-training", and there is no need to supplement the number of trainers.

2) Optimization Mechanism

After analyzing the reasons for this problem, it is found that there are too many operators purchased and not enough coaches, which leads to the need to increase the purchase amount of coach operators.

Return to the algorithm process of analyzing the operator, using the method of comparing demand step by step. If there is a demand gap next week, directly increase the purchase volume this week, and then update the data from the first week. When the data is updated this week, there may be such a situation that although the work required by the available supply is sufficient, the number of coaches will increase due to the increase in the purchase amount, and the number of coaches will not be enough in the next step, so the demand gap of this week will be accumulated to the last week. The solution is to buy quantity forward in advance.

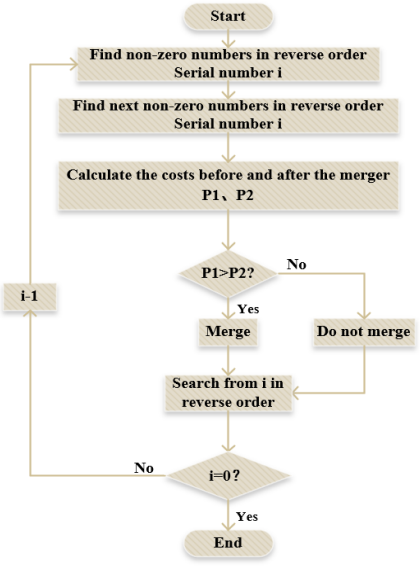


Figure. 4. Update flow chart of the Operator Hands after the first optimization

Based on the hypothesis, it can be seen that the changes of the purchase quantity and the number

of trainers this week will not lead to the changes of the available operators and the required operators this week. In addition, as the purchase quantity is advanced, the available number this week will only increase, so there will be no additional demand gap. In the previous example, when the run to this week cannot provide enough coaches, if one purchase of this week is brought forward to last week, the number of coaches available for this week increases by 1, the number of students that can be accommodated increases by 20, and the actual number of students decreases by 1. Since each time of the forward accumulation of demand gap in this model algorithm, an operation of re-running the update array from the first week will be added, so there is no need to set the purchase quantity of forward specifically, nor to consider whether the number of coaches in the previous week is sufficient, because these problems can be found during the re-running and solved with this improved idea. The flow chart, shown in Figure 4, only needs to be optimized when there are not enough coaches available.

3) Effect of Optimization

After the optimized model is loaded with the demand sequence, the purchase and idle sequence of the operator is output again, as shown in Figure 5. It can be found that the local problems analyzed above have been solved. Throughout the series, the other parts of the "buy one week and then idle one week" scenario have also been resolved, and have been "minimized" from a purchase volume perspective.

The final result after optimization is shown in Table 3. To compare the purchase plan of the operator, the red dot in Figure 6 represents the situation before optimization, and the blue dot represents the situation after optimization. It can be found that the optimization module does have an adjustment effect on the purchase plan. From a global perspective, due to the requirement of the overall number of operators needed, the sequence of weekly demand quantity is contingent, and a certain proportion of losses in the operation process, the total number of operators needed to be purchased is the same in the end. However, the operation and maintenance cost of operators is reduced by 495 yuan after optimization, which is because the "overbought" part in the early stage is transferred to the later stage. It reduces the maintenance cost. The optimizer does have an overall tuning effect along the way.

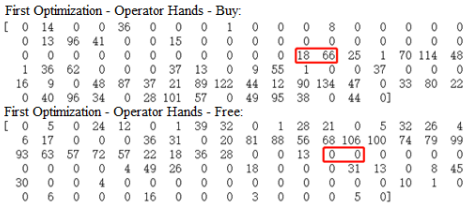


Figure. 5. The purchase and idle situation of the Operator Hands in the First Optimization Mode

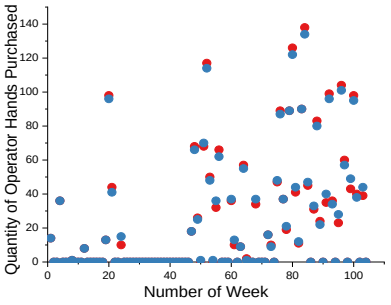


Figure. 6. Comparison of purchasing plans of Operator Hands before and after the first optimization

TABLE III. THE OUTPUT OF THE FIRST OPTIMIZATION

The output of the first optimization								
The purchase program of Container boat								
Week 1-8	0	0	0	5	0	0	0	1
Week 9-16	0	0	0	2	0	0	0	0
Week 17-24	0	0	2	19	0	0	0	7
Week 25-32	0	0	0	0	0	0	0	0
Week 33-40	0	0	0	0	0	0	0	4
Week 41-48	0	0	0	5	0	0	5	12
Week 49-56	0	2	15	18	0	4	8	13
Week 57-64	0	0	1	12	0	0	6	11
Week 65-72	0	0	2	11	0	0	0	10
Week 73-80	0	1	11	16	0	10	16	21
Week 81-88	0	9	18	24	0	6	11	19
Week 89-96	0	7	10	24	0	4	10	25
Week 97-104	1	8	14	20	1	7	16	0
Amount	484			Cost		91600		
The purchase program of operator hands								
Week 1-8	14	0	0	36	0	0	0	1
Week 9-16	0	0	0	8	0	0	0	0
Week 17-24	0	0	13	96	41	0	0	15
Week 25-32	0	0	0	0	0	0	0	0
Week 33-40	0	0	0	0	0	0	0	0
Week 41-48	0	0	0	0	0	0	18	66
Week 49-56	25	1	70	114	48	1	36	62
Week 57-64	0	0	0	37	13	0	9	55
Week 65-72	1	0	0	37	0	0	0	16
Week 73-80	9	0	48	87	37	21	89	122
Week 81-88	44	12	90	134	47	0	33	80
Week 89-96	22	0	40	96	34	0	28	101
Week 97-104	57	0	49	95	38	0	44	0
Amount	2290			Cost		311175		
Total Cost				402775				

B. Backward Combination

In this model, since the purchase quantity must be an integer, the relationship between the purchase quantity and price should be discrete, but in order to facilitate analysis, it is treated as a function. Preferential policies can reduce the cost compared with the original purchase price.

The main variable related to the preferential purchase plan is the purchase quantity. Obviously, it is necessary to consider "the more you buy at one time, the better it will be". You can try to concentrate the purchase quantity of the neighboring weeks into the same week, whether it will be cheaper. In the context of the application of this study, it is not enough to consider the centralized purchase, but the following points should also be noted:

a) Can only be the back of the weekly purchase volume forward, can not be merged in the future, otherwise it can not meet the demand for use;

b) Due to the advance purchase of more warranty costs should be considered;

c) For the operator, it is also necessary to consider whether the number of coaches is insufficient if the number of students is increased if the operator is purchased in advance.

1) Analysis of source

According to the preferential purchase method, it is obvious that the purchase plan should be selected forward and combined, and the situation of insufficient operator coaches should be considered at the same time. It is analyzed and optimized step by step here.

a) Purchase volume consolidation

Due to the existence of the purchase discount, it is ideal to get the discount through the combined purchase. However, it should be noted that the excess quantity purchased several weeks in advance will lead to the maintenance cost of this part in the waiting process. Therefore, the question of whether it is cost-effective to consolidate purchases forward needs to be considered.

Let the purchase quantity of one week (No. N) be a, the purchase quantity of the following week (No. N+s) be b, the interval between two weeks be s, and the combined quantity be i. The value range of the combined quantity is 0 (no combined purchase) to b (all combined purchase). It is necessary to consider the increased warranty cost due to the part purchased in advance after the forward consolidation purchase. The calculation method of the cost is as follows:

$$R_{pay} = f(a+i) + f(b-i) + 10is \tag{18}$$

$$C_{pay} = g(a+i) + g(b-i) + 5is \tag{19}$$

Since the function is segmented and there are many independent variables, it is difficult to find out the law directly from the mathematical point of view. Therefore, it is considered to test by simulation. In the selected two weeks of respective purchase volume, simulate the results of different merger volumes under different intervals between two weeks. The original purchase parameters of the two weeks involved in the calculation are shown in Table 4.

TABLE IV. ORIGINAL PURCHASE PARAMETER SETTINGS

The number of container boats									
Week. N	2			7			15		
Week. N+s	2	7	15	2	7	15	2	7	15
The number of operator hands									
Week. N	10			30			50		
Week. N+s	10	30	50	10	30	50	10	30	50

The test results are presented in Figure 7, simulating the combined purchase of container boats. Each small icon is the comprehensive cost from no combined purchase to all combined purchase (with increasing abscissa) under a certain value combination of A and b, simulating different intervals (differentiated by different color lines in the same graph). Figure 8 shows the operator's combined purchase simulated in the same way.

From the trend analysis in the figure, it can be found that the Container Boats and the operator have the same change rules. With the change of the combined volume, when the purchase volume is small for two weeks, the required cost of the combined purchase will increase, as shown in FIG. 8 (a) and 9 (a). However, when the purchase

quantity is relatively large, the cost trend of combined purchase increases first and then decreases. In some cases, the cost of all purchase is even less than that of non-combined purchase, as shown in Figure. 8 (i) and Figure. 9 (i).

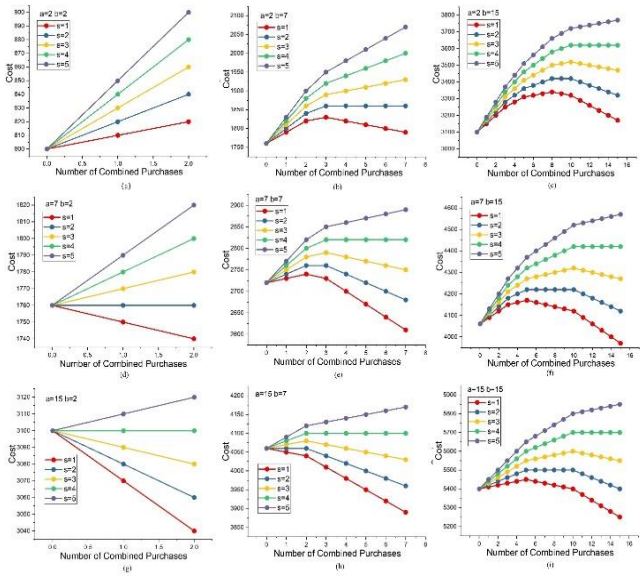


Figure. 7. Simulation of the combined purchase cost of Container Boats

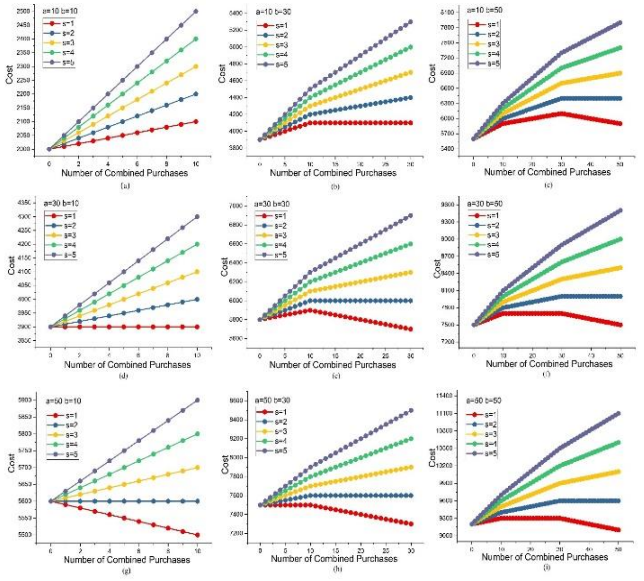


Figure. 8. Simulation of the combined purchase cost of Operator Hands

Then, when the minimum value needs to be selected, only one point between the starting point (no merge) and the end point (all merge) can be selected. Then only the two points need to be compared, without considering the middle case. A very important conclusion can be reached, that is, without considering the specific amount of consolidation, you only need to compare the case of no merger and the case of all merger, and choose which one costs less.

The operator calculates the cost without merging and the cost algorithm of all merging as follows:

$$C_{pay1} = g(C_a) + g(C_b) \tag{20}$$

$$C_{pay2} = g(C_a + C_b) + 5C_a s \tag{21}$$

The cost algorithm for calculating the cost without merging and the cost algorithm for all merging is as follows:

$$R_{pay1} = f(R_a) + f(R_b) \tag{22}$$

$$R_{pay2} = f(R_a + R_b) + 10R_a s \tag{23}$$

Since the merge is from the back to the front, in the modeling, the reverse order is adopted to determine whether the merge is necessary, and the merging may occur after a certain week and then forward.

b) Trainer number requirements and limitations of the operator

For container boats, the change in purchase volume does not affect the actual regular operation except for the change in the regular array. It is only necessary to update the purchase sequence by adding the judgment function. However, whether the combined purchase of operators can provide enough operators needs to be considered. If there are not enough free operators, then the maximum number of trainee operators that can actually be merged is the maximum number of trainee operators that the free operators can support. Compare the cost of merging with the cost of not merging and choose the cheaper option.

Here, there are two options for the operator's operation: one operation is completed when the cost is combined and the number of coaches is counted at the same time; Performing the merge operation first, and then considering the fallback if there are not enough coaches, will increase the

amount of computation. The second algorithm is used in this study.

The operation is performed between three weeks, and the week numbers are named l, m, and n in increasing order. According to algorithm 2, it is found that the merger is more cost-effective after performing the merger calculation from week n to week m, and then performing the merger calculation with week l, it is found that the merger can be performed again, and because there are enough free operators in week l, it is enough to meet the demand of purchase quantity. According to algorithm 1, when the number of coaches is moving from n to m weeks, there is a situation that the number of coaches is insufficient to backtrack, so the number of merges cannot reach the optimal level or even do not merge, but in fact, merging again to l weeks can solve this problem. So choose the second algorithm. Algorithm 1 is also logically feasible, but requires multiple iterations until the purchase sequence is no longer changing.

After performing the merge operation, perform the hand part rollback. Since not enough coaches can be provided, at least a certain number of coaches can be rolled back so that the number of coaches is just enough, or more can be rolled back. Combined with the analysis of the simulation results of the merging module, the forward calculation of the minimum rollback value (the more the rollback, the smaller the value of the abscissa in Figure. 10, 11, 12 and 13) shows that the value of the minimum cost in the rollback process must be a point between the starting state (minimum rollback) and the end state (all rollback). Therefore, You only need to consider which is less expensive between a minimum rollback (coach saturation) and a maximum rollback (total rollback).

When the number of free operators in a certain week is negative, the data of the current purchase number and the number of coaches in that week can be read at the same time to calculate the maximum number of coaches that can actually provide training:

$$coach\_total = coach + free \tag{24}$$

Each coach can undertake a maximum of 20 trainees, so the maximum number of purchases that can be undertaken is:

$$buy_{max} = 20coach\_total \tag{25}$$

The minimum number of purchases that need to be rolled back is:

$$buy_{back} = buy_{new} - buy_{max} \tag{26}$$

Read the original purchase number of the week in the original purchase data backup, and the original combined total is:

$$buy_2 = buy_{new} - buy_1 \tag{27}$$

Look for the weekly number of the nearest purchase, which is the weekly number of buy_2, and record the difference between the two weekly numbers for subsequent calculation. Note that you can't directly read the number of purchases from the source of the merge in the original array, because you can't rule out sequential merges near the source of the merge, but the rollback is gradual.

From this, we can calculate the cost of the total rollback $pay_{back-all}$ (no consolidation in the original case) and the cost of the minimum rollback value $pay_{back-\min}$ (the original case, keep buy_{back}, only transfer $buy_2 - buy_{back}$) :

$$pay_{back-all} = g(buy_1) + g(buy_2) \tag{28}$$

$$pay_{back-\min} = g(buy_{\max}) + g(buy_{back}) + 5s(buy_2 - buy_{back}) \tag{29}$$

Compare the two data and choose a rollback method that costs less.

2) Optimization Mechanism

a) Merge modules

To facilitate subsequent analysis and comparison, copy and reserve the original data before performing this operation.

According to the analysis, the module scans the purchase sequence in reverse order, and the specific process is shown in Figure 9.

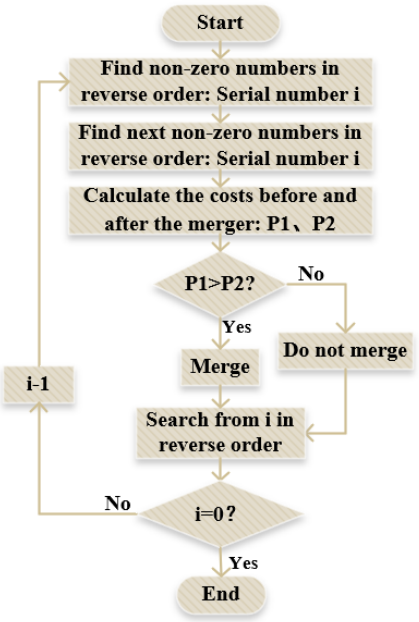


Figure. 9. Flow chart of merge purchase algorithm

Scan the purchase quantity sequence in reverse order scanning, find the first non-zero purchase quantity (let the ordinal number read is i), then look forward to the most recent non-zero purchase quantity (let the ordinal number read is j), and judge the cost before and after the merger. If the cost after the merger is higher and unchanged or higher, the merger will not be performed. If the cost is reduced after the merger, all the purchases in week i will be added up to week j, and the purchases in week i will be zero. Then continue scanning forward from the i-1 ordinal and repeat the above procedure.

After the merge part is complete, the purchase sequence has been updated, and updates of the other arrays have been performed since the first week.

b) Return module of the operator

Since the combination module does not consider the limitation of the number of coaches, all or part of the data needs to be rolled back if the number of free operators is negative after the update. The operation process is shown in Figure 10.

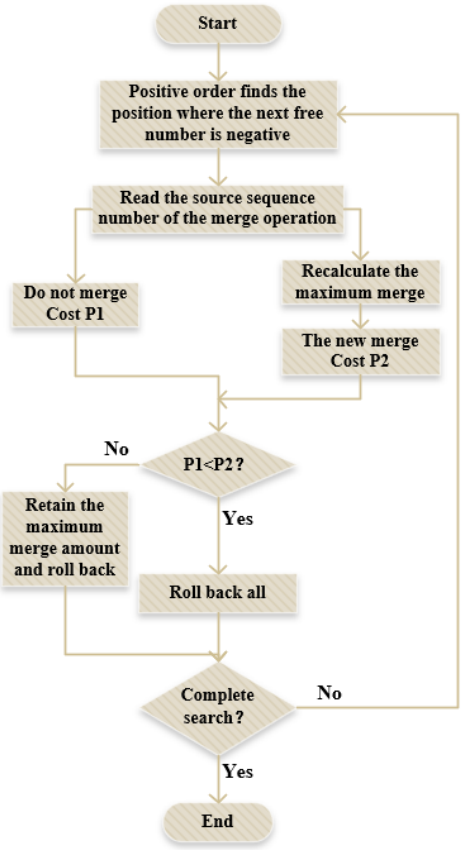


Figure. 10. Flowchart of the rollback algorithm

The positive sequence scans the array module that records the number of free operators every week, finds the first number less than zero (let the read ordinal number be i), scans the original purchase data array before merging on this basis, and finds the nearest non-zero number (let the read ordinal number be j), which is the ordinal number of merging the purchase number to i. The corresponding data is retrieved, and the cost is calculated in two ways: full rollback and only excess rollback, and the lower cost is selected to update the number of purchases. Since this operation must result in a change in the number of purchases, all the array data corresponding to the two ordinals must be updated with each rollback.

After the purchase sequence is updated, the other arrays are updated according to the rules from the first week.

3) Effect of Optimization

The running results after loading the reverse merge module are shown in Table 5.

TABLE V. THE OUTPUT OF THE SECOND OPTIMIZATION

The output of the second optimization								
The purchase program of Container boats								
Week 1-8	0	0	0	5	0	0	0	1
Week 9-16	0	0	0	2	0	0	0	0
Week 17-24	0	0	2	19	0	0	0	7
Week 25-32	0	0	0	0	0	0	0	0
Week 33-40	0	0	0	0	0	0	0	4
Week 41-48	0	0	0	5	0	0	5	12
Week 49-56	0	2	15	18	0	4	8	13
Week 57-64	0	0	1	12	0	0	6	11
Week 65-72	0	0	2	11	0	0	0	10
Week 73-80	0	1	11	16	0	10	16	21
Week 81-88	0	9	18	24	0	6	11	19
Week 89-96	0	7	10	24	0	4	10	25
Week 97-104	1	8	14	20	1	7	16	0
Amount		484		Cost		90140		
The purchase program of operator hands								
Week 1-8	14	0	0	36	0	0	0	1
Week 9-16	0	0	0	8	0	0	0	0
Week 17-24	0	0	13	96	41	0	0	15
Week 25-32	0	0	0	0	0	0	0	0
Week 33-40	0	0	0	0	0	0	0	0
Week 41-48	0	0	0	0	0	0	18	66
Week 49-56	25	1	70	114	48	1	36	62
Week 57-64	0	0	0	37	13	0	9	55
Week 65-72	1	0	0	37	0	0	0	16
Week 73-80	9	0	48	87	37	21	89	122
Week 81-88	44	12	90	134	47	0	33	80
Week 89-96	22	0	40	96	34	0	28	101
Week 97-104	57	0	49	95	38	0	44	0
Amount		2290		Cost		310525		
Total Cost				400665				

The simple greedy algorithm, the first optimization, the second optimization, and the three operation effects were compared and verified, mainly comparing the operator's purchase sequence and operation and maintenance cost, the container boat's purchase sequence and operation and maintenance cost, and the overall operation and maintenance total cost. The results are shown in Table 6. For the convenience of comparison, set the saving rate(SR):

$$\eta = \frac{Sum_{Before} - Sum_{After}}{Sum_{Before}} \times 100\% \quad (30)$$

TABLE VI. COMPARISON OF THE EFFECT BEFORE AND AFTER OPTIMIZATION

	OP-N	OP-C	CB-N	CB-C	TOTAL
Simple	2290	311505	484	91600	403105
Optimization 1	2290	311175	484	91600	402775
Optimization 2	2290	310525	484	90140	400665
SR total	0.6053%	SR 1	0.08187%	SR 2	0.5239%

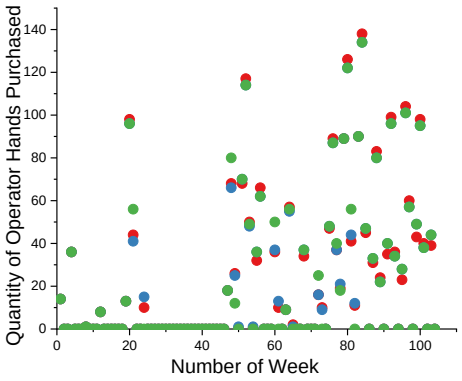


Figure. 11. The purchase of Operator Hands before and after optimization

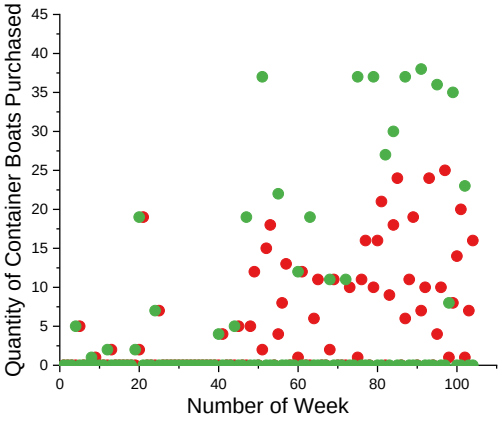


Figure. 12. The purchase of Container Boats before and after optimization

Through data comparison, it is found that the two optimizations are effective in saving operation cost requirements. Figure 11 shows the situation of the operator buying the scheme in the two optimizations, where red represents before optimization, blue represents after the first optimization, and green represents after the second optimization. Figure 12 shows the purchase plan of container boats before optimization and after overall optimization. The container boats were not involved in the first optimization, so the purchase plan of container boats after the first optimization is not presented.

C. Summary of optimization

Considering the minimum purchase amount, the reverse distribution mechanism is established to move the demand gap forward to reduce unnecessary purchase amount. Although the total purchase amount does not change under some special demand sequences, the maintenance waste

caused by "buying too early" is reduced. This module can achieve a saving rate of 0.08187%.

As for the preferential purchase scheme, the analysis found that "purchase in advance" could save more costs in some cases, so the reverse merger mechanism was designed to analyze and measure the costs of "merge" and "no merge" operation and choose a more favorable way. Among them, direct merging may lead to the operator "insufficient number of coaches", so a rollback module is designed. In the "coach oversaturation" section, the cost of the minimum rollback and the total rollback was calculated and the best choice was made. Compared with before loading this module, the saving rate of 0.5239% can be obtained. The two optimization modules work together to achieve a saving rate of 0.6053%.

Through the comparison of the chart, it can be found that the optimization module does have an adjustment effect on the purchase plan. Although the total number of purchases has not changed, the maintenance cost has been reduced due to the adjustment of the purchase plan. In other words, the purchase amount in the early stage has been adjusted to the purchase in the later stage, but the weekly work demand has not been affected. It should be noted that the total number of purchases does not change, which is also related to the demand sequence of the application scenario. In the subsequent simulation tests, the total number of purchases changes.

V. Extended Simulation Test

In order to more strongly verify the effect of the reverse correction mechanism, more array validation tests are performed in the original application context. Two test modes are planned: modifying application parameters and changing the requirements matrix.

A. Modifying Application Parameters

Under the application background of the study in this paper, although the addition of correction mechanism has a certain optimization effect, due to the limitations of environmental parameters, its effect is not obvious, so it is considered to modify different parameter ratios for further testing.

1) Change loss parameter

According to the application environment, there are many parameters besides the demand matrix, among which the main control parameters are more, and the appropriate change test should be done in combination with the actual application environment, so the loss related change test should be considered. In this application environment, the loss-related parameters were selected to change "damage rate(DR)" and "maximum number of instructions of operators (MN-O)", ten groups of test schemes with variable parameters were set up and the total operation and maintenance cost was tested, as shown in Table 7, Figure 13 and 14.

TABLE VII. Test results of loss parameter changes

DR	MN-O	Simple	Optimization	SR/%
0%	5	221290	217430	1.7443
	10	219205	216605	1.1861
	15	218565	216225	1.0706
	20	218260	216070	1.0034
	25	218000	215970	0.9312
	30	217895	216025	0.8582
	SR-average/%		1.1323	
10%	5	408780	405120	0.8953
	10	404815	402320	0.6163
	15	403680	401365	0.5735
	20	403105	400665	0.6053
	25	402755	400365	0.5934
	30	402535	399625	0.7229
	SR-average/%		0.6678	
20%	5	618710	611015	1.2437
	10	611970	606150	0.9510
	15	609710	604300	0.8873
	20	608975	603335	0.9261
	25	608230	602780	0.8960
	30	607465	602440	0.8272
	SR-average/%		0.9552	
30%	5	826950	816590	1.2528
	10	817045	810275	0.8286
	15	813980	807800	0.7592
	20	812330	806920	0.6660
	25	811305	805955	0.6594
	30	810660	805260	0.6661
	SR-average/%		0.8054	
40%	5	1030775	1019150	1.1278
	10	1018040	1011485	0.6439
	15	1014430	1008825	0.5525
	20	1012815	1007620	0.5129
	25	1011500	1006420	0.5022
	30	1011060	1005925	0.5079
	SR-average/%		0.6412	
SR-average-total/%			0.8404	

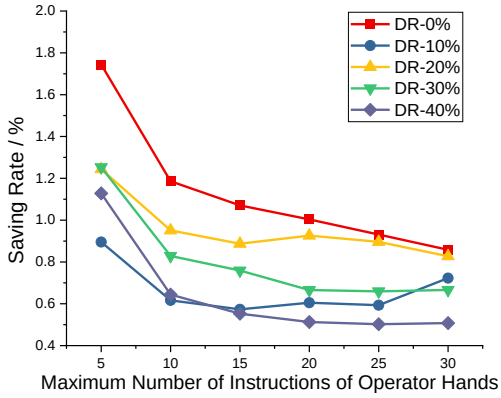


Figure. 13. The relationship between the MN-O and the SR with different DR

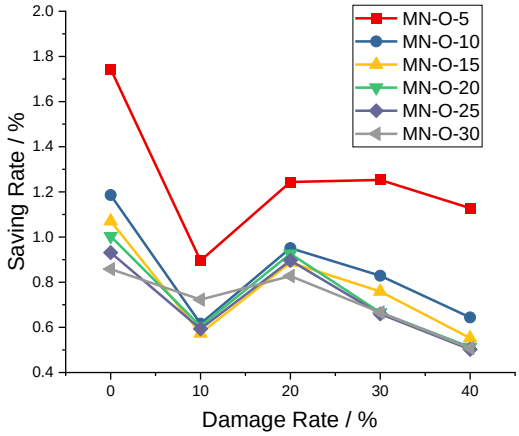


Figure. 14. The relationship between the DR and the SR with different MN-O

The test results show that under all parameter ratios, the total cost saving rate of the test results is greater than zero, indicating that the optimization algorithm is useful, the highest saving rate can be close to 1.8%, the lowest saving rate is more than 0.5%, and the average saving rate of all the test schemes is 0.8404%, and the related images show the same trend, the optimization algorithm effect is stable and has an obvious effect.

2) Change the Discount Purchase Scheme

A large part of the model optimization in this study is based on preferential policies, and combined purchase is also the core decision basis in the model optimization, so it is considered to adjust the optimization policies and then conduct comparative tests. The experimental test Settings and results are shown in Table 8, Figure 15 and 16. Since the Container Boats and operator do not substantially affect each other, they can change simultaneously.

Scheme 1 Original purchase policy

Scheme 2 Add a layer of piecewise function each

Scheme 3 Add two layers of piecewise functions respectively

Scheme 4 Reduce one layer of piecewise function respectively

Scheme 5 Reduce two levels of piecewise functions each (no discount)

Scheme 6 The preferential power of the original piecewise function is reduced

Scheme 7 The preferential power of the original subsection function continues to decrease

Scheme 8 The original segmented function preferential efforts to increase

Scheme 9 We will continue to give more preferential treatment to original piecewise functions

TABLE VIII. TEST OF PARAMETER CHANGE OF PREFERENTIAL SCHEME

Scheme	Container Boats	Operator Hands	Simple	Optimization	SR/%
1	$f(x)=\begin{cases}200x, & x\le 5\\ 100+180x, & 5<x\le 10\\ 300+160x, & x>10\end{cases}$	$g(x)=\begin{cases}100x, & x\le 20\\ 200+90x, & 20<x\le 40\\ 600+80x, & x>40\end{cases}$	403105	400665	0.6053
2	$f(x)=\begin{cases}200x, & x\le 5\\ 100+180x, & 5<x\le 10\\ 300+160x, & 10<x\le 15\\ 600+140x, & x>15\end{cases}$	$g(x)=\begin{cases}100x, & x\le 20\\ 200+90x, & 20<x\le 40\\ 600+80x, & 40<x\le 60\\ 1200+70x, & x>60\end{cases}$	397055	389930	1.7945

3	$f(x)=\begin{cases}200x, & x\le 5\\ 100+180x, & 5<x\le 10\\ 300+160x, & 10<x\le 15\\ 600+140x, & 15<x\le 20\\ 1000+120x & x>20\end{cases}$	$g(x)=\begin{cases}100x, & x\le 20\\ 200+90x, & 20<x\le 40\\ 600+80x, & 40<x\le 60\\ 1200+70x, & 60<x\le 80\\ 2000+60x, & x>80\end{cases}$	394265	379580	3.7247
4	$f(x)=\begin{cases}200x, & x\le 5\\ 100+180x, & 5<x\end{cases}$	$g(x)=\begin{cases}100x, & x\le 20\\ 200+90x, & 20<x\end{cases}$	414225	413355	0.2100
5	$f(x)=200x$	$g(x)=100x$	434295	433615	0.1566
6	$f(x)=\begin{cases}200x, & x\le 5\\ 50+190x, & 5<x\le 10\\ 150+180x, & x>10\end{cases}$	$g(x)=\begin{cases}100x, & x\le 20\\ 100+95x, & 20<x\le 40\\ 300+90x, & x>40\end{cases}$	418700	417880	0.1958
7	$f(x)=\begin{cases}200x, & x\le 5\\ 25+195x, & 5<x\le 10\\ 75+190x, & x>10\end{cases}$	$g(x)=\begin{cases}100x, & x\le 20\\ 60+97x, & 20<x\le 40\\ 140+95x, & x>40\end{cases}$	426193	425578	0.1443
8	$f(x)=\begin{cases}200x, & x\le 5\\ 150+170x, & 5<x\le 10\\ 450+140x, & x>10\end{cases}$	$g(x)=\begin{cases}100x, & x\le 20\\ 400+80x, & 20<x\le 40\\ 1200+60x, & x>40\end{cases}$	376065	365835	2.7203
9	$f(x)=\begin{cases}200x, & x\le 5\\ 200+160x, & 5<x\le 10\\ 600+120x, & x>10\end{cases}$	$g(x)=\begin{cases}100x, & x\le 20\\ 600+70x, & 20<x\le 40\\ 1800+40x, & x>40\end{cases}$	349025	329050	5.7231
SR-average %			1.6972		

According to the experimental scheme, the comparison images shown in Figure. 15 were drawn for groups 1-5, and the comparison images shown in FIG. 16 were drawn for groups 1 and 6-9.

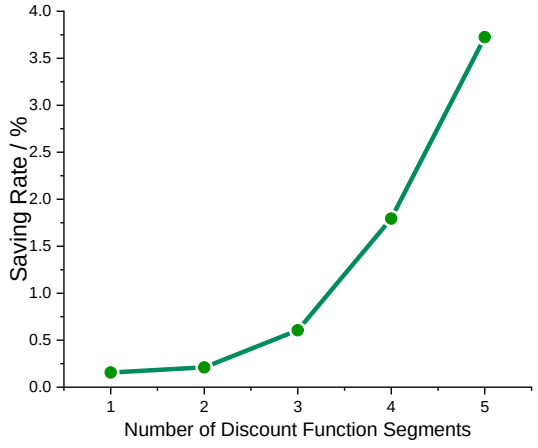


Figure. 15. Number of discount segments and SR

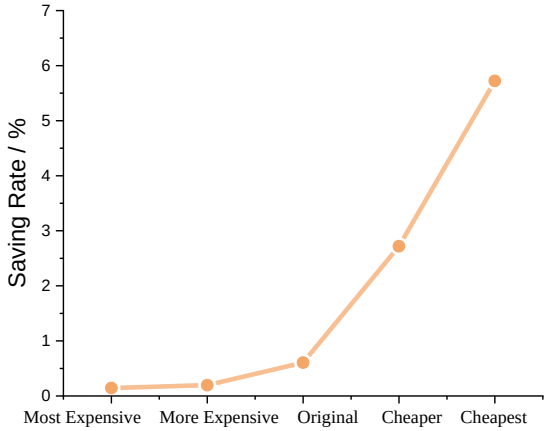


Figure. 16. Relationship between discount intensity and SR

The test results show that the saving rate is greater than zero in all cases, indicating that the optimization scheme is effective. In addition, with the increase of preferential intensity and the increase of preferential segments (also referred to as the increase of preferential intensity), the effect of the optimization scheme is getting better and

better, which is consistent with the optimization strategy "based on the lowest operating cost". At present, the best saving rate is close to 6%, the lowest is 0.14%, and the average saving rate is 1.69%.

B. Changing the sequence of requirements

It is considered to generate three groups of arrays from different sources as the weekly demand function in the application scenario. Due to the limitations of application environment modeling, a new demand data is randomly generated near the weekly demand number according to the Gaussian distribution rule based on the array of the original application scenario, namely:

$$N_{new} = random.gauss(N_{old}, 10) \tag{31}$$

If the generated data is less than 0, it is generated again.

Use this rule to generate 100 test arrays. Then input each group of data into the simple greedy strategy model and the optimized model, and compare the saving rate of each group of input data. The experimental results are shown in Figure 17.

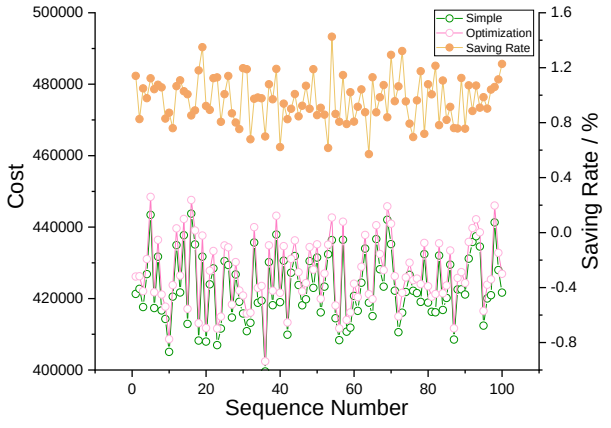


Figure. 17. Test results after changing the requirement matrix

In Figure 17, the red ring represents the operation result before optimization, the green ring represents the operation result after optimization, and the orange origin represents the total cost saving rate of this test. It is found that all test results of the newly generated demand matrix are in line with expectations -- the saving rate is greater than zero, indicating that the optimization strategy is universal in this application scenario. In addition, it can be found that the optimization effect is also affected by the demand matrix. In the above 100 groups of tests, the maximum saving rate is close to 1.6%, the lowest saving rate is less than 0.6%, and the average saving rate is 0.9677%. The overall "optimization" effect has been achieved.

C. Change the parameters and test again

In the experiment, it is found that due to the contingency of the original application scenario, constraint conditions and demand sequence, some optimization effects are not obvious. In the above simulation test, there is a better saving rate. Therefore, parameters with better optimization performance and reasonable performance were selected to modify the application scenario, and the simple model, the primary optimization model and the secondary optimization model were tested again.

The parameters that need to be changed are: the loss rate is 20%, the maximum number of instructions of the coach operator is 10, the preferential policies are selected as Article 9 in Table 5, and the demand sequence is selected as the group with the highest saving rate in the test in the previous section. The results of the run are shown in Table 9

TABLE IX. RETEST RESULTS OF CHANGING GLOBAL PARAMETERS

	OP-N	OP-C	CB-N	CB-C	TOTAL
Simple	4101	383245	939	146280	529525
Optimization 1	4080	381415	939	146280	527695
Optimization 2	4080	368045	939	137530	505575
SR total	4.5229%	SR 1	0.3456%	SR 2	4.1918%

In this group of tests, the "reverse purchase allocation" module expressed in optimization 1 showed obvious optimization effect on operator planning, which not only reduced the maintenance cost, but also clearly reduced the total number, indicating that this optimization module is effective. In addition, the optimization module 2 continues to maintain its original optimization effect and performs well.

Because of the contingency of parameter designation and requirement sequence, the effect of modules and models can not be expressed by a simple data, but should be analyzed concretely. Although the optimization rate (also known as saving rate in this paper) of the module in this application scenario is not large numerically, it can be seen from the variable parameter extension test and simulation test that the operation rate of the optimization module is 100 percent, that is, all the tests have optimization effect.

In the same application scenario, some researchers also choose linear programming model, overall programming model, ARIMA model, unitary linear regression model, etc. In contrast, the model proposed in this study does not focus on the expression of the overall function, but runs according to the time block, focuses on mining the influence relationship between variables, adopts the idea of greedy strategy to gradually select the optimal solution, and finally completes the model optimization module by gradually considering the "optimal condition" in the reverse optimization. The final operation results show that the model proposed in this study can achieve approximate or even better operation effect, and the model analysis and parameter variable analysis are more clear and concise, the model segmentation is clear, the action mechanism is clear, and the stability and optimization effect of the optimization module are excellent.

VI. CONCLUSIONS

In view of the application background tested in this paper, the superficial simple application of greedy strategy can indeed find very big drawbacks, especially under the influence of local optimal solution and after-effect, the so-called "optimal solution" produced in the early stage is often affected and changed. By analyzing the source, generation mechanism and solution direction of the problem and adding the "reverse correction" mechanism with "greedy strategy" as the core principle, it can meet the requirements of the application environment and optimize the results without destroying the original modeling logic. Some parameters of the application background were changed and tested regularly, and 100 groups of test sequences were randomly generated according to the original requirement sequence rules. Parameters and sequences with more obvious performance were selected to test the effect of the module again, and the optimization module had a better performance.

1) According to the requirements of application scenarios, the optimization strategy of this study proposes two optimization modules, reverse allocation and reverse merger. The saving rate of each module can be 0.08187% and 0.5239% respectively, and the comprehensive saving rate is 0.6053%. In the application scenario of this study,

the operating cost has reached hundreds of thousands or even millions, and the actual cost saved has reached tens of thousands.

2) The damage rate and the maximum number of instructions of the operator instructor were changed, and the highest saving rate was close to 1.8%, the lowest saving rate was more than 0.5%, and the average saving rate was 0.8404%.

3) When the parameters are changed regularly, the saving rate of the optimization module can also show a certain regular change, which shows that the optimization module has stability and scalability. It is not a simple numerical optimization, but an effective optimization from the mechanism of action.

4) Changing the purchase discount program for several times, the test results perfectly meet the optimization goal. The greater the preferential intensity of the purchase plan, the higher the optimized saving rate, the highest is close to 6%, the lowest is 0.14%, and the average saving rate is 1.69%.

5) Based on the original demand sequence, 100 groups of new demand sequence were randomly generated and tested using Gaussian distribution. Among them, the optimization efficiency is 100%, all the sequences can have the optimization effect, the best can be close to 1.6%, the average saving rate is 0.9677%, the optimization module has good effect and strong stability.

6) Select a set of parameters and demand sequences with good performance in the simulation test and in line with the actual situation, and test the optimization module effect again. The optimization module has obvious effect, the overall saving rate is more than 4.5%, and the cost is more than 20,000 yuan.

The application background of this research is a dynamic programming, multi-constraint and multi-objective topic. Its internal equation of state, state transition function and objective function are not a simple expression, with many variables and complex interconnections. The local optimal solution obtained by greedy strategy can directly influence each other, and the optimal solution cannot be obtained directly. However, this does not mean that greedy strategy cannot be used. In fact, in the solving idea of greedy strategy and dynamic programming, "decompose the problem and solve it step by step" is very conducive to analyzing the problem. Based on the preliminary greedy strategy model and through the analysis of the relationship between internal variables, the "reverse correction mechanism" proposed in this study can repair the shortcomings of the greedy algorithm and save tens of thousands of yuan of operating costs in the application environment of this study.

Compared with other modeling strategies in the same application scenario, the "compound optimization greedy strategy integrating reverse correction mechanism" proposed in this paper can not only obtain better operation results, but also dig deep into modeling analysis, module programming solutions and optimization strategies. It focuses more on the in-depth analysis of the mechanism of action and the analysis of the relationship between variables, so as to solve practical application problems by breaking the whole into pieces.

Greedy strategy is easy to be ignored because its local optimal solution does not mean the global optimal solution. However, the thinking of greedy strategy of "breaking the whole thing into pieces and breaking it one by one" is very conducive to analyzing the relationship between variables and

sorting out clear logical lines in a large number of variables and constraints. Greedy strategy should have more space to play, especially from the perspective of mathematical analysis, through the reverse correction of greedy strategy algorithm, repeated calls, comparative selection, etc., can get obvious and stable optimization effect.

In today's research environment where various kinds of high-order algorithm strategies and artificial intelligence machine learning strategies are blooming everywhere, sometimes simple problems are easy to be complicated. Traditional algorithms should not be abandoned directly, but can be modified or integrated with emerging technology algorithms to get more possibilities and give play to greater potential.

REFERENCES

[1] W.Shi, Application of data mining technology in biological sequence analysis and recognition algorithm [J]. Computer Programming Skills & Maintenance, 2020(07):77-79

[2] J.Yu. Application analysis of Python language in big data analysis [J]. China New Telecommunications, 2021, 23(21):78-79.

[3] S.Ou. A brief analysis of the application techniques of Python data analysis [J]. Youth of Shanxi, 2021(10):67-68.

[4] Z.Liu, J.H.Guo. Explore the application of Jupyter Notebook programming tools in project-based learning [J]. Chinese modern educational equipment, 2020(14):3.

[5] Y.Y.Wei, Y.F.Wu, Y.Y.Li. Research on the application of Python technology in data visualization [J]. Fujian Computer, 2022, 38(01):27-31.

Real-time Satellite Anomaly Data Tagging Based on DAE-LSTM

Caiyuan Xia
School of Computer Science & Engineering
Xi'an Technological University
Xi'an, 710032, China
E-mail: 2969714860@qq.com

Qianshi Yan
School of Computer Science & Engineering
Xi'an Technological University
Xi'an, 710032, China

***Abstract*—Spacecraft is the main carrier of human exploration of outer space, exploration and understanding of the Earth and the universe, and the development of spaceflight can promote human civilization andsocial development, and can meet the nee-ds of economic construction, scientific and technological development, security construction, social progress and other aspects. The current global number of satellites in orbit reaches 5,465, of which China has 541. The vigorous development of the space industry symbolizes the steady improvement of the country's comprehensive national power and overall technology. During the operation, the satellite in orbit needs to transmit data to the ground, these data may be subject to interference from various aspects, or even equipment failure, we find these data in real time is very important to reduce losses. The data transmitted by satellite has obvious temporal characteristics, and Long Short-Term Memory (LSTM) network has obvious advantages for processing temporal data, so this paper proposes a BER marking model based on the combination of LSTM network and self-coding technology. By comparing the data before and after noise reduction, a threshold value can be determined, and the BERs can be accurately distinguished by this method. After testing with real satellite temperature data, the accuracy of the model detection reaches a high level.**

Keywords-LSTM Network; Self Coding Technology; Satellite Data; Deep Learning

I. INTRODUCTION

Spacecraft are the main carrier for human beings to explore outer space and explore the understanding of the Earth and the universe. The development of spaceflight can promote human civilization and the development of society, and can meet the needs of economic construction, scientific and technological development, security construction, social progress and other aspects. According to UCS statistics, as of April 30, 2022, the current number of satellites in orbit around the world reached 5465, of which China has 541. The booming development of the space industry symbolizes the steady improvement of the country's comprehensive national power and overall technology.

The satellite will send some collected data or its own parameters to the ground in the normal operation process. The ground receiving station decodes and demodulates the received data and then transmits it to the user side, which then further processes the received data according to the demand. The data received on the ground may be affected by objective factors such as

electromagnetic interference, multipath effect and equipment failure, resulting in some false code points in the data. Very reliable telemetry data are needed for analyzing the operation status of spacecraft in orbit and for various application areas. The study of spacecraft anomaly detection technology can not only detect potential satellite anomalies in time, but also monitor satellite performance, detect satellite faults, locate the root cause of satellite faults, improve the safety of satellite operation in orbit, improve satellite operational life, and improve satellite work efficiency. Telemetry data fault diagnosis and abnormality detection technology has also become an important direction of research in the field of aerospace. The following characteristics exist for its parameter data: the boundary between data in anomalous mode and normal data is blurred and difficult to distinguish; the trend and period of normal data changes continuously as the time and position of the spacecraft in orbit shift; the definition of data anomaly is different in different scenarios and environments. Thus, it can be seen that it is challenging to be able to accurately and effectively represent the possibility of telemetry data as anomalous data.

In-orbit satellites will collect a data every time, which has obvious time-series and real-time. This paper adopts a network structure combining LSTM and self-encoder to mark data anomalies in real time for the purpose of real-time detection of anomalous data.

II. RELATED JOBS

The traditional methods for detecting anomalies in temporal data have been a common research topic, and the traditional methods for detecting anomalies in temporal data are detected by the proximity between the data, and when applying such methods, it is necessary to choose the appropriate proximity. If the distance is too large, it means that the similarity between them is low [1] and can be judged as anomalous. There are also density-based detection methods, the principle is to use the number of anomalous points is much less than the number of normal points, and the distribution of normal data has a certain similarity, so the data of low-density neighborhood is judged as anomalous data [2,3]. Deep learning-based data anomaly detection methods have also been widely proposed in recent years, and literature [4] used a combination of stacked self-coding neural networks (SAE) and long short-term memory neural networks (LSTM) for anomaly detection of process data in industry. The literature [5][6] used noise reduction self-encoder (SDAE) for anomaly detection, but did not consider the problem of real-time detection. The literature [7] uses Markov model, and the data dimension processing in the pre-model stage uses principal component analysis to downscale and human experience to extract features, this method does not make good use of the characteristics of the studied data itself, which leads to limited anomaly detection effect.

The detection method based on manual monitoring combined with thresholds is the most traditional and widely used method in practice, but the occurrence of many anomalies in telemetry data does not cause the value to exceed the limit, and its detection method has disadvantages such as high labor cost and poor scalability of thresholds. The application of anomaly detection methods based on expert systems is considered a more widespread and reliable method, applying expert knowledge (expressed in the form of some rules) to determine anomalies through judgment. However, due to the complex space environment of spacecraft, it is difficult for expert systems to establish perfect and comprehensive rules, and often unknown anomalies occur that cannot be detected by expert systems. Data-driven anomaly

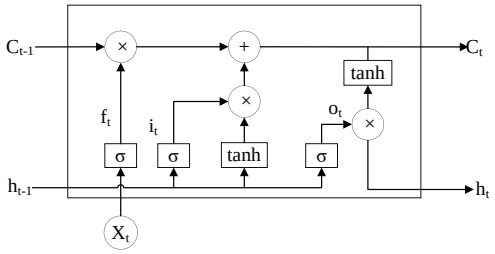
detection method based on machine learning related theory is used to model time series data and identify anomalous patterns that do not conform to normal data through the model output. It is the most intelligent and is currently the most popular method for research on data anomaly detection of spacecraft telemetry data, which is characterized by high data volume and high dimensionality, etc. Since it is targeted at analyzing the change pattern of the data itself and does not rely on expert experience, relatively few of them are currently implemented in ground monitoring systems.

RNN (Recurrent Neural Network) is the most typical model for processing time-series data, and LSTM (Long Short-term Memory), as a variant of RNN, can well solve the gradient disappearance and explosion problems that exist in the training process of RNN. Auto-Encoder (AE) is a kind of neural network that learns the effective representation of sample data by unsupervised learning, and can be used for noise reduction of data. By comparing the data before and after noise reduction of telemetry data, it is easy to find the false code points. In this regard, we construct a network model by combining the two to solve the problem of real-time detection of satellite data outliers.

III. LSTM SELF-CODING THEORY

A. LSTM network

Long and short memory neural network, a special kind of recurrent neural network (RNN), has the problem of gradient disappearance or gradient explosion as time increases in the original recurrent neural network, so it cannot handle data of long sequences. To address the shortcomings of RNN, an improved LSTM model is proposed. Figure 1 shows a single LSTM unit.



Figer 1. LSTM unit

In a single LSTM cell, when the input sequence is x_t :

$$f_t = \sigma(W_{hf} h_{t-1} + W_{xf} x_t + b_f) \quad (1)$$

$$i_t = \sigma(W_{hi} h_{t-1} + W_{xi} x_t + b_i) \quad (2)$$

$$o_t = \sigma(W_{ho} h_{t-1} + W_{xo} x_t + b_o) \quad (3)$$

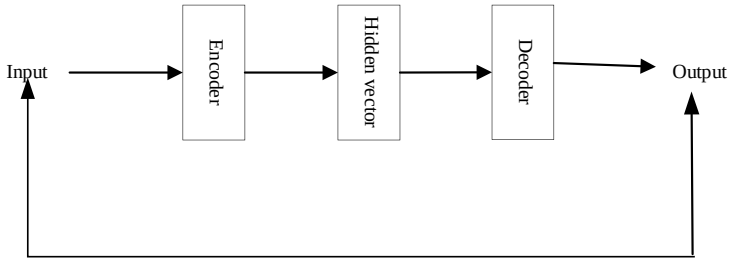
$$C_t = f_t \odot C_{t-1} + i_t \odot tanh(W_{hc} h_{t-1} + W_{xc} x_t + b_c) \quad (4)$$

$$h_t = o_t \odot tanh(C_t) \quad (5)$$

The above equation represents the forgetting gate (ft), the input gate (it) and the output of the output gate (ot), respectively, represents the cell state of the current LSTM cell, which will also be used as the input of the next cell, is the output of the current LSTM cell also as the next input. The formula represents the activation function generally uses the sigmoid activation function, W denotes the weight matrix between the forgetting gate and the previous layer, and "⊙" denotes the product operation of the elements. From the formula (5), we can see that the connection between the current cell state and the previous cell state is controlled by the forgetting gate, and if the forgetting gate is 0, it means that the previous data is cleared and will not affect the following operation anymore, which also avoids the problem of gradient disappearance or explosion.

B. Self-encoder principle

Self-encoder is a class of artificial neural networks used in semi-supervised and unsupervised learning, whose main function is to extract features and reconstruct data. A complete self-encoding network includes two parts: encoder and decoder, the role of the encoder is to map the input vector into a kind of hidden vector, and the decoder reconstructs the vector into the original input, the network structure is shown in Figure 2. In this paper, a noise reduction self-encoder (DAE) is used, which adds random error points to the input data and learns through training to recover the pure data as much as possible.



Figer 2. Self-coding network structure

Equation 6 is the encoding process from the input layer to the hidden layer, where x represents the input, h represents the hidden layer vector, and b is the paranoid term.

$$h = \sigma(W_1 x + b_1) \tag{6}$$

Decoding process from the hidden layer to the output layer:

$$Output = \sigma(W_2 h + b_2) \tag{7}$$

The corresponding mean square error loss function is：

$$Loss = \frac{1}{N}(Output - \mathrm{x})^2 \tag{8}$$

Through iteration after iteration, the training of the self-encoder is completed when the error stabilizes and the input and output are as similar as possible.

IV. REAL-TIME SATELLITE ANOMALY DATA TAGGING BASED ON DAE-LSTM

A. Pre-processing of data

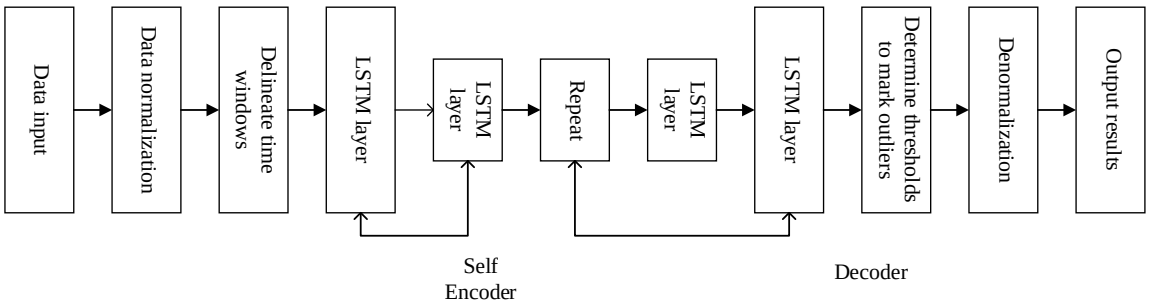
The data set in the experiment is from a model of satellite taken to the temperature data, in the data obtained, in order to save storage space near the same data was omitted, and in the use of the omitted data need to be restored, the data sampling interval is known 8.192s, so every 8.192s if there is time data missing that copy the previous data as the data value of the moment. The input requirement of the LSTM network is temporal data, so the data needs to be partitioned into multiple sets of time windows, and the length of the time windows also affects the accuracy of the model, for which time windows of different lengths are tested. In order to speed up the convergence of the network, the data are normalized before training, and for comparison during detection, we perform the inverse normalization process after detection, so that it still returns to the range before detection. The

normalization is done using the Z-Score normalization method $Z=\frac{x-\bar{x}}{\sigma}$, Retain the variance before normalization σ and the mean value $\bar{x}$ and other data, It is used to denormalize the data after detection, and we can use $x=Z*\sigma+\bar{x}$ to recover the data after denormalization. As the data we got is better quality data, there is no error code point in the data, we need to add error code point manually, also for the normal operation of the evaluation mechanism we save the position of the added error code point, in order to simulate the error code point of the satellite data as much as possible, we add the error code position is random, the number of error code point is set to 1/1000 of the total number of data with reference to the actual situation, considering the short time interval of the data collection of this type of on-board sensor, there may be continuous error code situation, we choose 1/5 of the error code point set as continuous error code point.

B. Experimental process analysis

1) Construction of the network model

By deepening the number of network layers of the self-encoder, more interference information can be suppressed and more effective data feature representation can be obtained. The encoder part of the experiment is a three-layer network structure, and the decoder part is also a three-layer network structure, which is built by using the keras. layers. LSTM network structure in TensorFlow, and the sequential approach is used to build the LSTM network structure in TensorFlow, using sequential approach to build the network model, the number of encoder and decoder is 64, 128, 128, 64, 128, 128. In order to prevent overfitting, a Dropout layer is added to the network model to randomly deactivate some neurons, and the Dropout ratio is set to 0.3. tanh is used as the activation function, mean square error (mse) is used as the loss function, and Adam optimizer is used to adjust the network to reach the optimum, the number of batch training samples is 32, and the number of experimental training rounds (The number of batch training samples is 32, and the number of training rounds (Epoch) is set to 15 for the best results. The overall structure of the program is shown in Figure 3.

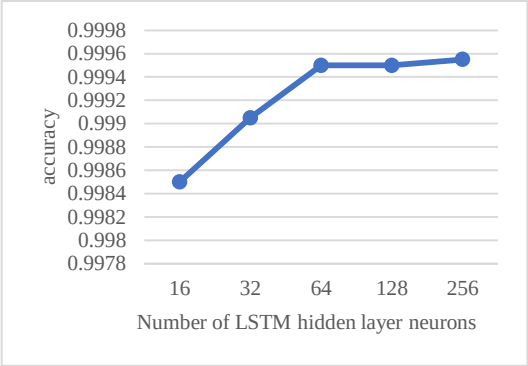


Figer 3. Program structure diagram

2) Number of neurons in the hidden layer

In the physical architecture of the LSTM, an LSTM cell contains four gate structures, and the implementation of each gate structure relies on multiple hidden function layer neurons, which are equivalent to four hidden layers, each of which is fully connected to the input, and the input is composed of two parts, one is the output of the LSTM cell at the previous moment, and another part is the input of the current sample vector, so

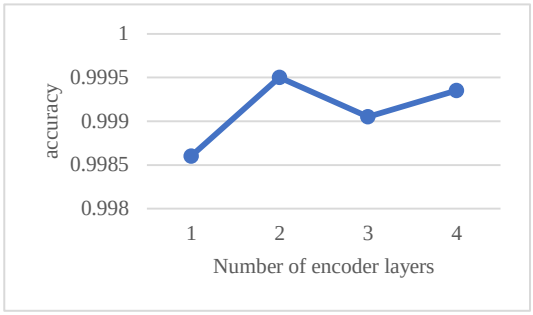
that by the After the internal computation of the LSTM unit, the final output of the current moment is obtained as part of the input of the LSTM unit in the next moment. The number of neurons in the hidden layer also has an important impact on the effect of the model. The more the number of neurons, the more complex the network and the higher the accuracy, but it also brings the problem of more computation. From the figure, we can see that the accuracy of the model is better when the number of neurons is 64 or more, and there is no need to increase the number of neurons to avoid the computational overload.



Figer 4. Effect of the number of neurons on the accuracy of the model

3) Determination of the number of network layers

The number of network layers of the self-encoder largely affects the effect of the model. The deeper the layers, the more theoretically the ability to fit the function is enhanced and the effect will be better, but in practice deeper layers may bring about overfitting problems and also increase the training difficulty, making it difficult to converge the model. Shallow neural networks do not have a high degree of feature abstraction, for which we need to determine the optimal network depth. As shown in Figure 5, the horizontal coordinates indicate the number of network layers for the encoder and the same number of layers for the decoder.



Figer 5. Effect of the number of network layers on the accuracy of the model

4) Model evaluation mechanism

Regarding how to measure the accuracy in the experiment, we used two metrics to measure it, which are the ratio of the number of abnormal data correctly classified as abnormal to the number of detected abnormal data (acc_n) and the ratio of the number of normal data and abnormal data both correctly classified to the total number of detections (acc). The details are calculated as follows.

$$TP = set(x\_l)\ \&\ set(x) \tag{9}$$

$$TN = \left(set(testx) - set(x\_l)\right)\ \&\ \left(set(testx) - set(x)\right) \tag{10}$$

$$acc = \left(len(TP) + len(TN)\right)\ /\ len(testx) \tag{11}$$

$$acc\_n\ =\ len(TP)\ /\ len(x\_l) \tag{12}$$

TP in equation (9) represents the anomalies correctly classified as abnormal, set(x_l) represents the anomalies added to the data during preprocessing, and we read them again during testing, set(x) represents the anomalies found by the model, and the two intersect to get the anomalies correctly found by our model.

TN in equation (10) represents the number of normal data correctly classified as normal, and set (testx) represents the test set data.

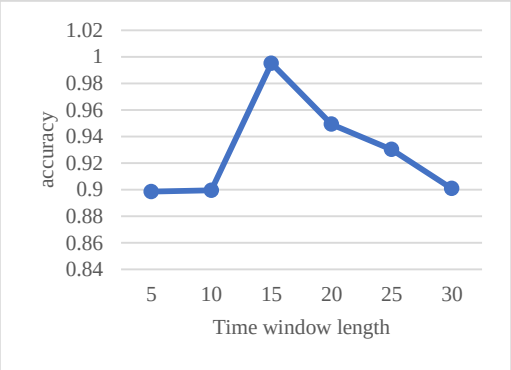
The acc in equation. (11) represents the accuracy of the model in correctly classifying abnormal data and normal data.

The acc_n in equation (12) represents the accuracy of the model in correctly classifying the anomalies as abnormal.

The accuracy rates used in the following experiments are all calculated by acc， In addition to measuring by accuracy, we also need to determine whether the network reaches optimality in the training process through the loss function, which determines the number of iterations. We choose the mse (mean squared error) loss function, which uses the difference between the data before and after training squared before taking the mean value.

5) Construction of time series

LSTM networks require the input data to be time-series data, given a set of multivariate time series of length n $X = \{X_1, X_2, ..., X_n\}$, where the observations at the moment t($1 \leq t \leq n$) are m-dimensional vectors $X_t = [x_1, x_2, ..., x_m], X$ It records the system state information at this time. The goal of time series data anomaly detection is to determine if there is an anomaly in the time series, and if so, to locate the anomaly. In order to improve the robustness of the detection model, a sliding time window of length k is used to segment the temporal data, and then obtain the set of subsequences $W = \{W_1, W_2, \ldots, W_n\}$, where the subsequence $W_t = \{X_{t-k+1}, X_{t-k+2}, ..., X_t\}$ is the window data at moment. Anomaly detection for data X_t can be translated into anomaly detection for subsequence W_t. Anomaly detection on the data can be translated into anomaly detection on the subseries. When we divided the data into time windows, we experimentally confirmed that different lengths of time windows have an impact on the accuracy of the model. As shown in Figure 6, in order to test the effect of the length of the time window on the accuracy rate, 10% of the error points are randomly added to the data, and it can be seen that the accuracy rate reaches a maximum of 0.995 when the length of the time window is around 15, so the length k of the time window is set to 15 in future experiments.

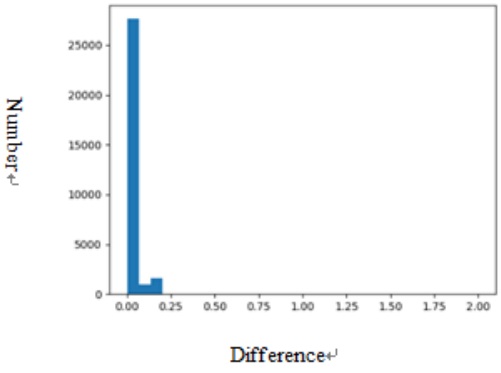


Figer 6. Effect of time window length on accuracy rate

6) Setting of the threshold value

In the experiment, we use the self-encoder for noise reduction of the data, and the bit error point will be close to the real data, which has little change. Using this feature of the model, we perform the difference operation on the data before and after training, and the difference value exceeds the threshold value we mark as the bit error point. To determine the threshold value, we counted the difference values of the training data and found that most of the differences were within the range of less than 0.25, as shown in Figure 7, for which we set the threshold value to 0.25. The

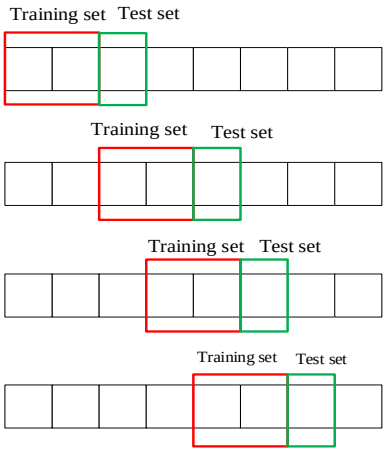
points exceeding this value were marked as false code points for output.



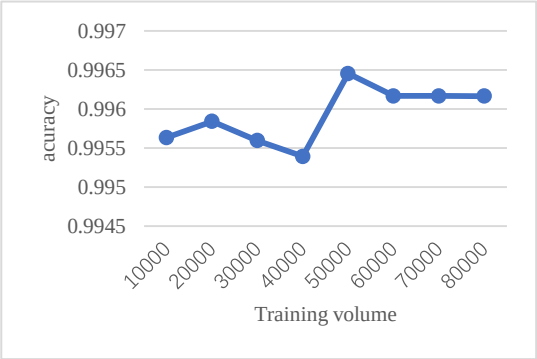
Figer 7. Difference statistics

7) Real-time testing

The application background of the model is different from other time-series data processing because satellite data is constantly sent to the ground, the amount of data generated is very large, we need not only the detection of historical data but also the real-time detection of data, if we use a period of data for testing will cause a certain degree of delay for our practical applications, we can not put all the data into the We cannot put all the data into the model for training and testing. For this reason, we adopt a training and testing approach based on the idea of a sliding window, in which a model is trained and saved before starting to detect the data, and when a certain amount of data is detected, training is started again, and the latest saved model replaces the previous one, and the model is updated once for the next data detection. When saving the model, we use the method of saving the network parameters and the network structure separately, as shown in Figure 8, which also reduces the requirement for hardware equipment. We also conducted special experiments on how much data to take for training each time to make the accuracy reach a high level, as shown in Fig. 9, the accuracy rate tends to stabilize after the training amount reaches 50,000, and the accuracy rate reaches about 99.6%, so we use 50,000 data for training in the experiment. One data point is tested at a time, and the network weights are updated once after the cumulative 50,000 data points are trained. A printing operation is performed after each test of 10,000 data, and the experimental results are printed out as shown in Figure 10.



Figer 8. Program execution proces

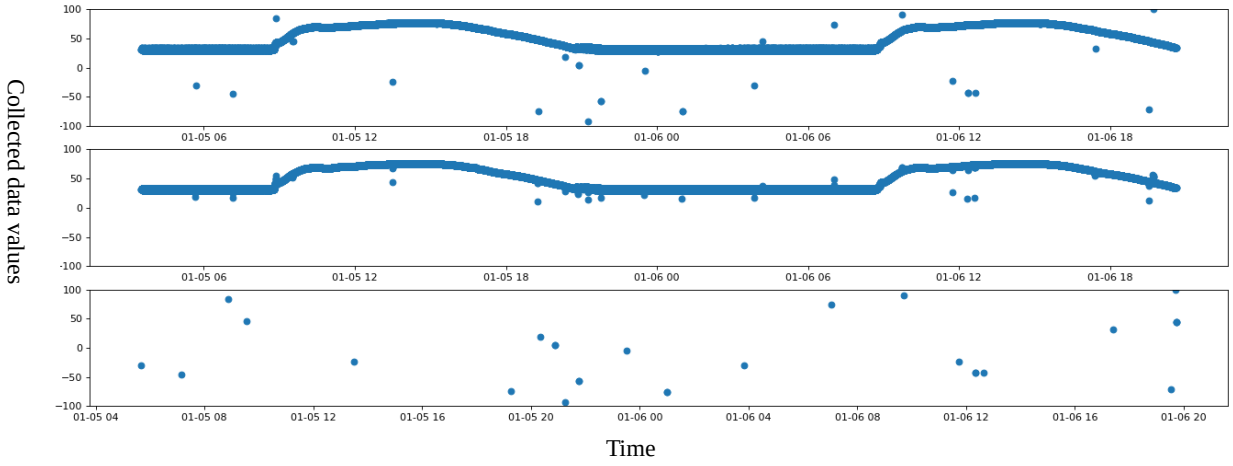


Figer 9. Relationship between training volume and accuracy rate

C. Analysis of experimental results

In the experiment, 50,000 data are used as training, one data is tested each time, and the data are iterated 15 times. As shown in Figure 10, the first image is the image before data processing, the second image is the image after noise reduction using the trained model, and the third image is the identified error code points. After the actual test, the accuracy of the model reaches

99.95%，It can be considered that our model achieves a better result, after which we would like to add other types of satellite data for comparative analysis.



Figer 10. Experimental results

V. CONCLUSION

Spacecraft are multi-functional and complex systems that can be used in many fields, such as navigation satellites and meteorological satellites, etc. Therefore, spacecraft are high-value targets, and it is of great importance to ensure their safe and stable operation. The only means to obtain the operational status of spacecraft is telemetry data, which can reflect the orbital operation of spacecraft and the operational status of each component. Due to the communication or sampling reasons, some of the telemetry data may contain wrong data (abnormal data), which can greatly interfere with the subsequent spacecraft analysis and even make the fault detection and health analysis of the spacecraft impossible. Therefore, certain means are needed to analyze the telemetry data to evaluate the accuracy of the data itself and prepare for further analysis of the spacecraft. However, there is no effective means to analyze the reliability of telemetry data. Therefore, by studying the characteristics of telemetry time series data and the existing time series data analysis methods, this paper focuses on the anomaly detection method of telemetry time series data with spacecraft telemetry data as the research object and uses the self-encoder related technology.

REFERENCES

[1] Kullback S., Leibler R. A.. On information and sufficiency [J]. The Annals of Mathematical Statistics, 1951, 22(1): 79-86

[2] Zhang L. W., Lin J., Karim R.. Adaptive kernel density-based anomaly detection for nonlinear systems [J]. Knowledge-Based Systems, 2018, 139: 50-63

[3] Liu F, Qi JP, Yu YW, et al. Density-based top-n local anomaly fast detection algorithm[J]. Journal of Automation, 2019, 45(9): 1756-1771

[4] Shang W-L, Shi He, Zhao J-M, et al. An SAE-LSTM-based method for process data anomaly detection [J]. Journal of Electronics, 2021, 49(08): 1561-1568.

[5] Liu Yongmei. Research on the anomaly detection method of satellite power parameters based on SDAE [D]. Harbin Institute of Technology, 2021.

[6] Guo Lei. Research on rolling bearing fault diagnosis based on SDAE and LSTM fusion [D]. Shenyang University of Architecture, 2020.

[7] Zhang, R. L., Wang, W., Guan, X. H.. Hidden Markov model-based anomaly detection for program behavior [J]. Journal of Xi'an Jiaotong University, 2005, 39(10): 1056-1059.

[8] Cuan-Ying,Wu-Yue. Research on time series anomaly detection method based on Seq2Seq depth self-encoder [J]. Modern Electronics Technology, 2022, 45(02): 26-30.

[9] Jin Yupeng. Research on key issues of air pollutant time-series prediction based on improved self-coding network [D]. Shanghai Normal University,2019.

[10] DU Xiaolei,CHEN Zhigang,XU Xu,et al. Research on bearing fault diagnosis based on wavelet convolutional self-encoder and LSTM network [J]. Mechatronics Engineering, 2019, 36(07): 663-668.

[11] Cao Lichang, Deng Dexin, Shen Jian, et al, Meng Haitao. Research on integrated monitoring technology for monitoring station data quality [C]. Proceedings of the 12th Annual China Satellite Navigation Conference - S06 Time Reference and Precision Timing. 2021:54-58.

[12] Chen, J.S., Li, J., Chen, W.G., et al. Wind turbine state anomaly detection method using sliding window and multiple additive noise ratio stacking noise reduction self-coding [J]. Journal of Electrical Engineering Technology, 2020, 35(02): 346-358.

[13] Chen J, Liu M, Xiong P, et al. Noise reduction of ECG signals based on convolutional self-coding neural networks [J]. Computer Engineering and Applications, 2020, 56(16): 148-155.

[14] Yuanyuan Wei, Julian Jang-Jaccard, et al. LSTM-Autoencoder based Anomaly Detection for Indoor Air Quality Time Series Data [J]. arXiv preprint arXiv: 2204. 06701v1, 2022.

[15] Victoria Bell, Divish Rengasamy, Benjamin Rothwell , Grazziela P Figueredo. Anomaly Detection for Unmanned Aerial Vehicle Sensor Data Using a Stacked Recurrent Autoencoder Method with Dynamic Thresholding [J]. arXiv preprint arXiv: 2203. 04734, 2022.

Research on Extraction Method of Financial Knowledge Based on How Net

Chaoyang Geng
School of Computer Science and Engineering
Xi'an Technological University
Xi 'an, 710021, China
E-mail:541211200@qq.com

Peng Liu
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 643973010@qq.com

JieJie Zhao
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail:13734100641@163.com

Dan Yang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 1330311378@qq.com

***Abstract*—In order to obtain the knowledge information of financial texts more efficiently and make the extracted information such as entity relation attribute more accurate, this paper studies the grammatical features of financial news texts and the semantic features of How Net, and puts forward the scheme of financial information extraction based on How Net. First, the phrase matching is carried out in the dictionary. Then the neural network is used for weighting, BiLSTM is used for character vector feature enhancement training, and then conditional random field (CRF) is used to complete named entity recognition, and then the relationship extraction of entity pairs from the dependency syntax is carried out to complete the research on the construction method of knowledge extraction of text in the financial field. The experimental results show that this model is superior to the other three models in entity recognition, and the overall performance is improved by about 1.2%. In relation extraction, the accuracy and recall rate of the model algorithm adopted in this paper are improved by 5% and 1.5% respectively, which shows that the improvement of the algorithm is effective.**

Keywords-Named Entity Identification in the Financial Sector; Relation Extraction; How Net

I. INTRODUCTION

Information in the financial field is directly or indirectly related to all fields of social life, and it has the characteristics of huge data, various types and rich text formats. At present, with the rapid development of artificial intelligence, the fintech industry has gradually turned from manual to intelligent. News texts in the financial field are mostly unstructured text information, which contains a large amount of knowledge. Knowledge extraction can gather the entity relationship attributes and other information in financial news texts together in the form of triples to obtain efficient information for financial practitioners.

The entity recognition of financial texts plays a vital role in completing the construction of knowledge extraction in the financial field. Therefore, the named entity recognition of financial information texts is one of the main research contents of this paper. At present, there are two new challenges in naming and recognition in the financial field: the first is to identify the name of the organization: the name of the organization in the financial text often contains some personal names, place names, different abbreviations or unknown words, and its naming rules are not restricted by a unified format; The second is the recognition of terms in the financial field: entities in terms in the financial field are characterized by complex types and irregular updates. The boundary of professional terms in the financial field cannot be properly divided by common word segmentation tools, and colloquial

expressions of financial terms cause ambiguity identification, such as track and leverage.

Entity relationship extraction technology extracts structured relational triples from unstructured or semi-structured financial texts, which can not only promote the construction of knowledge base, but also provide support and help for intelligent retrieval and semantic analysis of financial knowledge. At present, the key problem in entity relation extraction is that overlapping triples seriously affect the efficiency of triplet recognition.

By studying the grammatical features and semantic features of financial news text, this paper designs the entity relation extraction algorithm of financial news text integrating How Net, and completes the knowledge extraction in the financial field. This paper proposes a scheme of financial entity identification based on How Net and BiLSTM-CRF, then elaborates on feature selection and named entity annotation, and then completes the relationship extraction between entity pairs through a relationship extractor including dependency syntax module. On this basis, the entity relationship extraction flow chart of financial news text is built. Finally, the effectiveness of the scheme is verified by experiments.

II. INFORMATION EXTRACTION RELATED WORK

As one of the important basic technologies of information extraction, entity relation extraction is mainly aimed at identifying the required entities and the semantic relationships between entities from unstructured natural language texts. This technology can provide important support for the construction of knowledge base and knowledge graph and the realization of intelligent search and intelligent question and answer.

A. Named entity recognition

The main task of named entity recognition in the financial field is to use natural language processing and artificial intelligence technology to extract financial entities in professional fields other than the names of people, places and institutions, such as the company name, company name, financial product name, project name, etc.

In the field of financial data, people first began to study entity recognition through a rule-based and dictionary-based approach, and the entities studied were only the abbreviation of the name of the institution. Wang Ning et al. [1] summarized the structural features and context information of the company name, then constructed six knowledge bases for identifying the company name manually, and completed entity identification through two traversal search and identification. The experimental results showed that in the closed test, the accuracy of the recognition of the organization name could reach 97.3%, and the recall rate reached 89.3%. But in addition to being time consuming, this approach requires a strong background in finance, so the researchers focused on statistical machine learning methods.

Statistical machine learning is a method based on feature engineering. This method is to manually select and design relevant features, then vectorize these features, and then use mathematical function models to make predictions. Although the statistics-based machine learning method has achieved good results in the data sets of the general domain, it is still difficult to identify financial named entities in the field of finance, especially in the field of finance. Therefore, Wang et al. [2] proposed a financial named entity recognition method based on conditional random field and mutual information and information entropy. Since stock names often appear in financial texts, the author adopted domain dictionary to identify stock names. They integrated the rule features of company name suffix, place name prefix, title and subsequent words into a linear CRF classifier to identify the complete financial naming entity. The experimental results showed that the accuracy of the model increased by about 6% after adding mutual information and information entropy.

Because of its powerful learning ability, deep learning has been applied to the field of natural language processing in recent years. Li Qianwen [3] proposed the structure of Word-BiLSTM-Attention-CRF cyclic neural network model, which was based on the BiLSTM-CRF model, and then took the vector combining words as the input

vector, and then added the attention mechanism into the network structure. This method effectively improved the effect of named entity recognition in the financial field. Wang Guo-ming [4] proposed a Bi-LSTM and CRF model, namely BLaC model (Bert-bi-Lstm-attention-CRF), which takes BERT pre-trained word vector as input and integrates the Attention mechanism. Experimental results show that this model is more effective than other models in named entity recognition. He Yunqi et al. [5] improved the performance of entity recognition by combining CRF models with a series of syntactic and semantic features, and achieved good results.

B. Relationship Extraction

Relationship extraction is to retrieve the existing relationships among entities on the basis of extracting a large number of entities, which is one of the important tasks of information extraction. Currently, the mainstream entity relationship extraction methods can be divided into three categories: supervised method, statistical machine learning method and deep learning method.

For relation extraction of unstructured text, the earliest method adopted by researchers is artificial construction rules for extraction. For example, Yu Li et al. [6] adopted Bootstrapping method for relation extraction in geographical domain, which can automatically mine descriptors representing entity relations by considering various text features such as parts of speech and location. Make it suitable for large-scale relationship extraction. However, rule-based methods often require a strong background of domain knowledge, as well as in-depth observation and analysis of data, which requires a high labor cost.

The method based on statistical machine learning defines all possible categories in advance, then trains the classifier according to the annotated corpus, and uses the classifier to predict entity relationships. For example, Kambhatla [7] et al. used dependency parsing to obtain various syntactic and semantic features of the corpus, and then used the maximum entropy model to classify them, providing ideas for subsequent researchers.

With the development of deep learning, people gradually pay attention to the field of deep learning. Liu et al. [8] proposed a relationship extraction model based on dynamic long - and short-term networks, in which entity information, entity location information and semantic features were introduced. The accuracy rate reached 72.9%, the recall rate reached 70.8%, and the F1 value reached 67.9%. By integrating the attention mechanism with the improved convolutional neural network model, Zhao Pengwu [9] dynamically extracted entity relationships by focusing on contextual semantic information, and automatically extracted features from training data by using the improved convolutional neural network model. Then, through comparison and verification with multi-dimensional experimental results of different models and entity relationship extraction efficiency of different vector training sets, it is proved that the prediction accuracy and global performance of CNN+Attention model are superior to SVM, LR, LSTM, BiLSTM and CNN model.

III. INFORMATION EXTRACTION BASED ON HOW NET

A. How Net

How Net holds the idea of reductionism, that the meaning of all words can be composed of "meaning", that is, the most basic semantic unit that should not be divided. And How Net believes that meanings and words in real world scenarios can be represented with limited meanings. In How Net, semantic primitives are unique and can accurately represent the semantic information of words by marking some complex semantic relationships between them, such as subordination and upper and lower relation. Therefore, this paper adds semagrams to word representation learning, aiming to improve the quality of word representation learning through semagrams embedding.

How Net uses various relationships to express the meanings of many meanings, such as definitions and modifiers, and forms a complex hierarchy. In addition, the knowledge base mainly describes concepts, including more than 2000 semantic primitives. Knowledge System Description Language (KDML) is its special language, which can describe the relationship

between concepts and meanings in the knowledge base. Therefore, this paper proposes a knowledge extraction method based on How Net. Firstly, entity extraction is carried out by combining How Net and bidirectional long and short time memory network. Then, dependency parsing module is added to the model and the relationship between entity pairs is extracted by using it to form an information extraction framework based on How Net.

B. BiLSTM-CRF Model

The network structure of BilSTM-CRF model is shown in Figure 1, which is composed of word embedding layer, BiLSTM layer and CRF layer:

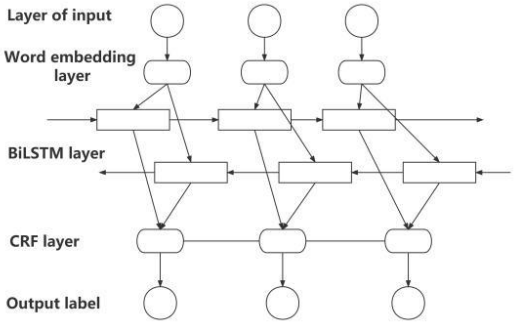


Figure 1. Network structure of BiLSTM-CRF model

1) Word embedding layer: This layer turns preprocessed words into low-dimensional dense vectors that can be processed by the model. The resulting vectors are distributed representations of words, which can represent semantic distance between words through the relationship between words and context. The process of this layer is to first train the word model with unmarked data, and then build a word vector matrix according to the word list, where v represents the size of the word list, d represents the dimension of the word vector, and then convert the text information into the id corresponding to the word list z, and map the id to the word vector matrix of the text, and s represents the number of words contained in the text.

2) BiLSTM network layer: BiLSTM is a bidirectional short and long time memory network, which is composed of forward LSTM and backward LSTM. Bidirectional text information can be extracted through BiLSTM. In this layer, the word vector matrix obtained by the word embedding layer is input into the forward LSTM and backward LSTM respectively, and then the forward LSTM output and backward LSTM output are combined according to the position to obtain the output of BiLSTM, which represents the score of each word corresponding to each category. For example, code the sentence "domestic crude steel production record high" and input "domestic", "crude", "steel", "output", "new high" forward to get five vectors $\{h_{L0}, h_{L1}, h_{L2}, h_{L3}, h_{L4}\}$. Input "New high", "output", "steel", "crude" and "domestic" successively in the back to get five vectors $\{h_{R0}, h_{R1}, h_{R2}, h_{R3}, h_{R4}\}$. And then to implicit vector before and after to the implicit vector according to the position of the splicing get $\{[h_{L0}, h_{R4}], [h_{L1}, h_{R3}], [h_{L2}, h_{R2}], [h_{L3}, h_{R1}], [h_{L4}, h_{R0}]\}$, $\{h_0, h_1, h_2, h_3, h_4\}$.

3) CRF layer: This layer is a conditional random field layer, which can "judge" the probability of the current state according to the state of the surrounding nodes. Such as X, Y is random variable, P (Y | X) is the probability distribution of under the condition of X Y, if random variable Y consisting of undirected graph G (V, E) is a markov random field, says the probability of undirected graph is a conditional random field (CRF). This layer can combine the output of BiLSTM layer as its input and take the state transition matrix as one of its parameters, and then learn and predict the label information to get the best label sequence of the text. Its calculation is shown in Equation (1).

$$score(X, y) = \sum_{i=1}^{n} A_{y_i, y_{i+1}} + \sum_{i=1}^{n} P_{i, y_i} \qquad (1)$$

Where, A is the transfer matrix, representing the probability of transferring from the label to the label, the probability of the first word being labeled as the label, and the probability fraction of the input sentence sequence being labeled as the label sequence of is. The maximum value can be obtained to obtain the maximum label sequence of the current sample sentence.

C. Feature Selection

The words in a sentence are the basis of named entity recognition, so this paper takes the grammar and semantics of the words as the basic features.

Grammatical features are composed of words, word structures, parts of speech and morphemes, which refer to the surface features of words and sentences. Semantic features are in-depth word mining to find the correlation between named entities. This paper uses How Net as semantic features to improve the effect of named entity recognition in financial texts.

Since many words are often polysemous or have multiple words with one meaning, researchers have combined the knowledge of linguistics, cognitive science and artificial intelligence to build a semantically based dictionary to solve the non-unique semantic relationship between words. Such as the 1980 s by Princeton university George Miller team to build a large dictionary WordNet, its nouns, verbs, adjectives and adverbs in the form of a synonym set (synsets) stored in the database, the main relationship between words are synonyms. Domestic dictionaries include How Net, a large language knowledge base marked by Dong Zhendong and Dong Qiang for decades. After years of development and expansion, this knowledge base contains almost all the words used in daily life, and has extremely important application value in terms of lexical similarity calculation, text classification and information retrieval.

How Net believes that words and meanings can be described by the original meaning, and the original meaning is unique, and the semantic relationship between the meanings is also marked with complex. By adding How Net into the entity recognition model as a semantic feature, this paper can deeply explore the relationship between entities. In How Net, the record of each entry is composed of four parts: word, example, part of speech and concept, as shown in Figure 2(a). Take "net profit" as an example. Its composition in How Net is shown in Figure 2(b)

W_X = WORD

E_X = Examples of Words

G_X = Part of speech

DEF = Definition of Concepts

(a)

NO.=094557
W_C=net profit
G_C=noun [2] [jing4 li4]
E_C=
W_E=net profit
G_E=noun
E_E=
DEF={wealth|money:domain={finance | financial}, quantity = {net net |}, {earn earn | : possession = {~}}

(b)

Figure 2. How Net composition

In figure 2 (b), the "net profit" by DEF is described, in DEF "wealth | money" is the first meaning of the original "net profit", "{finance | financial}, quantity = {net net |}, {earn earn | : possession = {~}}" is the first original righteousness "wealth | money" domain limit; How Net takes the first semantic as the main meaning to express the semantic concept of words, and the semantic is to reflect the essence of the concept. Therefore, the first semanthems of entity classes and event classes in How Net are added into the named entity recognition stage of this paper as semantic features to improve the entity recognition effect.

D. Dependency syntax module

One of the factors causing noise propagation in relation extraction is that the text is long and the words describing the relation in the text are obscure, so the effective information in the sentence cannot be better learned by the relation classification model. The dominant and dominated relation between the words forming a sentence is called dependency relation, which is represented by arcs and dominated by core verbs in dependency parsing. Therefore, the syntactic structure of a sentence can be represented by a structural graph with words as nodes and relations as edges.

One way to express the interdependence between words and syntactic structure information in a sentence is to use syntactic dependency tree, which takes key verbs in a sentence as root nodes and relations as edges to form a tree-like structure, which can clearly show the logical relationship between words in a sentence. In the syntactic dependency tree, no matter how far the physical distance is, as long as there is a mutual modifying relationship between words, the distance in the tree will be very close, which can solve the problem of irregular expression.

The syntactic dependency tree is usually trained as an input feature of text, and then the corresponding weight is obtained, which can play the role of predicting relations. In this study, dependency parsing is used to extract important information from the text. In this part, only dependency subtrees containing entity pairs in the sentence are used instead of the dependency tree structure of the whole sentence, and the dependency subtrees containing entity pairs are encoded into local feature vectors, which is convenient to reduce noise and strengthen the direct relationship features of entity pairs. This study uses syntactic dependency tree for analysis, constructs a graph representation of dependency tree, and then uses the graph representation to calculate the attention factor. It uses the bidirectional long and short memory network to train the semantic features of the text, and finally adds the semantic features and attention weights to realize the extraction of inter-entity relations.

For the purpose of logical relation between words in sentences and relation extraction, this paper firstly uses tools to obtain the expression of dependency tree of text. The dependency tree takes core words as the root, and then describes the relation between words and adds them to the tree according to grammatical rules. After adding the relation between all words of a sentence to the tree, the syntactic dependency tree of the sentence is formed. Syntactic dependency can well describe the relationship between words. Traditional machine learning algorithms usually convert the syntactic dependency into a vector and then combine it with the semantic vector of the text as the input of machine learning, but this method cannot achieve effective integration with the semantic vector. Therefore, this paper uses syntactic dependency tree to generate graph representation. Then the attention is calculated by using the graph representation. Finally, the attention vector is multiplied by the semantic vector to realize the real fusion of syntactic vector and semantic vector. The specific steps are divided into the following 4 steps:

1) Generate syntax dependency tree. The syntactic analysis tree of "Synchronized development of Shanghai Pudong and legal system construction" is obtained through word segmentation and syntactic analysis, as shown in Figure 3.

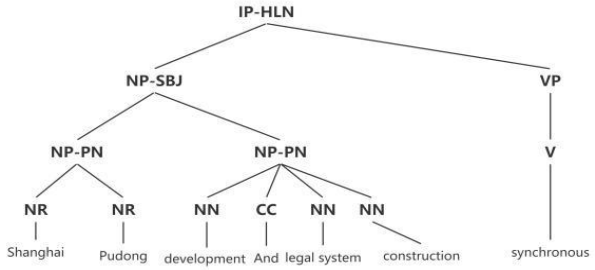


Figure 3. Syntactic analysis tree of "Shanghai Pudong Development and legal construction Synchronization"

2) According to the dependency tree results in FIG.3, the structure of directed graph is generated. In the directed graph, the specific dependency relationship in the graph is ignored, and only whether there is a relationship between words is concerned.

$$S_{ij} = \begin{cases} 1, \text{There is a dependency} \\ 0, \text{There is no dependency} \end{cases} \tag{2}$$

3) Use the two-dimensional table generated in step 2 to calculate the sentence attention distribution, as shown in Equation (3):

$$a_{t,i} = \frac{\exp(S_{t,i})}{\sum_{i=1}^{T} \exp(S_{t,i})} \tag{3}$$

4) The sentence's attention distribution is obtained, and the attention weight and the hidden vector trained by the bidirectional gated neural network are fused to calculate the semantic value, as shown in Equation (4) :

$$C_t = \sum_{i=1}^{T} a_{t,i} h_i \tag{4}$$

IV. EXAMPLE SIMULATION

The financial relationship data set in this pap er is 200,000 financial and economic news materi als crawled from the People's Internet as the corpu s, and then the BIO annotation system is adopted t o label the data set. The data set obtained in this st

udy is randomly divided into two parts, 80% of w hich are taken as the training set and 20% as the te st set.

A. *Experimental results of entity recognition*

In the feature extraction experiment, this stud y first extracts the Chinese word W_C from How Net and defines DEF (the first meaning of DEF), t hen takes W_C as the key and DEF as the corresp onding value to build the semantic dictionary requ ired by this paper, and finally uses the semantic di ctionary to conduct semantic annotation for the te xt in the financial field. In this study, HanLP was first used as a basic word segmentation tool, and t hen BiLSTM-CRF model integrated with How Ne t semantic features was used for analysis to obtain the experimental results of named entity recogniti on in this study. As shown in Table 1, feature extr action and BIO annotation results of "Internet Fin ance Association of China is registered and establi shed in Shanghai with the approval of Ministry of Civil Affairs" are obtained by using HanLp word segmentation tool and combined with How Net se mantics.

TABLE I. "CHINA... "IS VALID

Participles	The part of speech	How Net Semantics	BIO
China	ns	Local, national	B-ORG
Internet	n	The Internet	I-ORG
Finance	n	Business, money	I-ORG
Association	n	group	I-ORG
With	p	Function words	O
Ministry of Civil Affairs	nt	institutions	B-ORG
approval	v	Give one's consent	O
Shanghai	ns	place	B-ADDR
registered	v	record	O
establishe	v	To establish	O

In the experiment, Precision (P), Recall (R) a nd F1 (F-score) were used as evaluation indexes f or named entity recognition performance. As can be seen from Figure 4, this model is higher than th e other three models in terms of accuracy, recall r ate and F1 value. In the comparison experiment be tween BiLSTM model and BILSTM-CRF model, because BiLSTM model does not consider the infl uence of context and semantic environment on the label classification of words and words, the predi ction labels of each element are independent of ea ch other without dependence, and the CRF layer c an add some constraints to the final prediction lab els. These constraints can be learned automaticall y by the CRF layer from the training data set duri ng training. Therefore, the overall level of BiLST M model is lower than that of BILSTM-CRF mod el.

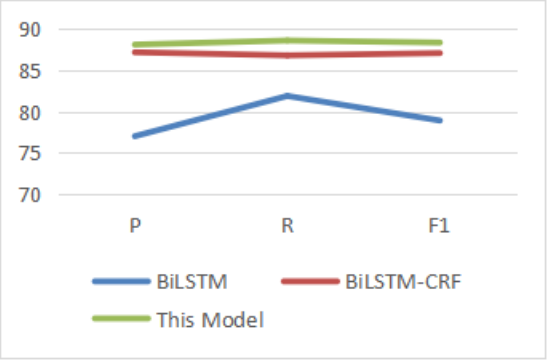


Figure 4. Comparison of P,R and F1 values

On the basis of the above advantages of the model, after adding How Net semantic features to the BiLSTM-CRF model, the overall performance is improved by about 1.2%, that is, the model can obtain more information than the comparison mo del, alleviating the imbalance between samples. T herefore, BiLSTM-CRF based on word embeddin g and How Net has achieved the best entity recog nition effect in each evaluation index.

B. *Relationship extraction results*

In this study, LTP of Harbin Institute of Tech nology was adopted as the support to build a synta ctic tree, and the machine learning training was co mpleted through Python programming. The param eters of the machine learning part were set as follo ws: Word vector dimension of training was set to 200, because this dimension was the best compreh ensive index of training time and effect. dropout v alue is 5.0; The learning rate was 0.01; Number of iterations: 200, hidden layer dimension is 200.

In this paper, Precision, Recall and F1 were u sed to evaluate the performance of the model algo rithm.

The algorithm in this paper is compared with the most classic LSTM algorithm, Bi-LSTM algo rithm and Bi-LSTM+self-ATT algorithm. Among them, Bi-LSTM+self-ATT algorithm is a bidirecti onal gated neural network integrated with self-atte ntion, which uses self-attention and attention is no t well integrated into the semantic features. The ac curacy and recall rate of the model algorithm used in this paper are improved by 5% and 1.5% respe ctively, which shows that the improvement of the algorithm is effective.

The experimental results are shown in Table 2.

TABLE II. COMPARISON OF EXPERIMENTAL RESULTS

Model	Precision	Recall	F1
LSTM	35.8	42.4	38.8
Bi-LSTM	57.5	55.4	56.4
Bi-LSTM+self-ATT	65.2	56.8	60.7
This algorithm	70.2	58.3	63.7

V. CONCLUSIONS

In this paper, a large number of entities are obtained by adding the semantic elements in How Net as features into entity recognition. Then, the syntactic dependency tree is used to add them into the relation extractor to extract the relationship between entities, and the knowledge extraction of financial texts is completed. In the aspect of entity recognition, the phrase is matched in the dictionary, then weighted by the neural network, the character vector is trained by BiLSTM, and then the conditional random field (CRF) is used to complete named entity recognition. By comparing with BiLSTM model and BilSTM-CRF model, it is proved that the algorithm improves the overall performance of named entity recognition by about 1.2%. In relation extraction, dependency syntax method and How Net semantic features are integrated to carry out relation extraction for the obtained entity pairs. The algorithm in this paper is compared with the most classical LSTM algorithm, Bi-LSTM algorithm and Bi-LSTM+self-ATT algorithm. It is proved that the accuracy and recall rate of this algorithm are improved by 5% and 1.5% respectively.

REFERENCES

[1] Wang Ning, GE Ruifang, YUAN Chunfa, HUANG Jinhui, LI Wenjie. Recognition of company names in Chinese Financial News., 2002(02):1-6.

[2] WANG S, XU R,LIU B,et al.Financial named entity recognition based on conditional random fields and information entropy[C]//2014 International Conference on Machine Learning and Cybernetics. Lanzhou: IEEE, 2014: 838-843.

[3] Qian-yu li. Financial entity oriented knowledge map construction research [D]. Yunnan university of finance and economics, 2020, DOI: 10.27455 /, dc nki. Gycmc. 2020.000492.

[4] Wang Guoming. Based on the deep study of the financial sector knowledge map construction research [D]. The northeast normal university, 2021. The DOI: 10.27011 /, dc nki. Gdbsu. 2021.000237.

[5] He Yunqi, Liu Suwen, Qian Longhua, et al. Disease Name Recognition Based on Syntactic and Semantic features [J]. Science China Information Science, 2018, 48 (11): 1546-1557.

[6] Yu L, Lu F, Liu X. A bootstrapping based approach for open geo-entity relation extraction [J].Acta Geodaetica et Cartographica Sinica, 2016, 45(5): 616-622.

[7] Kambhatla N. Combining lexical, syntactic, and semantic features with maximum entropy models for information extraction[C]//Proceedings of the ACL Interactive Poster and Demonstration Sessions. 2004: 178-181.

[8] Liu J, Ren H, Wu M, et al. Multiple relations extraction among multiple entities in unstructured text [J]. Soft Computing, 2018, 22(13): 4295-4305.

[9] Zhao Pengwu, Li Zhiyi, Lin Xiaoqi. Chinese Character Relation Extraction and Recognition Based on Attention Mechanism and Convolutional Neural Network [J/OL]. Data analysis and knowledge discovery: 1-16 [2022-05-26]. HTTP: / / http://kns.cnki.net/kcms/detail/10.1478.G2.20220511.1654.008.html

Research on Super-resolution Image Based on Deep Learning

Tong Han
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 110955443@qq.com

Chuang Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 10423848513@qq.com

Li Zhao
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 332099732@qq.com

Abstract—Image super-resolution is a kind of important image processing technology in computer vision and image processing. It refers to the process of recovering high-resolution image from low-resolution image. It has a wide range of real-world applications, such as medical imaging, security and others. In addition to improving image perception quality, it also helps improve other computer vision tasks. Compared with traditional methods, deep learning methods show better recons-truction results in the field of image super-resolution reconstruction, and have gradually developed into the mainstream technology. This article will study the depth in the super resolution direction is important method of types of introduction, combed the main image super-resolution reconstruction method, expounds the depth study of several important super-resolution network model, the advantages and disadvantages of different algorithms and adaptive application scenarios are analyzed and compared, this paper expounds the different ways in the super resolution to liquidate, Finally, the potential problems of current image super-resolution reconstruction techniques are discussed, and the future development direction is prospected.

Keywords-Super Resolution;Neural Network; Convolu-tional Neural Network; Deep Learning

I. INTRODUCTION

Image resolution is a set of performance parameters used to evaluate the richness of detail information contained in an image, including time resolution, spatial resolution and color level resolution, which reflects the actual ability of imaging system to reflect object detail information [1]. High-resolution images usually contain greater pixel density, richer texture detail, and higher reliability than low-resolution images. However, in practice, due to the constraints of acquisition equipment and environment, network transmission medium and bandwidth, image degradation model itself and many other factors, we usually can not directly obtain the ideal high-resolution image with edge sharpening and block blur [2]. The most direct way to improve the image resolution is to improve the optical hardware in the acquisition system. However, it is difficult to improve the manufacturing process and the manufacturing cost is very high, so it is often too costly to solve the problem of low image resolution physically. Therefore, from the perspective of software and algorithm, the technology of image super-resolution recons-truction has become a hot research topic in many fields such as image processing and computer vision. Image super-resolution reconstruction technology refers to the restoration of a given low resolution image into a corresponding high resolution image by a specific algorithm. To be specific, image super-resolution reconstruction technology refers to the process of reconstructing high-resolution images from a

given low-resolution image by using relevant knowledge in digital image processing, computer vision and other fields through specific algorithms and processing processes [3]. It aims to overcome or compensate the problems of blurred image, low quality and insignificant region of interest caused by the limitations of image acquisition system or acquisition environment [4]. A simple way to understand super-resolution reconstruction is to change a small image into a large one to make the image more "sharp".

The existing super-resolution reconstruction algorithms are generally divided into three categories: interpolation based methods, which are simple but provide too smooth reconstruction images, lose some details and produce ringing effect; The modeling-based method has a better reconstruction effect than the interpolation method, but when faced with a large amount of calculation, the calculation process is time-consuming, difficult to solve and greatly affected by the amplification factor [5]. Based on the learning method, this kind of algorithm solves the problem sensitive to the scale scaling factor and has the best reconstruction effect, which is the mainstream direction of current research.

Currently studying mainstream of super-resolution reconstruction algorithm based on the depth of the main network model is divided into three categories: convolutional neural networks and generate network and another rise nearly new method is to explore the inverse transformation of GAN, the three kinds of models can be a good image of high frequency information, improve the resolution of the images, closer to the original image.

II. NETWORK MODEL OF SUPER-RESOLUTION RECONSTRUCTION METHOD

At present, superresolution networks based on deep learning can be divided into three categories: (1) supersegmentation methods based on convolutional neural network model; (2) superscore method based on generative adversarial network model; (3) it is explored that the inverse transform of GAN [6]; Each has advantages and disadvantages

A. Superdivision method based on convolutional neural network direction

1）SRCNN：

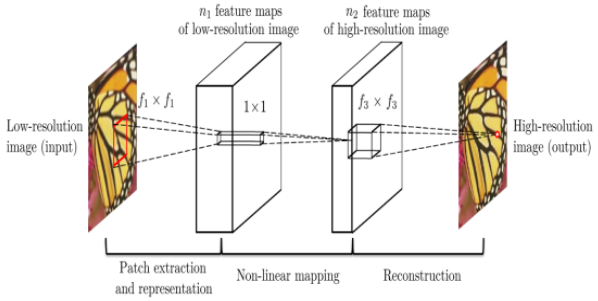


Figure 1. SRCNN network model.

SRCNN network consists of three modules: feature extraction, nonlinear mapping and final reconstruction [7]. These three modules correspond to three convolution operations. The first layer of CNN: feature extraction of input images. The purpose of block precipitation and representation is to obtain a series of feature maps from the input image Y. The second layer of CNN: nonlinear mapping of features extracted from the first layer. The nonlinear mapping corresponds to "convolution activation" operation [8]. The third layer of CNN: reconstruct the mapped features to generate high-resolution images. The Reconst-ruction process is also a convolution operation, but there is no activation function here [9].

SRCNN proposes to use conventional conv-olution collocation activation function to simulate the feature encoding process of the traditional SR method. From feature extraction to the final reconstruction, the convolution operation is used, which is very concise and efficient.

Loss function: SRCNN uses the mean squared error as a loss function to evaluate the difference between the network output and the true label.

The author conducted comparative experiments on 91images and mageNet datasets respectively. With the increase of iteration times, SRCNN obtained higher PSNR on ImageNet, which indicates that the increase of data volume can improve that the performance of the network [10].SRCNN first proposed to simulate the feature encoding process of the traditional SR method by

using a series of conventional convolution collocation activation functions. Compared with the traditional SR method, SRCNN is a simple end-to-end learning method with better network performance and inference speed.

2）ESPCN

The core concept of ESPCN is the subpixel convolutional layer. As shown in the figure above, the input of the network is the original low-resolution image. After passing through two convolutional layers, the feature image obtained has the same size as the input image, but the feature channel is r2(r is the target magnification of the image). The r2 channels of each pixel are rearranged into an r*r region, corresponding to an r*r size subblock in the high-resolution image, so that the feature image of size R 2*H*W is rearranged into a high-resolution image of size 1*rH*rW. Although this transformation is called sub-pixel convolution, there is no convolution operation. Through the use of sub-pixel convolution, in the process of image amplification from low resolution to high resolution, the interpolation function is implicitly included in the previous convolutional layer and can be automatically learned [11]. The image size is transformed only at the last layer, and the previous convolution operation is carried out on low-resolution images, so the efficiency is higher.

B. Based on generative adversarial network direction

1）SRGAN:

In this paper, generative adversarial network was used to solve the super resolution problem [12]. It is mentioned in this paper that the mean square deviation is used as the loss function when training the network. Although high peak signal-to-noise ratio can be obtained, the recovered images usually lose high-frequency details, which makes people unable to have good visual perception. SRGAN uses perceptual loss and adversarial loss to enhance the realism of the recovered images. Perceptual loss is the feature extracted by using convolutional neural network. By comparing the features of the generated image after convolutional neural network and the features of the target image after convolutional neural network, the generated image and the target image are more similar in semantic and style [13]. The original GAN text gives an example: the generation network G is the person who prints counterfeit banknotes, and the discrimination network D is the person who detects counterfeit banknotes. G's job is to make the counterfeit money he prints as much as possible to fool D, and D is to distinguish as best as possible whether the money he gets is the real money in the bank or the fake money printed by G. In the beginning, G was not good enough, and D was able to point out what was wrong with the bill. G After each failure to carefully summarize experience, strive to improve themselves, progress every time. Until the end, D could not judge the authenticity of the banknotes [14]. The work of SRGAN is that G network generates high resolution images from low resolution images, and D network determines whether the obtained image is generated by G network or the original image in the database. When the G-net can successfully fool the D-net, then we can complete the super resolution with this GAN.

In this paper, the SRResNet (the generating network part of SRGAN) is optimized by mean square error, and the results with high peak signal-to-noise ratio can be obtained. By calculating the perception loss on the high-level features of the trained VGG model to optimize SRGAN, and combining the discriminant network of SRGAN, the results with realistic visual effects can be obtained, although the peak signal-to-noise ratio is not the highest.

SRGAN provides a new loss function. In previous SR, MSE loss function is used to teach the network how to realize LR to HR, but this will smooth the details of the image. Although the PSNR is very high, the human eye does not have a good visual sense. GAN objective function can be defined as shown in Equation (1):

$$\min_{\theta_G}\max_{\theta_D} V(D,G) = E_{x\sim Ddata(x)}[logD(x)] + E_{x\sim D_Z(Z)}[log(1-D(G(x)))]. \quad (1)$$

Fixed, to learn adjustments, that is, in order to train a discriminator network. Then, we fix the discriminant network parameters to learn to adjust the parameters of the generating network. Its purpose is to make the parameters of the generating network as large as possible by adjusting them. That is to say, the training of the generator network is to make the output result of the discriminant network output a high score, so as to deceive the discriminator. Therefore, we can see that after the generator becomes stronger, the next discriminant network will continue to become stronger, increasing the ability to distinguish between true and false. The generator, in turn, will continue to increase the score of the fake (the output of the input after G) in the discriminator, and then the discriminator will continue to improve and iterate, and the two will fight and grow each other, and finally the trained generator network will be the network we want.

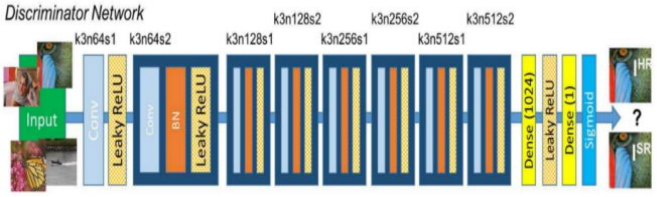


Figure 2. SRGAN generator and discriminator network structure diagram

SRGAN loss consists of two parts: content loss and adversarial loss, which are weighted and summed with a certain weight, as shown in Equation (2).

$$l^{SR} = l_X^{SR} + 10^{-3} l_{Gen}^{SR}$$
$$l_{MSE}^{SR} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta G}(I^{LR})_{x,y})^2 \quad (2)$$

In this paper, we define the VGG loss based on the ReLU activation layer of the pre-trained 19-layer VGG network to obtain the Euclidean distance between the feature representation of the image and the reference image. The feature map of a certain layer is proposed on the trained vgg, and this feature map of the generated image is compared with that of the real image.

$$l_{VGG/i,j}^{SR} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\varphi_{i,j}(I^{HR})_{x,y} \frac{n!}{r!(n-r)!} - \varphi_{i,j}(G_{\theta G}(I^{LR}))_{x,y})^2 \quad (3)$$

Equation (3) generates adversarial loss: generates a data distribution that the discriminator cannot distinguish. $G_{\theta G}(I^{LR})$ represents the probability that the image generated by the generator will be a natural image by the discriminator.

$$l_{Gen}^{SR} = \sum_{n=1}^{N} -logD_{\theta D}\left(G_{\theta G}\left(I^{LR}\right)\right) \quad (4)$$

Equation (4) is an improved generator loss function proposed in this paper. To minimize this expression is to maximize the probability that the generated image given to the generator by the discriminator is true. The goal is to fool the discriminant network by producing a high discriminant value.

2）ESRGAN:

In this paper, some improvements are made on the basis of SRGAN, including improving the structure of the network, the decision form of the adjudicator, and replacing a pre-trained network for calculating the perceptual loss. Three key parts of SRGAN are studied in detail: network structure; adversarial loss; Perceptual domain loss. And improve each item to get ESRGAN. ESRGAN makes several major improvements to SRGAN:

a) Introduce changes to the generator architecture.

b) The improvement of adversarial loss is mainly the use of elativistic GAN to make relative realness instead of the absolute value.

c) Perceptual loss is composed of features before activation (previously activated features).

d) Pre-train the network to optimize PSNR first, and then use GAN to fine-tune it.

To be specific, the paper proposed a resist-in-residual Dense Block (RRDB) network unit [15], in which the BN (Batch Norm) layer was removed.

In addition, let the discriminator predict the truth of the image rather than whether the image "is a fake image". Finally, the perceptual domain loss is improved by using pre-activation features, which can provide stronger supervision for luminance consistency and texture recovery. With the help of these improvements, ESRGAN gets better visual quality as well as more realistic and natural textures.

In the generator part, the author makes several changes to the generator G by referring to the SRResNet structure as the whole network structure: remove all BN layers; Turn the original block into residual-in-residual Dense Block (RRDB) which combines multi-layer Residual network and dense connections. Removing the BN layer has been shown to improve performance and reduce computational complexity [16].

The discriminator tries to estimate the probability that the real image is relatively more realistic than the fake image. In this paper, the loss function of the discriminator is defined as Equation (5):

$$L_D^{Ra} = -E_{Xr}\left[log\left(D_{Ra}\left(x_r, x_f\right)\right)\right] - E_{Xr}\left[log\left(1 - D_{Ra}\left(x_f, x_r\right)\right)\right] \quad (5)$$

The adversarial loss function of the corresponding generator is shown in Equation (6) :

$$L_G^{Ra} = -E_{Xr}\left[log\left(1 - D_{Ra}\left(x_r, x_f\right)\right)\right] - E_{Xf}\left[log\left(D_{Ra}\left(x_f, x_r\right)\right)\right] \quad (6)$$

Xf Is the image generated by the generator, Xr Is the input low resolution image.

The generator benefits from the gradient between the generated data in adversarial training and the actual data, and this adjustment enables the network to learn sharper edges and more detailed textures [17].

For generator G, its loss function is shown in Equation (7):

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta L_1 \quad (7)$$

For the discriminator, its loss function is shown in Equation (8):

$$L_D = L_D^{Ra} = -E_{x_r}[log(D_{Ra}(x_r, x_f))] - E_{x_f}[1 - log(D_{Ra}(x_r, x_f))]. \quad (8)$$

The ESRGAN proposed in this paper makes improvements on the basis of SRGAN, including removing BN layer, replacing the basic structure with RDDB, improving the discriminator discrimination target in GAN, and using the features before activation to form the perceptual domain loss function. Experiments have proved that these improvements are effective in improving the visual effect of the output image [18].

In addition, the authors also use some techniques to improve the performance of the network, including scaling of the residual information and smaller initializations. Finally, the authors use a network interpolation method to balance the visual effect of the output image with the PSNR and other index values.

3）REAL-ESRGAN

In the problem of single image super-resolution, many methods use the traditional Bicubic method to achieve downsampling, but this is different from the real world downsampling situation, which is too single.

Blind super resolution is designed to recover unknown and complex degraded low resolution images. According to the different down-sampling methods used, they can be divided into explicit modeling and implicit modeling. Explicit modeling: the classical degradation model consists of blur, down-sampling, noise and JPEG compression. However, the downsampling model in the real world is too complex to achieve the ideal effect through the simple combination of these methods.

Implicit modeling: It relies on learning the data distribution and using GAN to learn the degradation model, but this method is limited by the data set and cannot generalize well to the images distributed outside the data set. In the real world, image resolution degradation is usually a complex combination of many different degeneracies. Therefore, the network extends the CLASSICAL first-ORDER degradation model to the REAL-WORLD higher-order degradation model by using multiple repeated degradation processes, each of which is a CLASSICAL degradation model. However, in order to balance simplicity and effectiveness, the second-order degradation model is actually used in the code. However, because the high-order degradation model is adopted, the degradation space is much larger than that of ESRGAN, so the training is more challenging. So the network made two changes based on ESRGAN:

Use U-Net discriminator to replace VGG discriminator used in ESRGAN;

spectral normalization was introduced to make training more stable and reduce artifacts.

The authors of the article made the following study, A higher-order degradation model and sinc filters were proposed to model common ringing and overshoot artifacts.

Some basic changes (for example, the U-Net discriminator with spectral normalization) have been adopted to increase the discriminator's capabilities and training stability.

Real-esrgan trained on purely synthetic data is able to recover most real-world images, with better visual performance than previous works, and is more practical in the Real world. Disadvantages of Gan: Superresolution neural network based on gan has been a relatively mature scheme. However, such a method requires the generator not only to extract and retain the structure information of LR images, but also to generate as realistic and high-resolution texture information as possible, which is difficult for the generator, especially in high-magnification tasks. The resulting problems include but are not limited to: The generated image is blurred and the generated image has false texture information [19].

C. GAN Inversion

1) PULSE

The main purpose of single image superresolution is to construct a high resolution (HR) image from the corresponding low resolution (LR) input. In previous approaches, which are usually supervised, the training target is usually measured for the pixel-oriented average distance between superresolution (SR) and HR images. Disadvantages: Optimizing such metrics often leads to ambiguity, especially in areas of high variance (detail). We propose another scheme to simulate the super-resolution problem based on creating realistic SR images with the correct reduction in scale. In this paper, we propose a novel super-resolution algorithm to solve this problem, PULSE (photo upsampling through latent space exploration), which can generate high-resolution, realistic images that have not been seen in the literature before. It does this entirely in a self-supervised manner and is not limited to the specific degradation operators used during training [20], which is different from previous approaches that require training of the database of LR-HR image pairs for supervised learni-ng.Instead of traversing the LR image and slowly adding details, PULSE traverses the high-resolution natural image manifold, searching for images narrowed down to the original LR image [21]. This is formalized by a "scaled-down loss" that guides exploration of the latent space of the generative model. By exploiting the properties of high-dimensional Gaussians, we restrict the search space to ensure that our output is realistic. As a result, PULSE generates super-resolution images that are both realistic and correctly scaled down. We show extensive experimental results that demonstrate the effectiveness of our approach in the field of facial superresolution, also known as facial hallucinations. This paper also introduces the limitations and biases of the current implementation methods using adjoint model cards with relevant metrics. The proposed method outperforms the latest methods in perceptual quality, and has a higher resolution and scaling factor than before. Advantages and disadvantages: Firstly, a generative network is pre-trained, and then given an LR image with fixed parameters, the

model tries to explore a latent code z, which should be down-sampled from the HR image generated by the generative network to LR image. Then, by adjusting the z, the LR obtained from the generated HR downsampling is the closest to the real LR, and it can be approximated that the network generates the target SR image. Compared with the first model, the trained generative network can generate richer and more realistic texture information. This method sounds nice, but at high multiples, it is difficult to retain the structure information of the image just by code z, so the generated image will suffer Identity distortion. At the same time, in the process of image generation, for each SR image generated, it needs to be iteratively estimated several times, which is very time-consuming, so this method can not be applied to real-time tasks obviously.

III. SUMMARIZATION AND PROSPECT

Future improvement directions in the field of supersegmentation may include proposing more complex loss functions; Implement arbitrary super-resolution construction; While improving the performance, the pursuit of lightweight; Effective combination of various network modules; How to reduce the image quality of data sets, such as blind over segmentation technology to solve the problem of unknown degradation model.

In addition, the training data are difficult to obtain, at present most of the model using the simulation data, the process is difficult to imitate the real image is reduced, the real image degradation is not only a lower resolution, but also in the process, will introduce all kinds of image noise, so based on the sampling of the trained model easy to fitting, generalization ability is bad. It is difficult to generalize the model. For specific types of images, such as facial super segmentation, it is necessary to train the face-related model specially, and the general super segmentation model is often difficult to obtain good results.

Although the performance of the existing deep learning image super-resolution reconstruction algorithms has been greatly improved compared with the previous ones, far surpassing the traditional algorithms, there is still a lot of room for improvement. Looking into the future, the research on super-resolution can be carried out from the following aspects:

(1) Improve network performance. Improving the image effect after reconstruction has always been a hot issue for researchers, but for different use needs, the performance requirements of the network are also different. For example, in video surveillance images, it is necessary to reconstruct the image with good visual perception effect and high reconstruction efficiency. In medical image reconstruction, it is necessary to reconstruct the image with better texture details and ensure authenticity. Therefore, the reconstruction efficiency is improved and better visual perception is obtained

Fruit, better texture details, higher magnif-ication and other aspects are the focus of future research to continue to improve the performance of super resolution network.

(2) Application of image super-resolution in various fields. For example, in the aspect of video, we will continue to optimize video enhancement algorithms including super-score algorithm, create industry-leading image restoration and image enhancement technology, help customers improve video quality, reduce video playback cost, provide lower consumption, lower power consumption, better subjective quality, With more models and algorithms that save more bit rates, users can enjoy UHD video experience on different mobile phones and networks.

REFERENCES

[1] Murshed H,Wei Z,Ahmed M . Perfect Single Image (SR)Super-Resolutionwith,Deep Super Resolution Convolutional Neural NetworkandOpenCV Method [J]. IOSR journal of computer engineering, 2020(3):22.

[2] Ward C M, Harguess J, Crabb. Image quality assessment for determining efficacy and limitations of Super-Resolution Convolutional Neural Network (SRCNN)[C]//Applications of Digital Image Processing XL. SPIE, 2017, 10396: 19-30.

[3] Wang Lie, Yin Jin-wei. Small Object Detection Method Based on SRCNN and SSD network [J].Computer simulation, 2020, 37(3):5.

[4] Shen H F, Li P X, Zhang L P. Overview of image super-resolution reconstruction techniques and methods [J]. Optical Technology, 2009(2):7.

[5] Pu Jian, Zhang Junping, Huang Hua. Review of super-resolution algorithms [J]. Journal of Shandong University: Engineering Science Edition, 2009(1):6.

[6] ANWARS, KHANS, BARNESN. A deep journey into super-resolution: a survey[J]. ACM Computing Surveys (CSUR), 2020, 53(3):1-34.

[7] Kim J, Lee J K, Lee K M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2016

[8] Wang Jiaming, LU Tao. Satellite image super-resolution algorithm based on multi-scale residual deep neural network [J]. Journal of Wuhan Institute of Technology, 2018, 40(04):440-445.

[9] Wan Xuefen, Cui Jian, WANG Guanjun. Research on Image Super-resolution Reconstruction Processing Algorithm[C]// National Conference on Optoelectronics and Quantum Electronics Technology. Chinese Institute of Electronics, 2011.

[10] TIAN Yan, TIAN Jinwen, LIU Jian. Implementation for Super Resolution--An Improved Image Interpolation Based on Wavelet Implementation of Super-Resolution Technology -- An Improved Wavelet Interpolation Method [J]. Journal of Image and Graphics, 2003, 45(12):1422-1426.

[11] Jiang Hao, Wang Bofu, Zhuang Qiliang. Reconstruction of turbulent flow field based on super-resolution reconstruction method [J]. Experimental Fluid Mechanics, 2022(036-003).

[12] Wang Rong, Zhang Yonghui, Zhang Jian. Image super-resolution Reconstruction Method based on CNN [J]. Computer Engineering and Design, 2019, 40(6):6.

[13] Zhong Z, Chen Y, Hou S. Super-resolution reconstruction method of infrared images of composite insulators with abnormal heating based on improved SRGAN [J]. IET generation, transmission & distribution, 2022(10):16.

[14] Zou Penghui, Zeng Yijie, Duan Zhenghong. Research on image super-resolution Reconstruction based on SRGAN technology [J]. Science and Technology Trends, 2019(18):1.

[15] Liu Yiwen. Research on Low resolution face Detection Algorithm Based on Deep Learning [D]. University of Electronic Science and Technology of China, 2019.

[16] Hu Lei, Wang Zugen, Chen Tian, et al. An Improved super-resolution Reconstruction algorithm for SRGAN infrared Image [J]. Journal of System Simulation, 2021(033-009).

[17] Nagano Y, Kikuta Y. SRGAN for super-resolving low-resolution food images[C]//Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management. 2018: 33-37.

[18] Li J, Wu L, Wang S. Super resolution image reconstruction of textile based on SRGAN[C]//2019 IEEE International Conference on Smart Internet of Things (SmartIoT). IEEE, 2019: 436-439.

[19] Wang X, Yu K, Wu S. Esrgan: Enhanced super-resolution generative adversarial networks[C]//Proceedings of the European conference on computer vision (ECCV) workshops. 2018: 0-0.

[20] Wang X, Xie L, Dong C. Real-esrgan: Training real-world blind super-resolution with pure synthetic data[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 1905-1914.

[21] Menon S, Damian A, Hu S. Pulse: Self-supervised photo upsampling via latent space exploration of generative models[C]//Proceedings of the ieee/cvf conference on computer vision and pattern recognition. 2020: 2437-2445.

Construction of Driving Condition Based on Discrete Fourier Transform and Improved K-Means Clustering Algorithm

Shuping Xu
School of computer science
Xi'an University of technology
Xi'an, 710032, China
E-mail: 563937848@qq.com

Yueqiu Huang
School of computer science
Xi'an University of technology
Xi'an, 710032, China

***Abstract*—In view of the low execution efficiency and slow convergence speed of traditional clustering algorithms, the initial clustering center has a greater impact on the clustering results, which leads to the problem of reduced algorithm accuracy. This paper proposes an improved K-means algorithm (Grid-K-means), that is, the Grid density is used to determine the initial clustering center; According to the density, the grid points are sorted to eliminate the idea of noise grid points and invalid grid points, so as to improve the efficiency and accuracy of the algorithm. First, the discrete Fourier transform was used to filter the original data, and then the principal component analysis and the improved K-means clustering algorithm were used to reduce and classify the kinematics fragments respectively, so as to construct the driving conditions of the vehicle. The experimental results show that this method can effectively improve the construction accuracy and reduce the construction time, and the fitted driving conditions can effectively reflect the local actual traffic conditions.**

Keywords-Driving Cycle; Principal Component Analysis; Improved K-Means Algorithm

I. INTRODUCTION

Vehicle Driving Cycle (Driving Cycle), also known as vehicle test cycle, which describes the speed-time curve of the vehicle. It is an important general-purpose basic technology in the automotive industry, which is used to limit pollutant emissions and energy consumption, and to apply evaluation technology [1-2] for new vehicle development. In the early days, developed countries such as the United States, Japan, Europe, etc., all established vehicle driving conditions based on their cities. For example, Lin et al. [3] analyzed and compared the driving conditions constructed in California through a stochastic process; Michel et al. [4]. studied that different vehicles should be represented by different driving conditions. In recent years, domestic scholars have conducted in-depth research on the driving conditions of automobiles and have achieved some research results. For example: Gao Jianping [5], Peng Yuhui [6] and others use the K-means algorithm respectively to construct the driving conditions of passenger cars in Zhengzhou and Xiamen, but the K-means algorithm is easy to fall into a local optimal solution, which affects the stability and accuracy of the construction

conditions; Hefei university of technology, represented by professor Shi Qin [7] team, led by typical road in hefei, for example, was carried out on the road data acquisition, processing, and using the particle swarm optimization method of fuzzy clustering [8], although can avoid local optimization to some extent, but for large data, there will be a problem of poor local search ability; Cao Qian et al. [9] constructed the driving conditions of Changchun city based on Markov chain, but when the data amount was small and the error of calculating the state matrix was large, the accuracy of the final construction conditions would be resulted.

Up to now, domestic scholars generally adopt the k-means algorithm to construct their own cities. The common problem is that the k-means algorithm is sensitive to outliers, the parameter K (cluster number) is difficult to determine, and it is easy to fall into the local optimal solution. Although some other algorithms have appeared, in the case of large amounts of data, the algorithm is inefficient in execution and cannot efficiently classify large quantities of data. Therefore, this paper proposes the Grid-K-means algorithm to achieve the clustering classification effect, and the clustering effect is stable, accurate and efficient, and the constructed working condition can comprehensively reflect the local traffic situation.

II. DATA PREPROCESSING

The data collected in this paper includes vehicle running time, speed, latitude and longitude and other information, sampling frequency of 1Hz, and data is collected on different roads. The original data is denoised [10-11] by discrete Fourier transform, the steps are as follows:

Step1: set $f(x)$ as the original signal, and transform the time domain signal to the frequency domain through Fourier transform:

$$F(u)=\sum_{x=0}^{N-1} f(x)e^{-j2\pi ux/N} \tag{1}$$

In the above formula: $F(u)$ is the frequency domain signal, x and u are the discrete signal time domain variables and frequency domain variables respectively, and N is the number of discrete signals.

Step2: adjust the value of the high frequency signal to 0 or a smaller value, while the value of the low frequency signal remains unchanged

Step3: to achieve signal smoothing and noise reduction, the frequency domain is restored to the time domain by the inverse Fourier transform, the expression is as follows:

$$f(x)=\frac{1}{N}\sum_{u=0}^{N-1} F(u)e^{j2\pi ux/N} \tag{2}$$

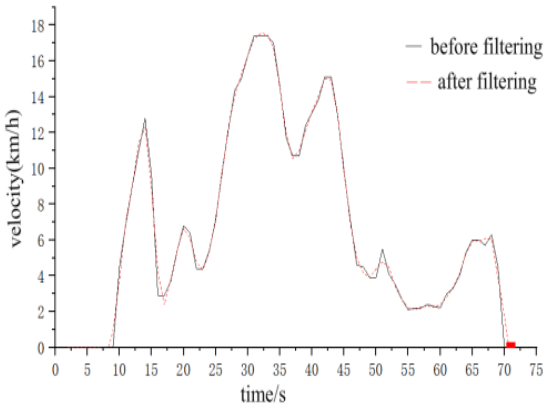


Figure 1. Contrast analysis of speed data filtering

From Figure 1, we can see that the discrete Fourier transform has a good filtering effect on the data, which can effectively eliminate the influence of noise on the data, thus avoiding the distortion in the calculated acceleration.

III. PRINCIPAL COMPONENT ANALYSIS AND IMPROVED K-MEANS CLUSTERING ALGORITHM

A. *Kinematics segment division and feature parameter extraction*

Kinematics segment refers to the vehicle speed range from the beginning of the idle state to the beginning of the next idle state [12]. As shown in Figure 2 below:

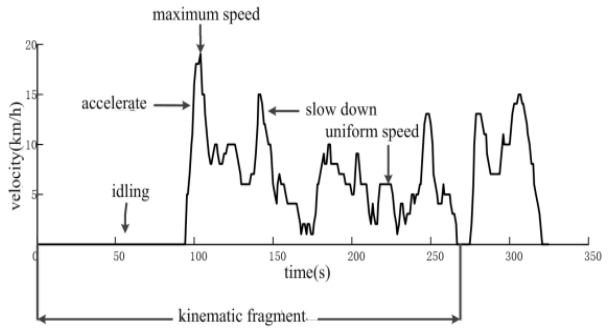


Figure 2. Schematic diagram of kinematics fragments

According to the domestic and foreign research sports state classification standards, the acceleration threshold can be selected as 0.1m/s2 to define four driving states:

1) Idling state: the engine continues to work, and its speed is zero during this process;

2) Deceleration state: the speed is not zero, and the acceleration during this continuous operation is less than or equal to -0.1m/s2;

3) Acceleration state: the speed is not zero, and the acceleration is greater than or equal to 0.1m/s2 during this continuous operation;

4) Constant speed state: the speed is not zero, and the absolute value of the acceleration during this continuous operation is less than or equal to 0.1m/s.

In order to fully and systematically reflect the driving characteristics of each feature and each operational segment, according to relevant literature[5-9], this paper selects 12 feature parameters to reflect each short-stroke feature:(segment duration) T/t ,(Driving distance) S/km ,(Average speed) $V_a/(km\bullet h^{-1})$,(Average driving speed) $V_x/(km\bullet h^{-1})$, (Idle time ratio) $T_i/\%$, (Acceleration time ratio) $T_a/\%$, (Deceleration time ratio) $T_d/\%$, (Cruise time ratio) $T_c/\%$,(Mean acceleration) $a_a/(m\bullet s^{-2})$, (Average deceleration) $a_d/(m\bullet s^{-2})$, (Standard deviation of acceleration) $a_{std}/(m\bullet s^{-2})$, (Standard deviation of velocity) $V_{std}/(km\bullet h^{-1})$.

Using Matlab to preprocess the 185726 sets of data to obtain 164089 sets of data, use R language programming to obtain 1585 kinematics segments, and then calculate the characteristic parameters of each kinematics segment respectively to obtain the number of samples (rows) x and characteristic parameters (columns), as shown in Table 1 below.

TABLE I. KINEMATICS FEATURE PARAMETER VALUES

Fragment number	T	S	V_a	V_x	T_i	T_a	T_d	T_c	a_a	a_d	a_{std}	V_{std}
1	119	203.29	6.15	7.60	0.06	0.17	0.08	0.03	0.42	-0.66	0.48	5.05
2	319	2320.7	26.19	33.63	0.14	0.26	0.19	0.11	0.30	-0.34	0.37	17.5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
78	243	346	7.03	12.23	0.04	0.21	0.08	0.04	0.38	-0.61	0.39	5.01
79	167	459.12	5.97	8.04	0.06	0.19	0.05	0.05	0.44	-0.59	0.28	4.91
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1584	169	829.04	17.66	23.65	0.13	0.19	0.09	0.07	0.36	-0.50	0.48	14.8
1585	486	6799.9	50.37	52.65	0.02	0.18	0.13	0.30	0.19	-0.20	0.29	16.5

B. Principal component analysis

There are 1585 kinematics segments, each with 12 parameters, forming a 1585*12 matrix. The classification of 12-dimensional data requires a lot of work. Too much data will make the subsequent calculation steps more complicated, and some parameters are not independent of each other. Therefore, it is necessary to use principal component analysis to reduce the dimension of the data, which has reached a small number of principal components. It can reflect many original feature parameters.

This article uses SPSS for analysis based on the preprocessed data of the data set. The contribution rate and cumulative contribution rate of each principal component are shown in Table 2 below.

TABLE II. PRINCIPAL COMPONENT CONTRIBUTION RATE AND CUMULATIVE CONTRIBUTION RATE

Serial number	eigenvalue	contribution/%	Cumulative contribution/%
1	5.5605	43.23	43.23
2	3.2315	26.83	70.06
3	1.9901	16.14	86.20
4	1.0434	6.14	92.34
5	0.53476	3.11	95.45
6	0.45796	2.01	97.46
7	0.38341	1.22	98.68
8	0.21525	0.78	99.46
9	0.17450	0.23	99.69
10	0.10126	0.19	99.88
11	0.09324	0.08	99.96
12	0.02532	0.04	100

It can be seen from Table 2 that the cumulative contribution rate of the first three principal components has reached 86.2%, and the eigenvalues of these three principal components are all greater than 1. According to statistical theory [13], in principal component analysis, in general, it is sufficient to select principal components with a cumulative contribution rate of 85%, which basically represents all the information of the 12 characteristic parameters of the fragment, so the first three principal components can be selected for cluster analysis.

Through SPSS software analysis, the load matrix of each principal component can be obtained, so the correlation coefficients of the first 3 principal components and 12 characteristic parameters can be determined. As shown in Table 3 below.

TABLE III. PRINCIPAL COMPONENT LOAD MATRIX

Characteristic parameters	M_1	M_2	M_3
Segment duration T	0.132	0.231	0.745
distance S	0.293	0.134	0.045
Average speed V_a	0.719	0.463	-0.025
Average driving speed V_x	0.478	0.615	0.112
Idle time ratio T_i	0.125	-0.351	0.843
Acceleration time ratio T_a	0.694	-0.156	0.060
Deceleration time ratio T_d	0.923	0.341	-0.123
Cruising time ratio T_c	0.641	0.435	-0.045
Mean acceleration a_a	0.014	0.623	0.033
Mean deceleration a_d	0.366	-0.433	-0.052
Standard deviation of acceleration a_{std}	0.445	0.267	-0.067
Standard deviation of speed V_{std}	0.387	0.215	0.034

The principal component loading represents the correlation coefficient between the principal component and the original variable. The larger the correlation coefficient (absolute value), the more representative the principal component of the variable. According to the principal component loading matrix in the above table, some parameters corresponding to a certain principal component can be obtained. As shown in Table 4 below.

TABLE IV. THE MAIN EIGENVALUES REPRESENTED BY THE FIRST THREE PRINCIPAL COMPONENTS

Category	Eigenvalue
Primary principal component M_1	average speed, acceleration time ratio, deceleration time ratio, cruising time ratio
Secondary principal component M_2	average speed, acceleration time ratio, deceleration time ratio, cruising time ratio
Third principal component M_3	segment duration, idle time ratio

According to principal component analysis, the first, second and third principal components were selected for further analysis.

C. Improved K-means clustering algorithm

For traditional clustering algorithm clustering, the initial clustering center has a greater impact on the clustering results, which is easy to reduce the accuracy of the algorithm, unable to handle noisy data, and the clustering results are unstable and relatively inefficient when processing large-scale data sets [14-15]. This paper proposes a Grid-K-means clustering algorithm based on Grid improvement, which introduces the idea of grid division into K-means algorithm, reduces the number of manually set initial parameters, speeds up the ability to process data, and reduces the time complexity. This improves the overall efficiency of the algorithm. The main steps of the algorithm are as follows:

Step1. Let the data set be S, and each sample point $x_i = \{x_{i1}, x_{i2}, \ldots x_{im}\}$ has m attributes (the data set is S has m dimensions)

Step2. Read all the sample points in the data set, and record the maximum and minimum values of each dimension $E_{xi} = \{x_{i\max}, x_{i\min}\}$

Step3. Calculate the data range of each dimension, $Range_i = x_{i\max} - x_{i\min}$, $i = 1,2,\ldots n$

Step4. set the grid number of each dimension (GridNum), $GridNum = k \times 9$,among them, $GridNum \le \sqrt{n/2}$

Step5. Determine the grid step size of each dimension in the data set: $Steps_i = Range_i / Gridnum$, $i = 1,2,\ldots n$

Step6. Convert the data set into grid points, $Grid_{ij} = \frac{x_{ij} - x_{i\min}}{Steps_i}$ $i = 1,2,\ldots n;\ j = 1,2,\ldots m$

Step7. set the dynamic threshold, if the grid density is less than the threshold, it is noise, and if the grid density is equal to zero, it is invalid grid points $Threshold = n/(3 \times GridNum^n)$

Step8. Sort the grid points, select K high-density grid points as the initial cluster centers ($Center_1$), and select other k-1 high-density grid points outside the interference radius R as the remaining initial cluster centers ($Center_2, Center_3 \ldots, Center_k$), The initial cluster center interference radius $R = a \times (GridNum/(k+1))$, among them $a \in [1,2]$.

Step9. Iteratively and circularly calculating the grid clustering center until the grid clustering center no longer changes ; $Center_k' = (\sum Grid_{ij} \times Grid_{i0}) / \sum Grid_{i0}$

Step10. Reversely calculating the corresponding cluster center of sample points $v = v_1, v_2, \ldots v_k$, the formula is as follows: $x_{ij} = Grid_{ij} \times Steps_i + x_{i\min}$

In order to verify the performance of the improved algorithm Grid-K-means proposed in this article, simulation experiments select 3 sets of data (20,000, 80,000, and 100,000) for verification, and the accuracy of clustering results and the performance of different algorithms are compared respectively, as shown in the following table 5 below.

TABLE V. TABLE 5 COMPARISON OF ACCURACY AND PERFORMANCE OF DIFFERENT ALGORITHMS

Data set	k-means		k-means++		Grid-K-means	
	Accuracy	Elapsed time	Accuracy	Elapsed time	Accuracy	Elapsed time
20000	92.76%	25.82s	93.16%	34.86s	95.78%	25.83s
80000	88.32%	43.83s	89.32%	49.16s	92.36%	40.69s
100000	87.24%	58.12s	87.65%	67.94s	90.12%	50.38s

It can be seen that the performance improvement of grid-K-means algorithm is not significant when dealing with small data sets; but as the amount of data increases, the accuracy of the clustering results and the clustering speed of Grid-K-means are both Far beyond k-means, and k-means++ algorithm.

The kinematics fragments are divided into 3 categories by using the improved K-means algorithm. Table 6. shows the average eigenvalues of each category after clustering. As can be seen from Table 6, the first category is traffic congestion, because the average speed and average driving speed are low and the idle time ratio is high; The second type is smooth road conditions, because the average speed and the average driving speed are higher, the idle time ratio is in the middle, and the cruising time ratio is in the middle; The third category is high-speed road conditions, because the average speed and average driving speed are the highest, idle time ratio is the lowest.

TABLE VI. AVERAGE FEATURE VALUE OF EACH CATEGORY AFTER CLUSTERING

Clustering categories	v_a	T_c	v_x	a_a	T	T_i
Class 1	6.23	0.12	11.06	0.16	26.15	0.59
Class 2	16.25	0.25	18.26	0.59	35.56	0.23
Class 3	23.03	0.49	30.51	0.45	48.12	0.18

IV. CONSTRUCTION AND COMPARATIVE ANALYSIS OF DRIVING CONDITIONS

A. Construction of working conditions

According to the construction of the city, the duration of the working condition is about 1200s. According to the time curve, the kinematic segments are selected from the three categories respectively to be 56,89,225, forming the final representative working condition. Fitting the representative working condition time-speed graph and working condition time-acceleration graph, as shown in Figure 3 and Figure 4.

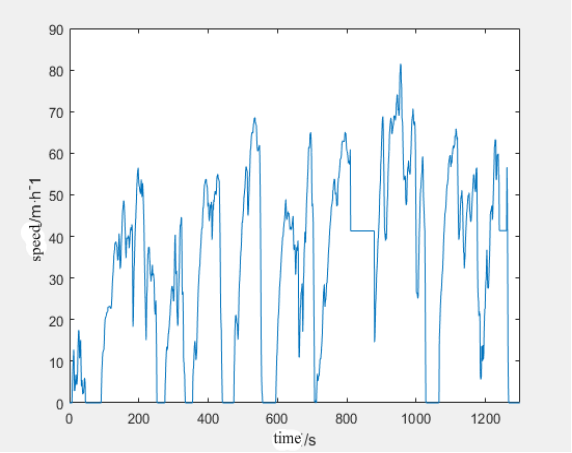


Figure 3. Time-speed curve of representative working conditions

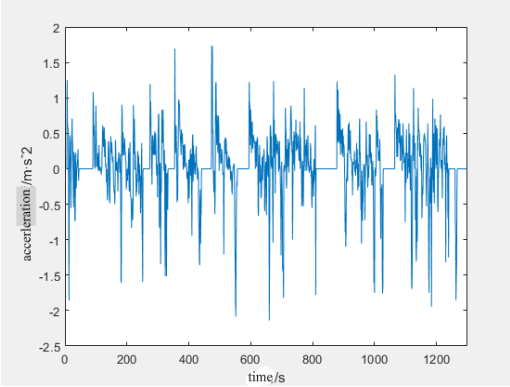


Figure 4. Time-acceleration curve of representative working conditions

B. Verification of Working Conditions

1) Error Analysis and Verification based on characteristic parameters

Comparing the characteristic parameters of the collected overall sample data with the characteristic parameters calculated based on traditional k-means, k-means++, and Grid-K-means, as shown in Table 7 below. Among them, E_s is the average relative error between the fitting representative working condition and measured values of all characteristic parameters of the test [16].

TABLE VII. CHARACTERISTIC PARAMETER RELATIVE ERROR

Characteristic parameters	Overall sample data	k-means		k-means++		Grid-K-means	
		Fitted value	Relative error/%	Fitted value	Relative error/%	Fitted value	Relative error/%
v_a	18.91	18.11	4.23	18.34	3.01	18.60	1.64
a_a	1.85	1.95	5.41	1.88	1.62	1.86	0.54
a_d	-2.36	-2.45	3.81	-2.43	2.97	-2.39	1.27
T_i	0.52	0.58	11.54	0.55	5.77	0.53	1.92
T_a	0.49	0.41	16.32	0.44	10.20	0.47	4.08
T_d	0.26	0.31	19.23	0.29	11.54	0.28	7.69
E_s			10.09		5.85		2.86

The average relative error of k-means is 10.09%, the average relative error of k-means++ is 5.85%, and the average relative error of all characteristic parameters of driving conditions constructed based on the improved k-means algorithm (Grid-K-means) is only 2.86%, and the relative error of the other eigenvalue parameters is less than 8%, so it shows that the Grid-K-means method can more accurately represent the driving conditions than the traditional method.

2) K-S test based on independent sample acceleration distribution

The driving acceleration of the whole sample is divided into three categories [17], （-∞,-0.64） m/s2，[-0.64,0.64] m/s2,（0.64,+∞）m/s2 ,and the sample speed of each type of acceleration is divided into six categories,[0,10） m/s，[10,20） m/s，[20,30） m/s，[30,40） m/s，[40,50） m/s，[50, +∞）m/s. In order to further test the similarity between acceleration distribution of representative working conditions and test data, K-S test of independent samples is carried out, and the test results are shown in Table 8 below. From this, we found that the acceleration distribution of the improved clustering algorithm Grid-K-means fitting working conditions is better than that of k-means++, which can better reflect the distribution frequency of acceleration.

TABLE VIII. SAMPLE K-S TEST

Method		Acceleration distribution（m/s2）		
		（-∞,-0.64）	[-0.64,0.64]	（0.64,+∞）
Grid-K-means	K-S value	0.68	0.57	0.53
	Similarity level	0.89	0.96	0.99
k-means++	K-S value	0.71	0.26	0.45
	Similarity level	0.79	0.44	0.82

V. CONCLUSION

In order to effectively construct vehicle driving conditions based on the city's own data, this paper uses discrete Fourier transform to preprocess the data, eliminate the noise in the data, and use principal component analysis to reduce the dimension of the data, and propose an improved k-means (Grid-K-means) optimization method to cluster the data, replace the initial parameter setting with the idea of dynamic grid division to improve clustering efficiency, speed up data processing, and reduce time complexity. Through experimental analysis: the average relative error between the driving conditions and the characteristic parameters of the experimental data is only 2.86%. The K-S test is used to verify and analyze the similarity between the fitting conditions and the experimental data. The results show that the constructed driving conditions are more accurate and more accurate and can better reflect the overall driving characteristics.

ACKNOWLEDGMENT

This research is partially funded by the Project funds in Shaanxi province University Student Innovation and Entrepreneurship Fund Project (S202210702077).

REFERENCE

[1] Li Mengliang, Li Wei, Fang Maodong, Liu Xiangmin, Du Chuanjin, Zhang Xingquan. The Parse Method of Actual Driving Cycle of Vehicle on Road [J]. Journal of Wuhan University of Technology (Transportation Science and Engineering Edition), 2003(01):69-72.

[2] Gao Jianping, Sun Zhongbo, Ding Wei, Xi Jianguo. Development and precision research of vehicle driving conditions [J]. Journal of Zhejiang University (Engineering Edition), 2017,51(10):2046-2054.

[3] Jie Lin, Debbie A Niemeier. An exploratory analysis comparing a stochastic driving cycle to California's regulatory cycle [J]. Atmospheric Environment, 2002, 36(38).

[4] Michel Andr é. The ARTEMIS European driving cycles for measuring car pollutant emissions [J]. Science of the Total Environment, 2004, 334.

[5] Gao Jianping, Ding Wei, Sun Zhongbo, Xi Jianguo. Construction of passenger car driving conditions in Zhengzhou [J]. Mechanical design and manufacturing, 2018(08):102-105.

[6] Peng Yuhui, Yang Huibao, LI Mengliang, Qiao Xueqi. Research on Construction Method of Urban Road Vehicle Driving Conditions Based on K-means Cluster Analysis [J]. Automotive Technology, 2017(11):13-18.

[7] Zhang Rui. Construction and Research of Urban Road Car Driving Conditions [D]. Hefei: Hefei University of Technology, 2009

[8] Shi Qin, Wang Nannan, Chou Duoyang. Application of Particle Swarm Optimization Fuzzy Clustering Method in Vehicle Driving Conditions [J]. Chinese Management Science, 2011, 19(02):110-115

[9] Cao Qian, Li Jun, Liu Yu, Qu Dawei. Construction of passenger car driving conditions based on Markov chain[J]. Journal of Jilin University (Engineering Edition), 2018, 48(05):1366-1373.

[10]Cao Qian, Li Jun, Liu Yu, Qu Dawei. Construction of driving conditions based on big data and Markov chain [J]. Journal of Northeastern University (Natural Science Edition), 2019, 40(01):77-81.

[11]DingYifeng, LI Jun,Gai Hongchao,Chen Hao. The Application of Wavelet Transform in Processing Vehicle Speed Data Processing [J]. Science Technology and Engineering, 2017, 17(28):274-279.

[12]Miao Qiang, Sun Qiang, Bai Shuzhan, Yan Wei, Li Guoxiang. The Application of Wavelet Transform in Processing Vehicle Speed Data Processing[J]. Journal of China Highway, 2016, 29(11):161-169.

[13]Fan Jincheng, MEI Canglin. data analysis. Beijing: Science Press, 2010.

[14]Xu, Tian-Shi, Chiang, Hsiao-Dong,Liu, Guang-Yi, et al.Hierarchical K-means Method for Clustering Large-Scale Advanced Metering Infrastructure Data [J]. IEEE Transactions on Power Delivery, 2017, 32(2):609-616.

[15]Wen Peng. Research on Improved K-means Algorithm and New Clustering Effectiveness Index in Cluster Analysis [D]. Anhui University, 2019.

[16]Shi Qin, Chou Duoyang, Wu Jing. Research on driving conditions based on principal component analysis and FCM clustering [J]. Environmental Science Research, 2012, 25(01):70-76.

[17]Jiang Ping, Shi Qin, Chen Wuwei. Construction method of urban road driving conditions based on Markov [J]. Journal of Agricultural Machinery, 2009, 40(11):26-30.

Research on Pilots ' Mental Workload Classification in Simulated Flight

Jinna Xue*
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail:xgenia@163.com
*Corresponding author

Changyuan Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: Cyw901@163.com

***Abstract*—The problem of human-computer interaction mental workload in flight driving has great reference value for the prevention of safety hazards in aviation driving. This paper analyzes and studies the classification method of mental workload in flight driving by designing different simulated flight experiment tasks. This study uses a combination of EEG signals and subjective evaluation, through the use of convolutional neural networks and long short-term memory network method of combining EEG signals for research and analysis. The accuracy of EEG signal classification is as high as 94.9 %. NASA-TLX evaluation results show that there is a positive correlation between task load difficulty and evaluation score. The results show that the combination of convolutional neural network and long short-term memory network is suitable for pilots ' mental workload classification. This study has important practical significance for flight accidents caused by pilots ' mental workload.**

Keywords-Mental Workload; EEG; Convolutional Neural Network; Long Short-Term Memory Network; Subjective Evaluation Method

I. INTRODUCTION

With the emergence and rapid development of the Internet, people's load on cognitive resources has also increased, and the demand for mental workload optimization has become increasingly prominent [1]. Studies have shown that too high and too low cognitive load will have a negative impact on job performance [2]. Pilots need to obtain more instrument data information during driving, and combine it with the working environment for comprehensive analysis, so as to make corresponding decisions quickly and accurately [3]. In the relevant operation process requires a high degree of concentration, may produce large mental labor [4], will not only affect the work efficiency, and even cause serious accidents. Thus, in human-computer interaction, mental labor gradually occupy the advantage. The intensity of mental labor can be measured by mental workload, which is mainly generated in projects where operators use limited mental resources for task processing [5]. Appropriate mental workload helps the operator to maintain good task performance, while too high or too low mental workload will adversely affect the performance in the operation.

According to the aviation accident research report, 60 % to 90 % of flight accidents occur in flights where the pilot 's mental workload intensity is too high and the stress task level is high [6]. Complex human-machine interaction operations such as air defense missiles, medical rescue and efficient driving are often accompanied by high-load human-machine interaction tasks [7]. For example, in the operation of missile weapons, it is necessary to achieve accurate and fast analysis of space intelligence such as the course and speed of combat targets, and to quickly realize the assessment of the threat of combat targets for the implementation of complex actions such as target tracking and attack [8-10]. Because there is a large degree of human-computer interaction tasks in this process. The high-intensity operation process can easily cause the mental operator to fall into an overload state, which greatly reduces the safety and reliability of the operator 's operation process and easily causes

safety accidents. Therefore, the study of mental workload is of great significance to the safety and reliability of human-computer interaction.

The concept of mental workload was first proposed in the 1980 s. Scholars have mainly conducted in-depth research and discussion on the causes, internal mechanisms, and measurement methods of mental workload [11]. In the past forty years, researchers have continued to pay attention to the mental workload of operators. The physiological measurement method has greater advantages in terms of sensitivity to changes in brain load and digital quantitative analysis, but there are greater difficulties in terms of post-processing of data and other aspects, a certain knowledge base is required for data analysis, and the highly sensitive measurement method is susceptible to interference from external factors. In response to these difficulties, some scholars have proposed the use of eye-movement metrics for the assessment of brain load [12-14]. Liu Yan [15] et al. conducted a correlation analysis of pilots' brain load in terms of typical issues such as fatigue and sleep, and obtained from the experimental results that the correlation degree coefficient between brain load and multidimensional fatigue measurements was around 0.5, while the correlation degree coefficient between brain load and Pittsburgh sleep quality was as high as 0.59. From the results of this study, it can be obtained that the degree of pilots' brain load is influenced by the degree of fatigue and sleep. Wang Lei et al. from Civil Aviation University of China [16] conducted a study on the brain load characteristics of pilots based on flight task context routes, in which they used different task difficulties to carry out relevant studies on the brain load characteristics of pilots. The experimental results are of great reference value for the study of pilot brain load classification methods. In this paper, we will draw on the relevant pilot brain load characteristics to investigate pilot brain load classification methods.

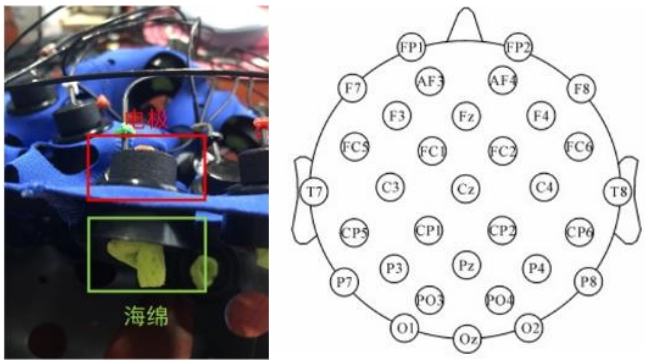
II. EXPERIMENT

A. Experimental preparation

The experiment was conducted under a simulated flight platform and eight male graduate student volunteers aged 22-30 years were recruited as subjects for data collection. The subjects were all in good health, had normal or corrected vision, were right-handed and had good sleep conditions. The subjects were familiarised with the details of the experimental operation and related precautions before the simulation. When conducting the experiments, only the subjects were left alone for each trial in order to avoid interference from the outside environment.

The equipment for this pilot mental workload classification study contains a DELL computer, three high definition displays (2560 x 1440 resolution), EEG signal synchronisation acquisition equipment and the test deployment platform is shown in Figure 1. The flight simulation platform contains a monitor, flight joystick and mainframe. The flight simulation was conducted using the DCS World digital battlefield game, using a free to fly Su-25T fighter aircraft designed and manufactured by Sukhoi in Russia.

Figure 1. Simulated flight environment experimental platform



(a) EEG signal acquisition device (b) EEG cap port diagram

Figure 2. ErgoLAB EEG device

The experimental equipment used in this experiment was the ErgoLAB human-computer environment synchronous platform EEG measurement system provided by Beijing Jinfa Technology Co. The system adopts the international standard 10-20 standard lead semi-dry electrohydraulic stage with an output bit 32-lead EEG signal; a wet sponge is used as the conduction medium, as shown in Figure 2(a).

B. Experiment design

In this simulated flight experiment task, each subject performed three main simulated flight experiments containing three mental-load flight tasks: control, low load and high load. The cockpit dial information that needed to be monitored during the simulated flight corresponding to the different difficulty tasks during the simulated flight was shown in Table 1. The flight information contained in the cockpit interface [17] includes indication of airspeed table, pitch scale information, no altimeter information, heading angle, roll angle, direction pod, landing gear status and engine status. During a flight task, the subject is required to respond accurately and quickly to the information presented in the interface.

TABLE I. INFORMATION ON THE DIALS TO BE MONITORED FOR DIFFERENT DIFFICULTY TASKS

High loads	Low loads	Control group
Indicated airspeed meter	Indicated airspeed meter	—
Pitching scale information	Pitching scale information	—
Altimeter information	Altimeter information	—
Directional angle information	—	—
Rolling corners information	—	—
Status of the steering compartment	—	—
Landing gear conditions	—	—
Engine status	—	—

C. The experimental process

For different subjects, the experiments are conducted one by one. Each subject is required to perform a control, low and high load flight for each experiment. In the same flight environment, the operator is prompted with different difficulty tasks based on the flight heading status and is required to quickly and accurately monitor the corresponding dial information.

The NASA-TLX Subjective Rating Scale [18] was used in the experiment. The scale of [0,100] was used to indicate the scale's score range to rate the subject's level of mental workload on six dimensions, with higher scores indicating higher levels of workload.

D. Data recording

During the experiment, simultaneous acquisition of EEG signals was performed using EEG acquisition equipment. At the end of each set of simulated flight experiments, subjects are required to complete the NASA-TLX measurement form, which is used to evaluate the subjective load of the subjects.

III. METHODS

In this study, a modified long and short term memory network was used to classify the EEG

signals for mental workload research, and the flight mental workload was studied and analysed by combining subjective evaluation methods.

A. *Data pre-processing*

During the experimental process, the hardware equipment will inevitably be disturbed by external factors. Therefore, the raw EEG signal needs to be pre-processed before the mental workload analysis is carried out, and the processing flow is shown in Figure 3.

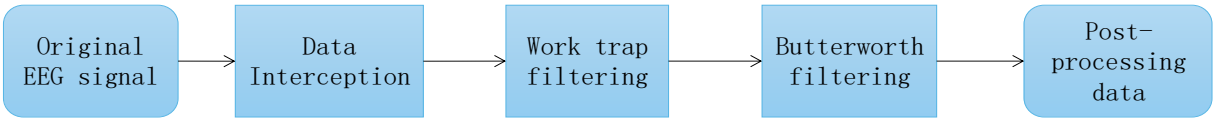


Figure 3. Figure. 3 Flow chart of data processing

Related studies have shown that most EEG signals are concentrated between the range of 0.05-50Hz, while event-related desynchronisation and event-related synchronisation (ERD/ERS) patterns are mainly manifested in μ rhythms (8-13Hz) and β beats (14-30Hz), so this study used Butterworth filters for 50Hz IDF trapping and 0.05-30Hz bandpass filtering, and Butter The Butterworth band-pass filtering removed the fine burrs of interference from the original data [19]. A comparison of the data before and after processing is shown in Figure 4.

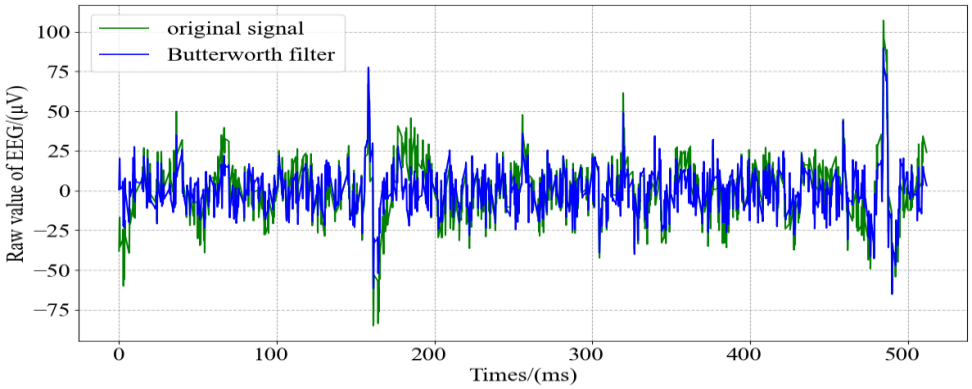


Figure 4. Comparison of EEG waveforms before and after data processing

B. *EEG features extraction*

According to recent research in cognition, psychology and psychiatry, EEG information about mental activity can be divided into five bands: delta (δ), theta (θ), alpha (α), beta (β) and gamma (γ), and the frequency range of each band is shown in Table 2.

TABLE II. TABLE OF FREQUENCY BANDS OF EEG SIGNALS

Frequency band range(Hz)	Band names
1-4	delta (δ)
4-8	theta (θ)
8-13	alpha (α)
13-30	beta (β)
>30	gamma (γ)

In this paper, the discrete Fourier transform is used for EEG signal feature extraction [20]. The power spectrum was estimated mainly using the Welch method with the power spectral densities of the σ, θ, α and β frequency bands, using a Hamming window with a window size of 256 and a 50% overlap of adjacent window segments. In calculating the power of different frequency bands, the sum of the power density of the band, i.e. the area within the corresponding frequency band of the corresponding power density curve, can be calculated as shown in equation (1):

$$\begin{cases} P_\delta = \int_1^4 \mathrm{psd}(f)df \\ P_\theta = \int_4^8 psd(f)df \\ P_\alpha = \int_8^{13} psd(f)df \\ P_\beta = \int_{13}^{30} psd(f)df \end{cases} \quad (1)$$

Where, $psd(f)$ denotes the power spectral density function and f indicates the frequency of the signal. This paper uses the power under the four fundamental frequency bands spectral densities for mental workload classification judgement.

C. Network model

The network model proposed in this paper is shown in Figure 5. The model is a new input convolutional neural network in front of the long and short-term memory network for local feature extraction of the EEG signal. The extracted local features are then fed into the long and short-term memory network for training. The loss function for model training is L1-loss and its expression is:

$$L1-loss = \sum_{i=1}^{n} \left| x_{true} - x_{predicted} \right| \quad (2)$$

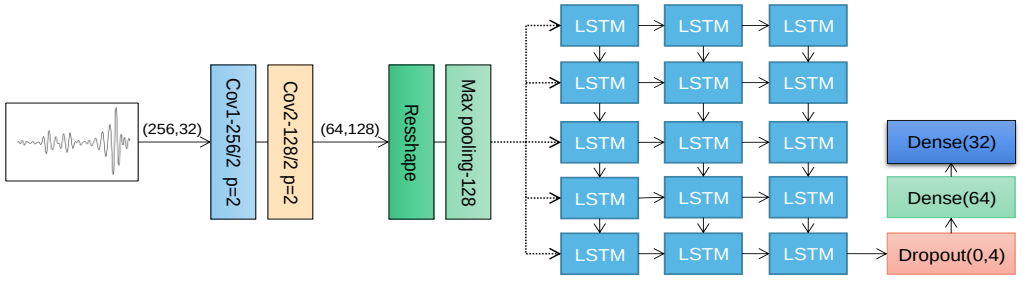


Figure 5. Network model diagram

D. LSTM networks

The Long Short-Term Memory (LSTM) is a recurrent neural network (RNN) [21]. The LSTM contains a three-part structure of input gates, forgetting gates and output gates, which is shown in Figure 6 and mainly outputs (0,1) through the sigmoid activation function values between.

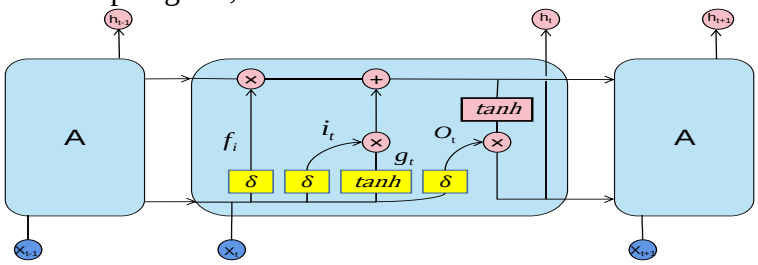


Figure 6. LSTM structure diagram

The role of the forgetting gate is to control the discard rate of the delivered information, as implemented in equation (3).

$$f_t = \sigma(W_f^T \times s_{t-1} + U_f^T \times x_t + b_f) \quad (3)$$

Where δ indicates the activation function, usually set to a sigmoid function, w_f^T indicates the forgetting gate weight matrix, U_f^T is the weight matrix between the input and hidden layers of the forgetting gate, and b_f indicates the bias of the forgetting gate, s_{t-1} is the previous moment's output value and x_t is the current moment's input value. The closer the forgetting gate output value is to 1, the more information is retained, and the closer it is to 0, the less information is retained.

The input gate determines at what percentage of the current moment's input information is fed

into the memory information stream, and is calculated similarly to the forgetting gate, as shown in equation (4).

$$i_t=\sigma(W_i^T \times s_{t-1}+U_i^T \times x_t+b_i) \quad (4)$$

The output gate is mainly used to control the amount of information updated by the next layer of the network and is implemented as shown in equation (5).

$$O_t=\sigma(W_O^T \times s_{t-1}+U_o^T \times x_t+b_o) \quad (5)$$

$$s_t=O_t \times \tanh(C_t) \quad (6)$$

σ is usually a sigmoid function. This structure expands the effectiveness of the action of the current amount of information so that it can both suppress the current information and output it normally through equation (6) combined with the memory information to obtain the output value at the current moment.

IV. Results and Analysis

Based on the results of the NASA-TLX subjective evaluation table obtained statistically, the sensitive assessment index scores for changes in mission difficulty can be obtained as shown in Figure 7. According to the ANOVA results, the NASA-TLX subjective evaluation scores show an extremely significant difference ($p<0.01$) and become larger as the difficulty of the mission increases and the load level increases.

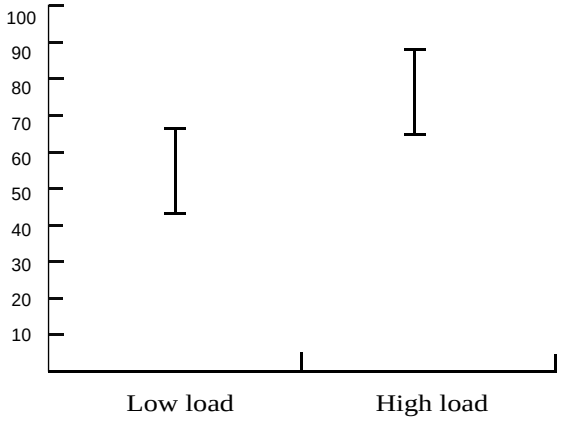


Figure 7. NASA-TLX evaluation results

Based on the results of the NASA-TLX subjective evaluation table obtained statistically, the sensitive assessment index scores for changes in mission difficulty can be obtained as shown in Figure 7. According to the ANOVA results, the NASA-TLX subjective evaluation scores show an extremely significant difference ($p<0.01$) and become larger as the difficulty of the mission increases and the load level increases.

TABLE III. MODEL TRAINING PARAMETERS

Name of the layer	Hyperparameters	Input vector	Output vector
EEG_Input	None	(256,1)	(256,1)
Conv1D	Kernel size=5 stride = 2	(256,1)	(128,64)
Conv2D	Kernel size=3 stride = 2	(128,64)	(64,128)
Reshape	None	(64,128)	(8192,1)
Max Pooling	pool size=128	(8192,1)	(64,1)
LSTM	units=32	(64,1)	(32,1)
Dense	units=64	(32,1)	(64,1)
Dense	units=32	(64,1)	(32,1)

This paper presents a study related to the use of a modified long and short term memory network for mental workload classification methods. In conducting the model training process, the pre-processed data was used as the input data for the model, where 80% of the data was used to conduct the model training and 20% of the data was used to conduct the test. Details of the parameter settings of the model are shown in Table 3.

The number of times the model was trained was set to 500 and the learning rate was set to 0.001. The experimental results showed that the training accuracy could basically reach over 90% around the 120th epoch. In the first 20 epochs and 70~80 epochs, the discriminant accuracy improved faster, and the experimental result accuracy is shown in Figure 8; the loss value converged more rapidly, and the experimental loss rate is shown in Figure 9. In this model training, the final accuracy rate can reach 94.9%. Compared with other traditional methods, this experimental model has greater accuracy.

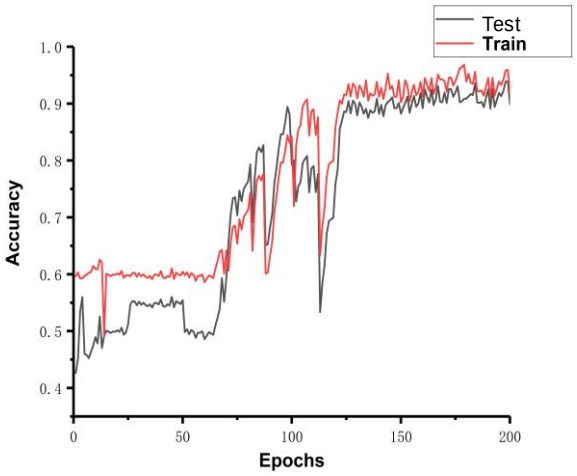


Figure 8. Training and test set accuracy

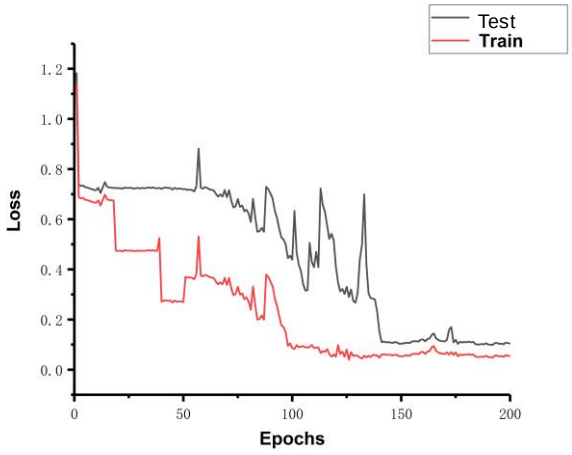


Figure 9. Loss rate for training and test sets

V. Conclusions

Piloting in aviation is a major hazard and pilots are under great stress during the flight. The study of the pilot's mental workload is conducive to a better assessment of the pilot's driving state, as well as to the effective prediction of phenomena that exceed the mental workload, which is conducive to the rational design of the human-computer interaction system in flight driving, thus reducing the phenomenon of high or low mental workload during flight driving and ensuring the safety of flight driving.

This paper uses a combination of convolutional neural networks and long and short-term memory networks to classify mental workload with an accuracy of 94.9%. The results of the analysis show that there is a positive correlation between the difficulty of the flight task and the subjective evaluation score. In summary, this study has some reference value for the classification of pilot mental workload.

Acknowledgment

This work is supported by The National Natural Science Foundation of China (No. 52072293)

References

[1] Previtali Federica, Picco Eleonora, Gragnano Andrea, Miglioretti Massimo. The Relationship between Work, Health and Job Performance for a Sustainable Working Life: A Case Study on Older Manual Employees in an Italian Steel Factory [J]. International Journal of Environmental Research and Public Health, 2022,19(21).

[2] Mohanavelu K, Poonguzhali S, Adalarasu K, Ravi D, Vijayakumar Chinnadurai, Vinutha S, Ramachandran K,Srinivasan Jayaraman. Dynamic cognitive workload assessment for fighter pilots in simulated fighter aircraft environment using EEG [J]. Biomedical Signal Processing and Control, 2020, 61.

[3] Sun Li, Sun Youzhao. Dynamic functional allocation of aircraft cockpit based on load balancing[J]. Measurement and Control Technology, 2022, 41(01):11-15+27.DOI:10.19708/j.ckjs.2021.04.230.

[4] Zhang P, Wang X, Zhang W, et al. Learning spatial–spectral–temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment [J]. IEEE Transactions on neural systems and rehabilitation engineering, 2018, 27(1): 31-42.

[5] Mental workload: Its theory and measurement[M]. Springer Science & Business Media, 2013.

[6] Li D, Wang X, Menassa C C, et al. Understanding the impact of building thermal environments on occupants' comfort and mental workload demand through human physiological sensing[M]//Start-Up Creation. Woodhead Publishing, 2020: 291-341.

[7] He Xiaomei, Li Jianyu. Research on the causes analysis and design strategies of cognitive load in human-computer interaction[J]. Packaging Engineering,2020, 41(10):24-30.DOI:10.19554/j.cnki.1001-3563.2020.10.005.

[8] Rainieri G, Fraboni F, Russo G, et al. Visual scanning techniques and mental workload of helicopter pilots during simulated flight[J]. Aerospace Medicine and Human Performance, 2021, 92(1): 11-19.

[9] Longo L. Mental workload in medicine: foundations, applications, open problems, challenges and future perspectives[C]//2016 ieee 29th international symposium on computer-based medical systems (cbms). IEEE, 2016: 106-111.

[10] Dias I M. Work zone crash analysis and modeling to identify the factors affecting crash severity and frequency [M]. Kansas State University, 2015.

[11] Mansikka H, Virtanen K, Harris D. Comparison of NASA-TLX scale, modified Cooper–Harper scale and mean inter-beat interval as measures of pilot mental

workload during simulated flight tasks [J]. Ergonomics, 2019, 62(2): 246-254.

[12] Howell N, Devendorf L, Tian R, et al. Biosignals as social cues: Ambiguity and emotional interpretation in social displays of skin conductance[C]/ /Proceedings of the 2016 ACM Conference on Designing Interactive Systems, Brisbane: ACM, 2016: 865-870.

[13] WY, Yuan XG, Liu ZQ. Analysis of relations between changes of pupil and mental work loads [J]. Space Medicine & Medical Engineering, 2007, 20(5): 364-366.

[14] Hsu CK , Lin SC, Li WC. Visual movement and mental-workload for pilot performance assessment[C]/ /International Conference on Engineering Psychology and Cognitive Ergonomics , Switzerland: Springer International Publishing, 2015: 356- 364.

[15] Liu Yan, Cheng Baoquan, Yu Xin, Gu Haowu, Wang Rui, Xian Jishu, Quan Yulian, Li Fei, Feng Hua, Chen Tunan, Wang Lihua. A typical correlation analysis of brain load, sleep and fatigue in helicopter pilots [J]. Journal of the Army Medical University, 2022, 44(11): 1087-1093. DOI: 10.16016/j. 2097-0927. 202201151.

[16] Wang Lei, Wang Shuo, Zou Ying, Zhang Mengxi, Wu Jianjun, Li Shu. A study on the brain load characteristics of airline pilots based on task context [J/OL]. Journal of Safety and Environment:1-9 [2022-11-20].DOI:10.13637/j.issn.1009-6094.2021.2153.

[17] Wei Zongmin, Wanyan Xiaoru, Zhuang Damin . Measurement and evaluation of mental workload for aircraft cockpit display interface [J]. Journal of Beijing University of Aeronautics and Astronautics, 2014, 40(1) : 86-91.

[18] Grier R A. How high is high? A meta-analysis of NASA-TLX global workload scores[C]//Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Sage CA: Los Angeles, CA: SAGE Publications, 2015, 59(1): 1727-1731.

[19] Chu Hongzuo, Jiao Xuejun, Jiang Jin, Cao Yong, Yang Jiehong, Huang Mengying, Wang Lizhi, Li Xiugan. Characteristics and recognition of personalized EEG response to mental workload [J]. Manned Spaceflight, 2021, 27 (06): 710-718.DOI: 10.16329 / j.cnki. zrht. 2021.06.006.

[20] Chen Qichao, Zhang Xuejun, Huang Wanlu. EMD fusion PSD, CSP EEG feature extraction method [J].Computer technology and development, 2019, 29 (05): 126-130.

[21] Jia Dongwei, Gan Yong, Wang Yifan. Model study based on improved LSTM-CRF combined with attention mechanism [J]. Network security technology and application, 2022 (11): 39-40.

Fabricate the Auto-aquaculture Structure with Android Monitoring System

Zhiyao Zhang
School of Electronic Engineering
Xi'an University of Posts and Telecommunications
Xi'an, China
E-mail: cokebjer@163.com

Shujuan Chang*
School of Electronic Engineering
Xi'an University of Posts and Telecommunications
Xi'an, China
E-mail: 364945696@qq.com

Yueyao Tian
Dept of Electronic Engineering
Xi'an University of Posts and Telecommunications
Xi'an, China
E-mail: 3247882993@qq.com

Lei Tian
School of Electronic Engineering
Xi'an University of Posts and Telecommunications
Xi'an, China
E-mail: tla02@126.com

***Abstract*—Based on the Android monitoring system, the automated fish feeder has been developed. The system provides a convenient and reliable solution for fish farmers. This system includes a fish feeder that distributes food at predetermined intervals. The Android application allows farmers to monitor and control the feeding process remotely. The application displays the current feeding schedule. At the same time, the users can adjust the frequency and amount of food dispensed. The alarm function can send the notification information to the farmer's mobile phone if the feeder experiences any issues or requires maintenance. By automating the feeding process and providing real-time monitoring, the system can help farmers optimize fish growth and health while reducing the time and effort required for artificial feeding.**

Keywords-Automated Fish Feeder; Android; Remote Control; Real-Time Monitoring

I. INTRODUCTION

A. Project context

Automation is defined as self-regulating control of equipment, system or processes without human intervention [1]. Surely, the scientific wizardry of automation suggests that the day will surely come when everything is automated. Some fish breeding center and fish shop the fish and breeding system is manually operated. The owner of some fish breeding center and pet shop hire a person/s to feed the fish twice a week and to monitor the aquarium and the fish [2-3].

In addition, developing an Automated Fish feeder with web-based monitoring system that will greatly assist fish breeders and the productivity of the farm was to help them operate more without bearing too much of a cost on other things [4-5].

The project was also very convenient on the part of the fish breeders when on vacation and for those living a busy lifestyle because some fish are sensitive normally need to be fed once or twice a day [6-7]. Traditional feeding is done manually.

The traditional way to feed fish is by hand feeding and monitoring, the manual monitoring of the fish needs a one or more person to monitor the fish activity and the to monitor the breeder fish [8]. The breeder fish is very sensitive so need to monitor it daily.

The Automated fish feeder with web-based monitoring system is a self-feeding machine that can automatically fed the fish regularly [9-10]. In order to feed automatically the servo motor was control by microcontroller board or the Arduino board [11]. The servo motor is set or program to feed the fish in accurate time [12]. The self-feeding machine has two sensors for monitoring purposes; the first sensor is water temperature sensor the

sensor that was sense the water temperature inside the tank and the output of the water temperature is Fahrenheit. Secondly, the water turbidity sensor, this sensor can sense the turbidity of the water, or the clarity and turbidity of the water in the tank [13]. The two sensors were ease the task of fish breeders for their monitoring for their fish especially for breeder fish that is so very sensitive to the temperature and water turbidity.

The fish farmers and fish pet enthusiast, which are the main target of the development of the project, could significantly benefit an easy access of information inclined to feeding and aquarium monitoring relevant to aquaculture progress in the country [14], it is reasonable to innovate the original process of acquiring feeding and monitoring management knowledge into a new technical perspective.

B. Objective of the project

The primary objective of the project is to develop a prototype that would automatically feed the fish powered by Arduino. This includes having precise processes such as time setting and adjusting feed release. The monitoring on the prototype water condition is to ensure that the fish inside the aquarium is healthy at its right water temperature and turbidity [15].

This prototype is rising to encounter the following objectives specifically:

1) To design an automated fish feeder with based monitoring system.

2) To generate results based on the data provided by the Arduino sensors.

3) To create a web-based monitoring system

4) To test the functionality for the following features:

5) To gather recommendation from the students and respondents.

6) To evaluate problems encountered.

C. Scope and limitation of the Project

The scope covers the lists of capabilities that an Arduino-based fish feeder can perform while limitations are the operations that the device Automation of fish feeding was mainly control by the system which is powered by Arduino and sensors, the servo motor program to release feeds once a day with specific time period. The prototype only focuses on tropical fish such a gold fish, carp, koi, and molies for breeding. Desktop apps focus only on water temperature and turbidity. The prototype is design with feedback sensors which are the Waterproof Temperature Sensor Thermal Probe Wire Thermometer DS18B20 and turbidity sensor to monitor the fish using web-based and desktop application. While feed container is mainly controlled by Servo motors to aid the dispensing of feeds from the container.

The prototype would be capable of releasing feeds at a given time. The output of the prototype was display in the automated fish feeder desktop application and automated fish feeder web-based application. The prototype can only measure temperature ranging from 0-99 degree Celsius and the result of turbidity sensor based (NTU) Nephelometric Turbidity Units. The instrument used for measuring it is called water turbidity sensors, which measures the intensity of light scattered at 90 degrees as a beam of light passes through a water sample and the water turbidity sensor can only measure water turbidity ranging 0-900 turbidity units.

II. REVIEW OF RELATED LITERATURE

A. Aquaculture

The aquaculture is the farming of freshwater and marine plants and animals. Evidently, all aquaculture is done in water and, because it is a farming activity, involves the considerations of property or the farmer who owns the products and activity or work is done in order to raise the animals or plants. Sometimes, the terms "aquiculture" and "aqua farming" are also used. This activity was done in many water sources types such as river, pond, lake and others. Today's industrialist take part in this activity by investing a large amount of money in managing, inventing and also marketing the output of aquaculture which promise a good potential as a profit source to gain back a good income to them or their company.

B. Fish feeder system

The fish feeder system is a device or an electronic gadget that has been developed or designed to dispense the exact or right number of pellets at an exact time. However, this particular system has also demonstrated the ability or functionality to accurately repeat the task every day, and is therefore highly efficient and productive in the field of fish farming in the long run. This device fed fish following the right schedule and amount pre-defined by user, therefore avoiding the issue of overfeeding.

Visit the livestock aquaculture in spots, revealed that the entrepreneur hire employee to feed fish in ponds. Through the interviews, most entrepreneur think that feeding fishes by using automatic fish feeding system is more easy, useful and more effective, although it is need a high cost to develop at initial stages. Besides that, many entrepreneur or fish culturists did not know about the existence of the system or machine of fish feeding.

C. Development of Automatic Feeding Machine for Aquaculture Industry

According to the department of process and food engineering, faculty of engineering university pure Malaysia the Aquaculture is a growing industry with a great potential towards the contribution of the country’s total fish requirement. Serious efforts have been done to develop and improve the production of fish by rearing high value fish in tanks or ponds. Under the Third National Agricultural Policy (1998-2010), the target is to annually produce 1.93 million tons of fish worth approximately RM8.3 billion by the year 2010. Consequently, the development of an automatic fish feeding machine can be very beneficial to the growth of the aquaculture industry. This device was developed to overcome labor problems in the industry and introduce a semi-automatic process in the aquaculture industry. It is capable of distributing various forms of dried fish food (such as pellets, sticks, tablets or pellets) to a tank or pond in a controlled manner within a set time period. The automatic fish feeder is controlled by a digital timer and it is capable of feeding the fish in accordance with a pre-determined time schedule without the presence of an operator, and at a feeding rate of 250g/min. The feeder can be adjusted to the desired height and conveniently moved around to be positioned adjacent to the pond or tank. Meanwhile, its hopper can be covered and easily dissembled to change the size of the hopper to accommodate different capacities of feed. This automatic fish feeder can be implemented in aquaculture system to convenience to fish culturists.

D. Development of an automatic fish feeder

The automatic fish feeder was designed, fabricated and tested. It eliminates major problems associated with manual feeding in aquaculture. The machine was powered electrically by one horse power motor. The overall dimension of the machine is 62*45*45cm. The hopper carrying capacity is 5.5kg/volume of hopper with a variable discharging chute. The timer was designed with a 24hour time step at user's specified discharge duration. Test results at the discharge time of 60 minutes showed that 85.5kg of feeds were evenly distributed across the pond. Less than 3% feed loss was recorded due to breakage and fragile nature of feed. The machine has 86.9% efficiency and adequately manages and preserves feed under harsh conditions.

E. Automatic Feeding Control for Dense Aquaculture Fish Tanks

According to the efficient CV system to continuously monitor the fish-eating activity. Detect excess feed, and automatically control the feeding process. A two-class classifier is learned to distinguish whether fish are actively consuming feed or not. To detect the amount of feed floating on the water surface, we propose a novel two-stage approach.

First, a supervised learned correlation filter is applied to the test frame in order to detect every individual feed. Second, a Support Vector Machine (SVM) classifier is deployed as a refinement step of the correlation filter output, which attempts to suppress falsely detected feed while preserving true feed. Furthermore, we propose to detect feed in an optimum local region only, rather than the entire frame whose accuracy

and efficiency are both less than ideal. Using the particle filter technique, the local region is estimated by maximizing the correlation between the number of locally detected feeds and the actual number of feeds in the entire frame. Finally, based on continuous measurements from fish activity and feed detection, various actions take place to control the feeding process.

III. RELATED STUDIES

A. Automatic Fish Feeder System Using Microcontroller

This device developed combines mechanical and electrical system in controlling fish feeding activity as shown in Figure 1. Page 15. The pellets controlled by DC motor which located under the pellet storage. A control system was then attached to this device allowing the fish to be fed at the right cycle time as required or predefined by user or entrepreneur. Timer was employed in this device to control the motor rotation attached to sphere former, which dispense the pellets into the water. The pellets dispensed into the marking area of the pond based solely on the rotation speed of the motor itself. The controller came with a keypad giving user more option in determining the suitable speed for the motor depends on their cattle.

B. Turbidity Units of Measurement

A wide variety of probes are available to measure turbidity the degree to which light is scattered by particles suspended in a liquid. The measured turbidity, however, this depends on the wavelength of the light and the Angle the detector is at. Turbidity detectors don't all use the same light source for a variety of reasons，angles of measurement to detect the scattered light, and signal processing strategies. As a result, measurements from different makes/models of turbidity probes may not be comparable to one other.

Turbidity is measured in NTU: Nephelometric Turbidity Units. The instrument used for measuring it is called nephelometer or turbidimeter, which measures the intensity of light scattered at 90 degrees as a beam of light passes through a water sample.

The unit used in the ancient times was JTU (Jackson Turbidity Units), measured with the Jackson candle turbidimeter. This unit is no longer in standard use.

C. The Arvotec Feeding Robot

The Arvotec Feeding robot was developed in order to meet the customer requirement. The Arvotec feeding robot improves the feeding efficiency and saves labor time. One feeding robot supplies many tanks, eliminating the need for a feeder at each tank. A high feed turnover rate through the hopper eliminated rancidity or other 16 storage problems. The Arvo-Tec T Drum feeder has a very high accuracy, whilst remaining at a competitive price. The feeder is multifunctional and is suitable for start feeding in hatcheries to on-growing on tanks, ponds and cages. Feed amounts are automatically calculated separately to each tank according to automatically updated biomass data, incoming water temperature and oxygen content. The system is easy to use with a menu driven display in control unit or an optional MS Windows based PC connection. The feeding program is controlled by a microprocessor mounted on the robot, which can be connected to a PC for monitoring and control by a centrally managed Network Control System. The disadvantages on this robot are it high cost of manufacturing and difficult to operate this robot uses battery power supply 24-volt direct current and speed 16 minutes per meters.

The standard feeding system utilizes one or more feed storage silos, a regenerative or positive displacement blower, one or more frequency-controlled dosing augers, rotary air lock (sluice) hopper, feed injector and a rotary selector valve and diverter valves. The operator interfaces and programs the feeding system with a standard Personal Computer (PC) programmed with our Auto Feeder Software. In basic operation feed from the storage silo empties into a feed dosing auger. The auger moves this feed into a sluice-hopper-injector system. The injector introduces the feed into the main transport pipe. The feed is then picked up by air from the blower and moved into the distribution valve where it is directed into the individual feed pipes travelling to the specific tanks, raceways or net-pens on the site.

All aspects of the feeding operation such as feed rates, meal times, feed types, etc.

D. Aquarium Water Temperature

To make temperature reading easy, consider purchasing an adhesive temperature strip that can be applied to the outside of your aquarium glass. They are inexpensive and available at your local fish retailer. Monitor the sun-Too much sunlight can cause algae growth. If you are installing a new aquarium, place it away from a window. If your existing aquarium is near a window, use a shade to reduce the amount of ambient light. Avoid heat never place an aquarium above or near a heat source or air conditioner. You want to keep the area's temperature as stable as possible. For all species of fish, there is a middle ground in which they are happiest, and they must be kept within that range for optimum health. Examples: Tropical fish 22-26.6° C (optimal is 25.5° C), Common Goldfish 18.5-22.5 ° C, Fancy Tail Goldfish 20-22.5° C. Submersible heaters are the most popular. Some are adjustable and others are pre-set at 78° for your convenience. Be sure to purchase a Pour tank. Never remove a heater from the water if it is still turned on. Unplug it first, let it cool, and then remove it.

E. Feeding Systems for Sea Cages

The AQ300 Adaptive feeding system uses the Infra-Red waste feed sensor and controls automatic feeders and feed cannons on sea cages. The AQ300 is versatile having managed feeding on 18 species of cultured finfish and it also monitors key environmental parameters and transmits data locally or globally through the mobile phone network.

A recent focus for AQ1 was the development of video-based feeding systems. Popular with Tuna and Kingfish farmers the AQTV-Pro connects with one or two single or Pan/Tilt underwater cameras and transmits video images of fish feeding via WLAN to a feed boat/barge or by 3G (internet) to anywhere in the world.

F. Feeding Systems for Pond

Understanding the needs of pond farmers to control feeding accurately, AQ1 has been developing passive acoustics (hydrophone) technology over the last 4 years. The SF500 Sound Feeding System contains smart algorithms which analyses fish feeding sounds and intensity and algorithms to control feed output to match this feeding intensity precisely.

IV. METHODOLOGY

A. Technical Background

Nowadays some of aquaculture has improve their feeding management by using high-tech gadgets or machine for their fish, but still, they sometime need manual survey, which is referring to human job to manage their machine. There are many jobs that require human labor to perform tasks such as cleaning the feeder, refilling pellets, and even repair or maintenance procedures. Thus clearly, the problem here is to invent a low cost and easy maintenance machine that can help them to feed their fish. The feeding device should be simple in construction and operation, reliable in operation, and relatively inexpensive to purchase and operate. This solution also must have good efficiency in the way it operates and serves as a dependable worker.

Some automated fish feeder sold in the market with the high price that the most fish breeders can't afford. And other fish breeders need to hire a person in order to monitor and to feed the fish, with that problem the fish breeders need to invest more money in order to monitor and to feed the fish. The automated fish feeder can be set time to feed the fish just like a commercial automated fish feeder but the different is, the propose project has a monitoring system. The web-based monitoring system will allow fish farmers to see the results of the sensors. The prototype consists of two sensors; first, the Water temperature sensor which is important to the fish breeder because some fish are more sensitive in the water temperature, second, the turbidity sensors which is important too. The turbidity of the water is very vital to some fishes.

The turbidity is the haziness of a fluid causes by large number of individual particles that are generally invisible to the naked eye, similar to smoke in air. Though invisible and odorless, nitrate can have a deadly effect on a tank's

inhabitants, but there are simple and effective methods for keeping it at a tolerable level.

B. Details of the Technology to be used

The technology use in designing the automated fish feeding and monitoring system are:

TABLE I. LIST OF SOFTWARE TO BE USED

List of Software to be used	
Visual Studio	The visual studio 2008 or higher the software to develop the web side and desktop application.
Adobe Photoshop	The Adobe photoshop CC the software to be used by the proponent to design the GUI of the system.
MySQL	The MySQL the database management system that runs as server.
Google Sketch-up	Google sketch-up the software that use by the proponent.
Arduino	The Arduino is an open-source platform

TABLE II. LIST OF HARDWARE TO BE USED

List of Hardware to be used	
Arduino Board	Microcontroller board control all the devices and sensor attach on it.
Servo Motor	Device main function is to dispense certain amount of feeds.
Water Temperature Sensor	This sealed digital temperature probe lets you precisely measure temperatures.
Aquarium	Prototype being installs.

C. Circuit Diagram

A Circuit Diagram for checking the water temperature with Arduino & water resistant DS18B20 or the sensor can detect the water temperature.

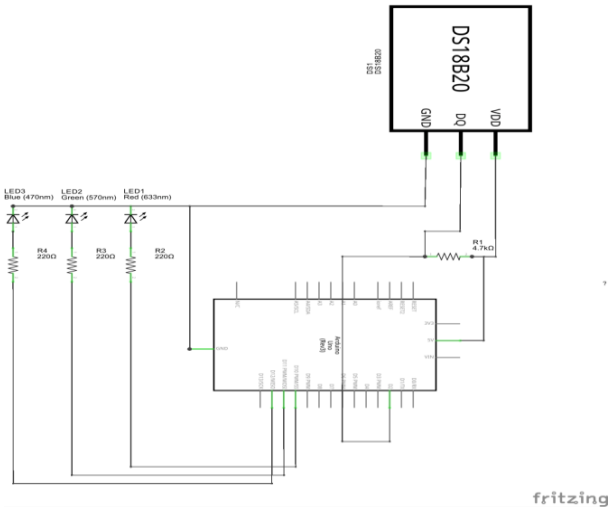


Figure 1. Circuit Diagram Temperature sensor DS18B20

The Circuit diagram for Turbidity temperature this sensor can sense the turbidity of the water. Turbidity is measured in NTU: Nephelometric Turbidity Units.

The instrument used for measuring it is called nephelometer or turbidimeter, which measures the intensity of light scattered at 90 degrees as a beam of light passes through a water sample.

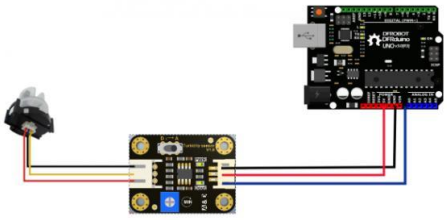


Figure 2. Water turbidity sensor SKU: SEN0189

The highlighted red rectangular shape shows the result of the gathered data from the turbidity sensors.

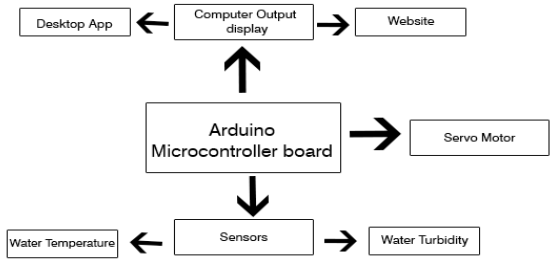


Figure 3. Hardware Block Diagram

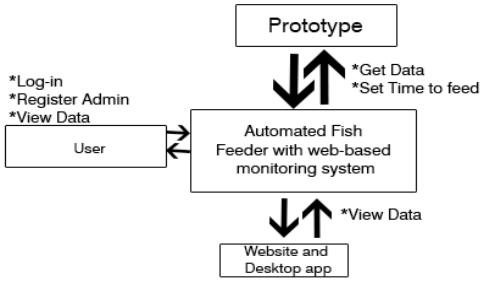


Figure 4. Software block Diagram

D. Capstone Specification

The automated fish feeder and Web based monitoring system is design to be easily learned and adapt by the anticipated users. The users are intended to undergo proper training and orientation in this project. On the other hands, the maintenance and the sustainability of the project will be done by the developer. Generally, the concepts before the users will become an efficient user of the automated fish feeder with Web based monitoring system control application.

The automated fish feeder consists of the following hardware; first the servo motor serves as a feeds dispenser, second the arduino sensors which are the water temperature sensor is to sense the temperature of the water turbidity sensor sense or to detect the water turbidity. The data gathered from sensor will be sent to the Web-site and allowing the fish breeder to monitor their fish through Web.

The proponent develops a monitoring application or web-site that to show the data produce from the sensors.

The user can monitor the Water Temperature and Turbidity in two ways first in offline using the Desktop application and second the web-based monitoring system.

The user can able to see the information of the sensor and the information of the prototype.

The proponent develops a desktop application in able to get data from the prototype and save into the Database.

E. Features of the Desktop Application

Inside this Panel is the System Information, visit website panel, developer info and user manual.

Manage Admin panel this panel can add admin account in able to be login in the desktop application.

Setting panel this panel is the main tool for the desktop application the module to connect or to fetch data from the sensors.

F. Result of the Sensors

Water Temperature

The results of the Water Temperature sensor is automatically sent to the computer. The data gathered from the sensors was stored in the database.

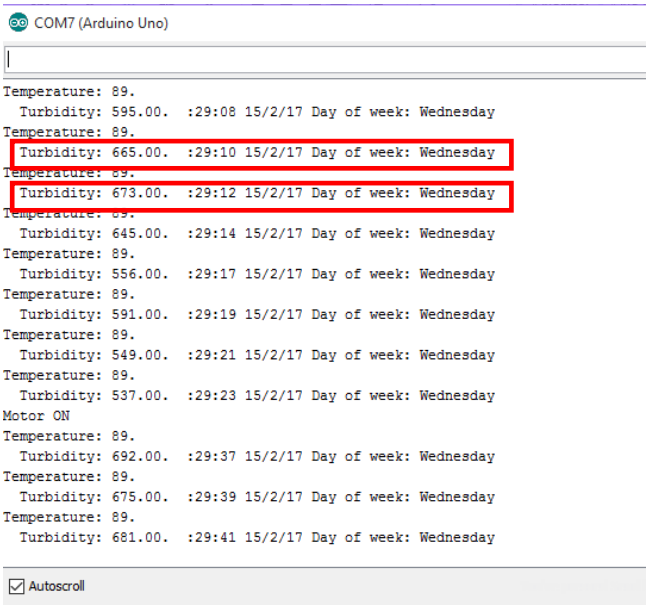


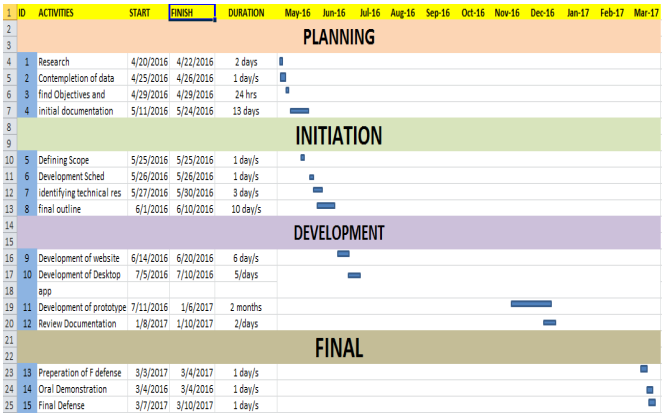
Figure 5. Generated result from the water temperature sensor

G. Schedule Feasibility

The automated fish feeder and Web based monitoring system was appropriate and practical to its time frame in creating the documents, fabrication of the prototype and the development of application.

The list of task with its corresponding duration in Gantt Chart/Schedule of activities in below must be followed to achieve specific goals in exact schedules.

Gantt Chart



	ID	ACTIVITIES	START	FINISH	DURATION
		PLANNING			
	1	Research	4/20/2016	4/22/2016	2 days
	2	Contemplation of data	4/25/2016	4/26/2016	1 day/s
	3	find Objectives and	4/29/2016	4/29/2016	24 hrs
	4	initial documentation	5/11/2016	5/24/2016	13 days
		INITIATION			
	5	Defining Scope	5/25/2016	5/25/2016	1 day/s
	6	Development Sched	5/26/2016	5/26/2016	1 day/s
	7	identifying technical res	5/27/2016	5/30/2016	3 day/s
	8	final outline	6/1/2016	6/10/2016	10 day/s
		DEVELOPMENT			
	9	Development of website	6/14/2016	6/20/2016	6 day/s
	10	Development of Desktop	7/5/2016	7/10/2016	5/days
		app			
	11	Development of prototype	7/11/2016	1/6/2017	2 months
	12	Review Documentation	1/8/2017	1/10/2017	2/days
		FINAL			
	13	Preperation of F defense	3/3/2017	3/4/2017	1 day/s
	14	Oral Demonstration	3/4/2016	3/4/2016	1 day/s
	15	Final Defense	3/7/2017	3/10/2017	1 day/s

Figure 6. Gantt chart for the project timeline

V. Design and Interface

A. Output and user interface design

The figure in the next page shows the sample data from the water temperature and turbidity sensors.

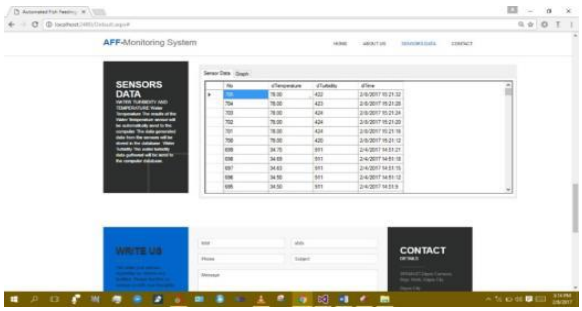


Figure 7. Water temperature and Turbidity Panel.

B. Sample Gathering of data

The table below shows the sample data between Arduino sensors and the manual process.

C. Configuration and Testing Description

First the use clicks the button monitor now to monitor the prototype activity. Then, the login form will pop-op for security purpose. Only admin or the user can use the desktop application. The Sensors Data Tab will view the Data Gathered from the sensors. The User or the Admin can add, edit, delete in Manage Admin Module.

VI. Conclusion

In this paper, an automatic water feeding device is developed with Android system. The system provides a convenient and reliable solution for aquatic farmers. Feed time is set by timer and food is distributed at predetermined intervals. It provides users with remote monitoring and alarm functions, which not only allows users to control the frequency and amount of feeding in the water in real time, but also allows the system to send information to the user's mobile phone when there is a problem. In the normal use of the process of automatic feeding, the system can help farmers to optimize the growth and health of fish, while greatly reducing labor costs. It can provide technical support for intelligent aquaculture.

Acknowledgment

This work was partly supported by the undergraduate innovation and entrepreneurship plan 202211664099X and 202211664052.

References

[1] Myung-Jae Lee, Holger Rücker, Woo-Young Choi, Effects of Guard-Ring Structures on the Performance of Silicon Avalanche Photodetectors Fabricated With Standard CMOS Technology, IEEE Electron Device Letters, Vol. 33, Issue.1, 2012, pp.80-82.doi:10.1109/LED.2011.2172390.

[2] Chee Hing Tan, Shiyu Xie, Jingjing Xie, Low Noise Avalanche Photodiodes Incorporating a 40 nm AlAsSb Avalanche Region, IEEE Journal Of Quantum Electronics, Vol. 48, Issue. 1, 2012, pp. 36-41. doi:10.1109/JQE.2011.2176105.

[3] Kang Yu-zhuo, Mao Lu-hong, Xiao Xin-dong, Xie Sheng, Zhang Shi-lin, Design and simulation of a novel CMOS superimposed Photodetector, Optoelectronics Letters, Vol. 8, Issue.4, 2012, pp. 02549-0252. doi:10.1007/s11801-012-1150-z.

[4] Da Silva, M. F.; Fraytag, J.; Schlittler, M. E.; Marchesan, T. B.; Dalla Costa, M. A.; Marcos Alonso, J.; do Prado, R. N., Analysis and Design of a Single-Stage High-Power-Factor Dimmable Electronic Ballast for Electrodeless Fluorescent Lamp. IEEE T Ind Electron 2013, 60 (8), 3081-3091. doi:10.1109/TIE.2012.2203774.

[5] Fayomi, C.J.B., Wirth, G.I., Achigui, H.F., & Matsuzawa, A. (2010). Sub 1V CMOS Bandgap reference design techniques: a survey. Analog Integrated Circuits and Signal Process, 62,141-157. doi:10.1007/s10470-009-9352-4.

[6] Jong Tae, H.; Moon Sang, J.; Dae Ho, K.; Jun Hong, L.; Min Ho, J.; Jong Ha, S., Off-the-Line Primary Side Regulation LED Lamp Driver with Single-Stage PFC and TRIAC Dimming Using LED Forward Voltage and Duty Variation Tracking Control. Solid-State Circuits, IEEE Journal of 2012, 47 (12), 3081-3094. doi:10.1109/ISSCC.2012.6177014.

[7] JUN Luo, Qian Wang, Xianghai Xiao. A modified artificial bee colony algorithm based on converge-onlookers approach for global optimization

[J]. Applied mathematics and computation, 2013, 219(20):10253-10262. doi:10.1016/j.amc.2013.04.001.

[8] Lixia Han, Shujuan Jiang, Shaojiang Lan. Novel electromagnetism-like mechanism method for multiobjective optimization problems [J]. Journal of Systems Engineering and Electronics, 2015,26(1):182-189. doi: 10.1109/JSEE.2015.00023.

[9] FAN Chao-dong, OUYANG Hong-lin, ZHANG Ying-jie, etc. Optimization algorithm based on kinetic-molecular theory [J]. Journal of Central South University, 2013,20(12:3504-3512. doi:10.1007/s11771-013-1875-2.

[10] Li Y, Liu T, Hu J, et al. Topical Co-Attention Networks for hashtag recommendation on microblogs[J]. Neurocomputing, 2019, 331: 356-365. doi:10.1016/j.neucom.2018.11.057.

[11] Shin Y , Lee S J , Park J . Composition pattern-oriented tag extraction from short documents using a structural learning method[J]. Knowledge and Information Systems, 2014, 38(2):447-468. doi:10.1007/s10115-012-0594-6.

[12] Xiaobing X , Gang Z , Yongzhong H , et al. Dynamic evolution of collective emotions in social networks: a case study of Sina weibo[J]. Science China, 2013(07):150-167. doi: 10.1007/s11432-013-4892-8.

[13] Goonetilleke O , Sellis T K , Zhang X , et al. Twitter analytics: a big data management perspective.[J]. ACM sigkdd explorations newsletter, 2014, 16(1):11-20. doi:10.1007/s11704-013-3903-7.

[14] Das, P.; Pahlevaninezhad, M.; Drobnik, J.; Moschopoulos, G.; Jain, P. K., A Nonlinear Controller Based on a Discrete Energy Function for an AC/DC Boost PFC Converter. IEEE T Power Electr 2013, 28 (12), 5458-5476. doi:10.1109/TPEL.2012.2232681.

[15] Da Silva, M. F.; Fraytag, J.; Schlittler, M. E.; Marchesan, T. B.; Dalla Costa, M. A.; Marcos Alonso, J.; do Prado, R. N., Analysis and Design of a Single-Stage High-Power-Factor Dimmable Electronic Ballast for Electrodeless Fluorescent Lamp. IEEE T Ind Electron 2013, 60 (8), 3081-3091. doi:10.1109/TIE.2012.2203774.

Research and Simulation of Negative Group Delay and Superluminal Propagation

Yueyao Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 1278730453@qq.com

Xinzhuo Li
School of electronic information Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail:1352580810@qq.com

Zhongsheng Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: wzhsh1681@163.com

Han Shen
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: sunny_shine_zj@163.com

***Abstract*—In recent years, negative group delay circuits have attracted much attention due to their propagation characteristics and wide application prospects. In the history of human exploration, the exploration of the speed of light has never stopped. The theory of relativity points out that the speed of light in vacuum is the limit speed of signal propagation. However, it is found through research that phase velocity and group velocity appear faster than the speed of light, which does not violate the causal relationship. This paper first introduces the related concepts of negative group delay and superluminal phenomenon, the second focuses on the principle of negative group delay and superluminal phenomenon in-depth analysis and research, finally using the principle of Multisim software, the bandwidth of two different job, different structure of circuit design, the virtual simulation experiment to negative group delay phenomenon and measurement data. It is of great significance to explore the field of faster-than-light and negative group delay in today's rapidly developing information age, and it can try to meet the high requirements for signal transmission. In the future, the interdisciplinary research direction of this research topic also has great development space.**

Keywords-Superluminal Phenomenon; Phase Velocity; Group Velocity; Signal Transmission; Simulation Study

I. INTRODUCTION

For a long time, most of the research on linear signal transmission system focuses on amplitude-frequency response, but pays less attention to phase-frequency response. In recent years, with the development of communication and space detection, the transmission capacity and channel transmission characteristics of signals are increasingly demanding. Scholars have found that compared with the information transmission of amplitude in the process of signal transmission, the phase can carry more useful information with higher stability, and group delay is a very important parameter. The study of group delay and superluminal speed is of great significance in space testing, navigation system, military positioning and so on.

Sommerfeld and L. Brillouin discovered the negative wave velocity in classical electromagnetic wave theory, and pointed out that in the irregular dispersion region, the group velocity can exceed the speed of light in vacuum, and can even be negative. At present, there are two main directions in the research of "superluminal physics". One is the superluminal phenomenon realized by particles through

DOI: 10.2478/ijanmc-2023-0050

anti-dispersion regulation, and the other is the superluminal phenomenon realized by electronic circuits. Human's exploration and measurement of the speed of light can be traced back to 1607, when Italian scientist Galileo Gallei put forward the scientific idea that "the speed of light is finite" [1], although his attempt was unsuccessful. In 1907, Einstein discussed the concepts of negative group velocity and negative time. In 1970 G. Garrett and D. Mc Dumber showed that faster-than-light group velocities and even negative group velocities could be observed experimentally while keeping the pulses undistorted. N. Budko et al. discovered negative wave velocities in 2009 while studying the near-field region of an antenna (source), with some waveforms showing retrograde to time. In 2011 Liang Zhang et al. observed a time advance of 221.2ns before the input signal. Hyper-luminal physics can be tested in two broad categories, namely hyper-luminal phenomena of particles regulated by anomalous dispersion and negative group velocity phenomena in electronic circuits, of which negative group velocity accounts for about half. The research of superluminal speed in China began roughly in the 1980s.In 1984, Tan Xue sheng put forward the standard space-time theory, which allows the existence of superluminal motion without violating the theory of causality. In 2003, Huang Zhixun et al. completed group velocity hyper-luminal experiment at shortwave, which was the first hyper-luminal experiment in China. Meanwhile, in 2002, Liu Liao pointed out that the negative velocity means that the delayed light pulse becomes the advanced light pulse, which challenges the absoluteness of the time sequence of causality and also has an impact on relativity. Up to now, the study of superluminal physics has a short history and its development is not yet mature, but it has already taken preliminary shape. As a newly emerging science with promising signs, its vitality and development potential is obvious and very necessary. For example, in 2012, the United States Space Administration (NASA) organized the "Star-ship Centennial Symposium", in which the so-called warp propulsion of the superluminal universe solution was discussed; In recent years, many Chinese scientists have joined in the field of faster-than-light research and made contributions. In 2019, Xia Hualing is the first person to utilize the fidelity concept in quantum mechanics to define the signal integrity of the signal through the dispersion circuit, and is the first person to apply the cascaded dual-band electronic amplification circuit to obtain abnormal dispersion region, and systematically proves that fidelity change in normal and abnormal dispersion cascaded electronic circuit depends on different truncation parameters or pulse shape on the electrical signal [8].

In this paper, the definition of phase velocity and group velocity is introduced, and the idea of group velocity exceeding the speed of light is in line with relativity is proved by quantum hypothesis and wave dispersion relationship. At the same time, the production mechanism of negative group delay and circuit characteristics are elaborated in detail. A virtual simulation experiment is carried out. Two kinds of electronic circuits of high frequency band and low frequency band are used to measure the pulse results by using Multisim simulation software. The results are consistent with the theoretical calculation. It is of great significance to explore the field of superluminal speed and negative group delay in today's rapidly developing information age, so we can try to meet the high requirements of signal transmission.

II. APPLICAIOTN SCENARIOS

A. Concepts related to superluminal velocities

In the vibration and wave of the optical part of physics, the two basic physical quantities, phase velocity and group velocity, play an important role in the related research and exploration.

Phase velocity refers to the phase moving velocity or phase velocity of a wave, or phase velocity for short. Phase velocity is defined as the advancing velocity of a constant phase point of an electromagnetic wave [2].In other words, this is the speed at which the wave will transmit the phase that any frequency component has. In which the electric field is:

$$E\ \ z,t\ \ = E_m \cos\ \ \omega t - \beta z \tag{1}$$

The phase velocity of electromagnetic wave shall be:

$$\upsilon_p = \frac{dz}{dt} = \frac{\omega}{\beta} \tag{2}$$

Where beta is the phase constant, which determines whether the phase velocity is dependent on frequency.

The group velocity is the advancing velocity of an envelope wave at a constant phase point. In a complex or modulated signal, there are thousands of frequency components, and a simple phase velocity cannot accurately describe the propagation speed of a signal. The transmission of the signal depends on the modulation of the wave, and the speed of the modulated wave is the speed of the signal transmission. Given two traveling waves with both amplitudes, they can be expressed as E_m:

$$E_1 = E_m e^{j(\omega+\Delta\omega)t} e^{-j\ \beta+\Delta\beta\ z} \tag{3}$$

$$E_2 = E_m e^{j(\omega-\Delta\omega)t} e^{-j\ \beta-\Delta\beta\ z} \tag{4}$$

Where, $\omega+\Delta\omega$ and $\omega-\Delta\omega(\Delta\omega << \omega)$ is the angular frequency, $\beta+\Delta\beta$ and $\beta-\Delta\beta$ is the phase constant.

Then the resultant wave of the two traveling waves is

$$E = E_1 + E_2 = 2E_m \cos(\Delta\omega t - \Delta\beta z) e^{j(\omega t-\beta z)} \tag{5}$$

It is concluded that the amplitude of the synthesized wave is modulated, which is called envelope wave，where C is a constant.

$$\Delta\omega t - \Delta\beta z = C \tag{6}$$

The group velocity is defined as

$$\upsilon_g = \frac{dz}{dt} = \frac{\Delta\omega}{\Delta\beta} \xrightarrow{\Delta\omega<<\omega} \upsilon_g = \frac{d\omega}{d\beta} \tag{7}$$

The trigonometric part of equation (5) can be approximated as the envelope signal, and the difference in peak time between the output pulse and the input pulse is determined by the maximum value of the envelope of the output pulse and the input pulse [8]. When the output pulse distortion is small, the time delay of the envelope peak can be approximately viewed as the group time delay of the pulse. Theoretically, we know that the group delay is determined by the derivation of the phase over-frequency on the pulse carrier frequency, but due to the finite width of the pulse spectrum, the actual group delay is slightly different from the theoretical prediction.

B. Group velocities exceeding the speed of light are relativistic

In 1905, Einstein put forward the light quantum hypothesis, which explained the photoelectric effect more accurately, and won the Noble Prize in Physics in 1921.In 1904, de Broglie proposed that physical particles also have waves. He used analogy to make the wave-particle duality of light no longer limited to particles and quanta, but extended to general matter particles [10]. In quantum mechanics, it is called matter wave theory, also known as "de Broglie matter wave theory", where the relationship between energy E, momentum p, frequency ν, wavelength λ and Planck constant h:

$$\lambda = \frac{h}{p} \tag{8}$$

$$\nu = \frac{E}{h} \tag{9}$$

According to Einstein's mass-energy formula:

$$E = mc^2 \tag{10}$$

Arrive at:

$$\upsilon_p = \lambda \cdot f = \frac{h}{m\upsilon} \cdot \frac{mc^2}{h} = \frac{c^2}{\upsilon} \tag{11}$$

Where υ_p is the de Broglie wave phase

velocity, is the particle velocity υ:

$$\upsilon_g = \frac{d\omega}{dk} = \frac{d\ h\omega}{d\ hk} = \frac{dE}{dp} \tag{12}$$

Where υ_g is the group velocity, and according to the relationship between energy and the momentum:

$$\upsilon_g = \frac{dE}{dp} = \frac{p}{m} = \upsilon \tag{13}$$

That is, the particle velocity is equal to the group velocity, and then there is

$$\upsilon_g \upsilon_p = c^2 \tag{14}$$

Einstein's special theory of relativity has mentioned a very important idea: the speed of light cannot be exceeded, that is, the speed of particles must be less than the speed of light, so it can be deduced. $\upsilon < \upsilon_p\ \upsilon_p > c$ In other words, the de Broglie wave phase velocity can exceed the speed of light. Doesn't the de Broglie wave phase velocity faster-than-light violate special relativity? The fact is that the phase velocity super-lumen was discovered in the early days of Einstein's theory of relativity, but he did not think that the phase velocity super-lumen would affect the relatively sound theory of relativity, because a monochromatic wave with constant frequency cannot transmit information, and the phase velocity refers to the speed at which a monochromatic wave with constant frequency travels [3]. In order to transmit information, A slower-changing wave packet needs to be modulated onto a sinusoidal wave [6], and the propagation speed of such a packet is called the group velocity. Whether the phase velocity and the group velocity can travel faster than the speed of light actually depends on the dispersion relationship of the medium through which the wave travels. In anomalously dispersive media, the group velocity of waves can also exceed the speed of light [7].For the analysis of dispersion, according to the definitions of group velocity and phase velocity derived above:

$$\upsilon_g = \frac{dz}{dt} = \frac{\Delta\omega}{\Delta\beta} \tag{15}$$

$$\upsilon_p = \frac{dz}{dt} = \frac{\omega}{\beta} \tag{16}$$

The relationship between group velocity and phase velocity can be obtained:

$$\upsilon_g = \frac{d\omega}{d\beta} = \frac{d\ \upsilon_p\beta}{d\beta} = \upsilon_p + \frac{\omega}{\upsilon_p} \cdot \frac{d\upsilon_p}{d\omega}\upsilon_g \tag{17}$$

It follows that:

$$\upsilon_g = \frac{\upsilon_p}{1 - \frac{\omega}{\upsilon_p}\frac{d\upsilon_p}{d\omega}} \tag{18}$$

As can be seen from the above equation, group velocity and phase velocity are not equal in general, and the classification discussion is as follows:

a) No dispersion: if the group velocity is equal to the phase velocity, $\frac{d\upsilon_p}{d\omega} = 0$, then the phase velocity is independent of the frequency;

b) Normal dispersion: if the group velocity is less than the phase velocity, $\frac{d\upsilon_p}{d\omega} < 0$, the phase velocity is inversely proportional to the frequency;

c) Abnormal dispersion: if the group velocity is greater than the phase velocity, $\frac{d\upsilon_p}{d\omega} > 0$, the phase velocity is proportional to the frequency..

According to the above analysis of the dispersion of the relationship between the physical quantities. Related, there is the Cherenkov effect refers to the speed of light in a medium is smaller than the speed of light in a vacuum, particles in the medium may travel faster than the speed of light in the medium, in which case radiation will occur. But that's not really superluminal, which is superluminal in the sense of exceeding the speed of light in a vacuum.

III. Concepts and Mechanisms Related to Negative Group Delay

A. Negative group delay

The group velocity representing the transmission speed of the envelope signal in the medium can be expressed as:

$$\upsilon_g(\omega) = \frac{c}{n_g(\omega)} \tag{19}$$

Where, ω is the angular frequency, c is the speed of light in vacuum, and n_g is the group refractive index.

Therefore, the group time delay required for electromagnetic wave to pass through a medium of length z can be expressed as:

$$\tau_g(\omega) = \frac{z}{\upsilon_g(\omega)} = \frac{z n_g(\omega)}{c} = \tau_c n_g(\omega) \tag{20}$$

Where, $\tau_c = \frac{z}{c}$ is the transmission delay of light in a vacuum of length z.

According to the above two formulas, electromagnetic wave transmission in the medium has the following three conditions:

a) $n_g > 1$, the group velocity of the wave is less than the speed of light, $\upsilon_g < c$, $\tau_g < \tau_c$, τ_g is positive;

b) $0 < n_g < 1$, the wave group velocity exceeds the speed of light, $\upsilon_g > c$, $\tau_g < \tau_c$, τ_g is positive;

c) $n_g < 0$, then the group velocity of the wave is negative, and $\upsilon_g < 0$, $\tau_g < \tau_c$, τ_g is negative.

The above three cases can be shown in Figure 1, the Curves A, B and C in the figure correspond to cases a), b) and c) respectively. According to the above analysis and Figure 1, negative group velocity and exceeding the speed of light both mean that the transmission delay of envelope signals in medium is smaller than that in vacuum, namely the so-called "super-lumen phenomenon" [9]. However, different from the superluminal group velocity, negative group delay also means that the output envelope peak appears earlier than the input envelope, which is a strange phenomenon.

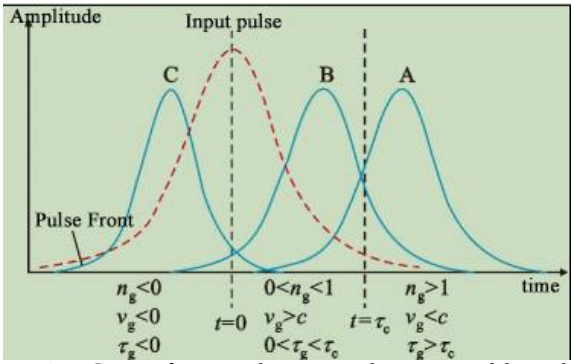


Figure. 1. Group refractive index, group velocity, group delay and the relationship between the speed of light

A. Negative group delay generation mechanism and circuit characteristics

Negative group delay phenomenon is essentially due to the abnormal dispersion characteristics of the medium. [4] Bolda et al. were the first to theoretically prove that the phenomenon of negative group delay in dispersive media occurs at the extremum of wave attenuation.

The difference between the transmission group delay defined in a medium of length z and the group delay in a vacuum is:

$$\Delta\tau(\omega) = \tau_g(\omega) - \tau_c = \left(\frac{1}{\upsilon_g} - \frac{1}{c}\right) \cdot z \tag{21}$$

Using the properties of analytic functions, it can be proved that:

$$\Delta\tau(\omega) = \frac{z}{\pi} p \int_{-\infty}^{+\infty} \frac{k(\tilde{\omega}) - k(\omega) - (\tilde{\omega} - \omega)\frac{dk(\omega)}{d\omega}}{(\tilde{\omega} - \omega)^2} d\tilde{\omega} \tag{22}$$

Where P is the Cauchy principal value and $k(\omega)$ is the imaginary part of the wave number. In the above equation, if $\frac{dk(\omega)}{d\omega} = 0$, the maximum or minimum value of $k(\omega)$ is taken, the corresponding integral result is negative or positive. Since $k(\omega)$ represents the absorption attenuation of the wave, there is at the attenuation maximum, $\Delta\tau(\omega) < 0$, indicating that the group velocity in the medium is greater than the speed of

light or is negative. This proof not only shows the condition for the existence of negative group delay, but also shows the relationship between negative group delay and amplitude response. At present, most negative group delay circuits rely on energy resonance or coupling to generate negative group delay and have obvious amplitude attenuation extremum characteristics, which is consistent with the above proof conclusions.

B. The relationship between negative group delay and superluminal speed

When we consider a system, it is easy to understand that supports a vacuum path (length L) and a lumped system (delay t_d), which is located at the end of the path [11]. The total time requires a pulse to pass through this system:

$$t_{total} = \frac{L}{c} + t_d \tag{23}$$

The corresponding total speed $\upsilon_g = \frac{L}{t}$ satisfies the relation:

$$\frac{1}{\upsilon_g} = \frac{1}{c} + \frac{t_d}{L} \tag{24}$$

For $\upsilon_g < c$ (positive delay) and $t_d < 0$ (negative delay), there are two cases. In $-t_d < \frac{L}{c}$ case, υ_g greater than c (in the narrow sense of superluminal), while in $-t_d > \frac{L}{c}$, υ_g becomes negative (negative group velocity). In the latter case, the lumped part's contribution is superior to the free propagation path.

In general, superluminal is considered a propagation effect [15] But in many cases, it seems more appropriate to discuss the negative group delay of a lumped system. We take experimental parameters from it, where we observe a negative delay of 60ns.We note that the pulse length $cTw = c \times 2\mu s = 600m$ is much longer than the cell length L = 6 cm long [14]. Therefore, we can safely use the lumped approximation. Since we can eliminate the carrier frequency with a slowly changing envelope approximation, the wavelength of light will no longer come into play. It should be emphasized that the second term in the equation.(4) or (5), whether it is positive or negative, dominates the first term. For typical atomic experiments, the bandwidth $\Delta\omega$, from MHz to GHz, roughly determines, $|t_d| \sim \frac{\Delta nk0L}{\Delta\omega} \sim \frac{1}{\Delta\omega}$, μs to ns, while the passage time is less than of ns. In this case, forcing a speed to be assigned, as in the example above, $\upsilon_g = -1/300\ c$, will cause some confusion [12].

We try to design a simple circuit to carry out a virtual simulation experiment to realize this phenomenon, to explore further.

IV. THEORETICAL CALCULATION AND SIMULATION EXPERIMENT

A. Circuit characteristic description and measurement method

Signals can be expressed in the time domain and frequency domain, and the effects corresponding to the circuit can also be described in the time domain and frequency domain. The field of group delay focuses on the change of signals in the frequency domain, so only the transfer function of the circuit and its related amplitude response, amplitude Angle response, and group delay need to be considered. The corresponding relation derived from the transfer function is shown as follows:

$$V_{out}(\omega) = H(\omega)V_{in}(\omega) \tag{25}$$

$$H(\omega) = \frac{V_{out}}{V_{in}} \tag{26}$$

The transfer function can be expressed in a complex form, namely:

$$H(\omega) = a + ib \tag{27}$$

Where a and b are represented as the real part and the imaginary part respectively, and at least one of them contains variables. From this, the amplitude response and amplitude angular response of the transfer function are:

$$A(\omega) = |H(\omega)| = \sqrt{a^2 + b^2} \quad (28)$$

$$\phi(\omega) = \arg[H(\omega)] = \arctan(\frac{b}{a}) \quad (29)$$

The group delay is the derivative of the angular response with respect to the angular frequency and takes a negative value. When the angular frequency is determined, the group delay also has a fixed value:

$$t_d = -\frac{d\phi(\omega)}{d\omega}\Big|_{\omega_0} \quad (30)$$

In an ideal state, the amplitude response is expected to be 1, so that the pulse does not deform, and the group delay can be fixed if the amplitude angular response is linear with respect to the angular frequency [16].

In the measurement of the circuit, there are two kinds of measurement: sinusoidal pulse measurement and modulated pulse measurement.

According to the description of the group velocity theory above, the group delay in envelope signals is represented by the change of the envelope peak. The input and output pulses are placed on the same time axis for comparison, and the time delay time can be obtained by comparing the position of the envelope peak on the time axis. Therefore for a modulated pulse, the group delay value can be directly obtained by comparing the peak value of its envelope signal [13].

In addition to the modulation pulse measurement, a single sinusoidal pulse (analogous to the monochromatic light dispersion effect) can also be measured. At this time, the phase shift of the pulse, namely the amplitude Angle response, can be obtained directly from the time axis, while the group delay needs to continue to be derived by its derivative. This measurement method can be used to conduct experiments from the theoretical reasoning sequence, analyze the experimental principle, but also more convenient for the measurement of low-frequency working bandwidth circuit.

A. Theoretical calculation results

In this experiment, two kinds of circuits are selected for analysis and testing. They correspond to the working bandwidth of low frequency band and high frequency band.

a) Low frequency RC loop

The working circuit of low frequency band consists of an amplifier and two groups of resistors and capacitors. The circuit diagram is shown in Figure 2. As shown in the Figure of the paper [11].

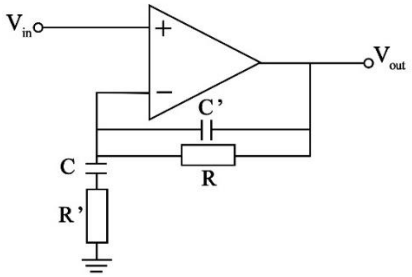


Figure. 2. Low frequency circuit

The transfer function of this circuit is:

$$H_{low}(\omega) = 1 + \frac{i\omega RC}{(1 + i\omega R'C)(1 + i\omega RC')} \quad (31)$$

After appropriate parameters are selected, the corresponding amplitude response, amplitude angular response and group delay relationship can be obtained as shown in FIG. 3, 4 and 5, where the green line represents the ideal condition, and the yellow line represents the actual theoretical data of the circuit. It can be found that the negative group delay phenomenon of the circuit is mainly in the low frequency band, and each data is symmetric relative to the frequency of 0Hz. Although the frequency less than zero is meaningless. When working in low frequency bands, the lower the frequency, the greater the negative group delay, and there is no longer a negative group delay phenomenon around 10Hz.

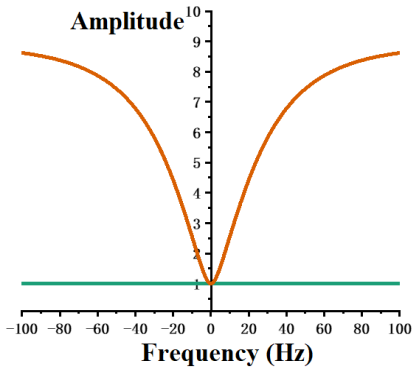


Figure. 3. Theoretical curve of amplitude response of low frequency circuit

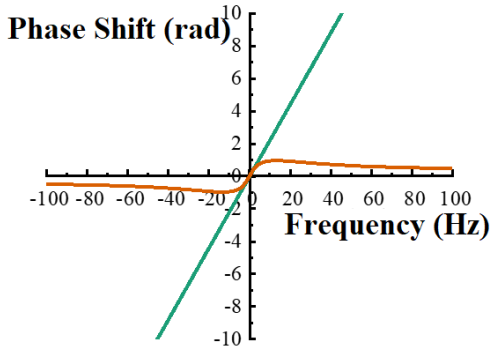


Figure. 4. Theoretical curve of amplitude-angle response of low-frequency circuits

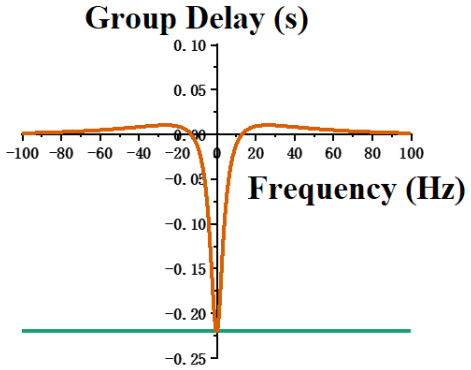


Figure. 5. Theoretical delay curve of low-frequency circuit groups

b) High band RLC loop

A circuit that realizes the phenomenon of negative group delay in high frequency band increases the use of inductors compared to low frequency circuits, and the transfer function of this circuit is relatively more complex. The first-order circuit diagram and its transfer function are shown in Figure 6, Eq. (32) and (33), while the second-order circuit diagram and its transfer function are shown in Figure 7, Eq. (34), (35) and (36).

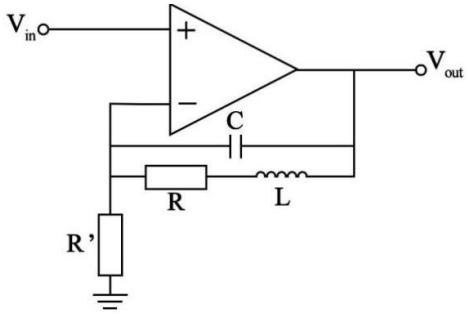


Figure. 6. First-order high-frequency circuit

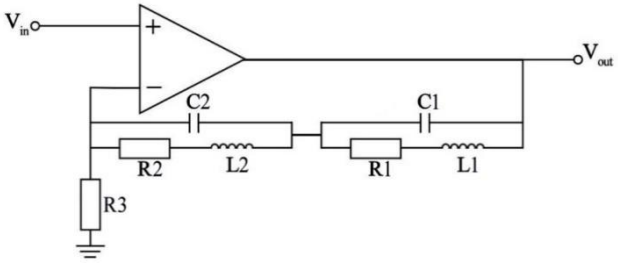


Figure. 7. Second order high frequency circuit

$$H_{high}(\omega) = 1 + \frac{Z}{R'} \tag{32}$$

$$Z = \frac{1}{i\omega C + \frac{1}{R + i\omega L}} \tag{33}$$

$$H_{high}(\omega) = 1 + \frac{Z_1 + Z_2}{R_3} \tag{34}$$

$$Z_1 = \frac{1}{i\omega C_1 + \frac{1}{R_1 + i\omega L_1}} \tag{35}$$

$$Z_2 = \frac{1}{i\omega C_2 + \frac{1}{R_2 + i\omega L_2}} \tag{36}$$

In the same way, appropriate parameters of electronic components are selected to draw the relationship diagram of amplitude response, amplitude Angle response and group delay, as shown in FIG. 8, 9 and 10, where the red curve is the characteristic curve of first-order circuit, and the green curve is the curve of second-order circuit. The operating range of this circuit is at high frequency, so it does not need to consider the situation around zero point, so it does not consider the negative frequency range of the image. Yellow

represents the theoretical data of the actual circuit. Due to the polymorphism of this circuit, the linear ideal state of the phase shift and group delay part is no longer of reference significance, but the ideal amplitude is still expected to be approximately 1. It has a very different performance from the low-band circuit, and the maximum negative group delay point is roughly 23kHz and 36kHz position, there are two points.

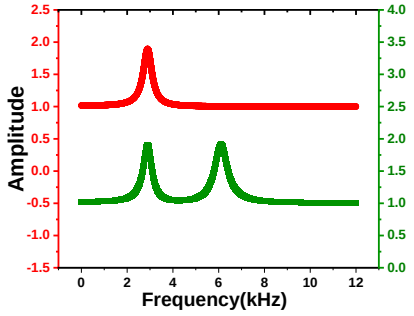


Figure. 8. Theoretical curve of amplitude response of high frequency circuit

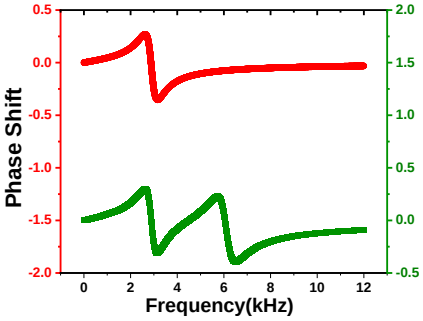


Figure. 9. Ideal amplitude-angle response curve of high frequency circuit

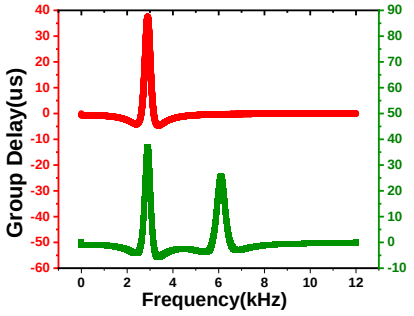


Figure. 10. Ideal delay curve of high frequency circuit group

B. Simulation test

Multisim simulation software is selected for simulation test. The software has the function of simulating various components, which can choose whether to produce noise or not. The built-in function generator and oscilloscope also meet the experimental requirements. In this study, sinusoidal pulse and modulated pulse are used as output pulses, and experiments are carried out in both low frequency and high frequency circuits. The low frequency circuit tests sine and modulation. The first order high frequency circuit tests the modulated pulse, and the second order high frequency circuit tests the sinusoidal pulse.

a) Low frequency circuit test

Figure. 11 shows the construction of the low-frequency working circuit in the simulation software, Figure. 12 is the screenshot of the 0.5Hz sinusoidal pulse used for testing, and Figure. 13 is the 1Hz carrier modulated pulse used as the simulation result of the test signal, where red is the input pulse and green is the output pulse. It can be found that the sinusoidal pulse has obvious phase shift phenomenon, and the envelope group delay phenomenon of the modulated pulse is also obvious.

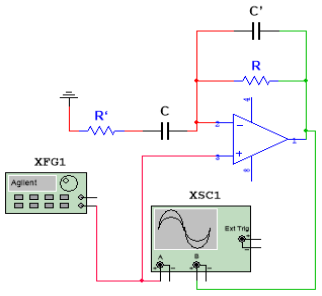


Figure. 11. Low frequency working circuit in Multisim analog system

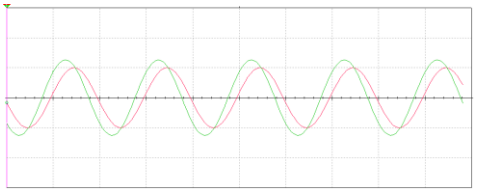


Figure. 12. Test result of sinusoidal pulse of certain frequency in low frequency circuit

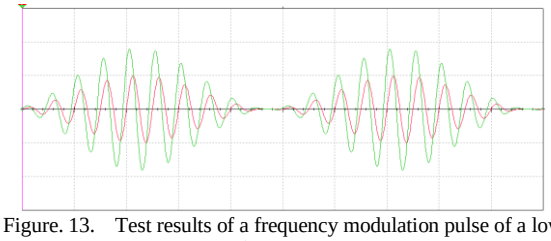


Figure. 13. Test results of a frequency modulation pulse of a low frequency circuit

Simulation software is used to measure the

pulse conditions under different pulses and measure the relevant data, and compare with the theoretical data.

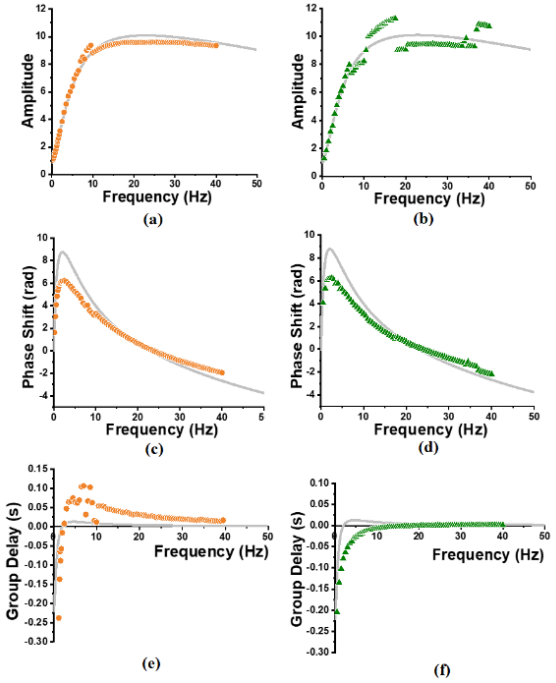


Figure. 14. Low frequency circuit test results

Measured from 0Hz to 40Hz, the group delay under each measurement frequency was obtained, and compared with the theory, plotted in FIG. 14, where the gray curve is the theoretical curve, the orange dot is the sinusoidal pulse test result, and the green triangle is the modulation pulse test result. (a) and (b) represent amplitude conditions, (c) and (d) represent phase shift conditions, and (e) and (f) represent group delay conditions. Through analysis, it can be found that the overall data trend is in line with the theoretical reasoning, and the error is also within the acceptable range. The main reasons for the errors are the sampling rate of the test software (affecting the program fitting) and the influence of the test frequency step size (especially the sine pulse needs to be differentiated).

b) High frequency circuit test

The test of high-frequency circuit shows the input and output of sinusoidal pulse and modulated pulse in the simulation software, in which the gray curve is the theoretical curve, the green triangle is the modulation pulse test first-order high-frequency RLC amplifier circuit, and the orange dot is the sinusoidal pulse test second-order high-frequency RLC amplifier circuit. Select some feature points in 0Hz-80000Hz for testing. Due to the sampling accuracy, test frequency compensation, fitting error and other problems, there are some errors, the final test results are in line with the theoretical situation on the whole.

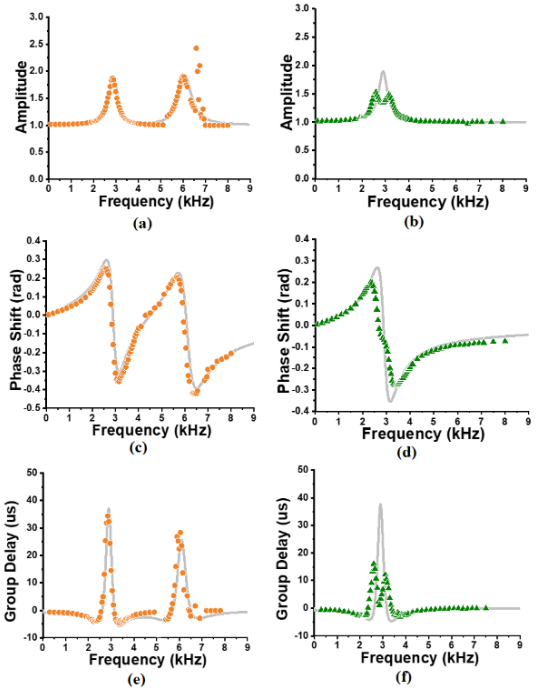


Figure. 15. Test results of high frequency circuit

When we used the Multisim simulation software for simulation test, the problem of image "breakdown" occurred many times, which was manifested as interruption of the output pulse curve, as shown in Figure 16. The reason was that the chip voltage was not enough to support or was related to frequency distortion. When the input signal of the amplifier circuit is a multi-frequency signal, if the amplifier circuit has different gain amplitudes for different frequency components of the signal, it will lead to distortion of the output waveform, which is called amplitude distortion; If the relative phase shift changes, it is called phase distortion, and the two are collectively referred o as frequency distortion. Frequency distortion is caused by the linear reactance element of the circuit, and frequency distortion is characterized by the absence of new frequency components in

the output signal that are not present in the input signal.

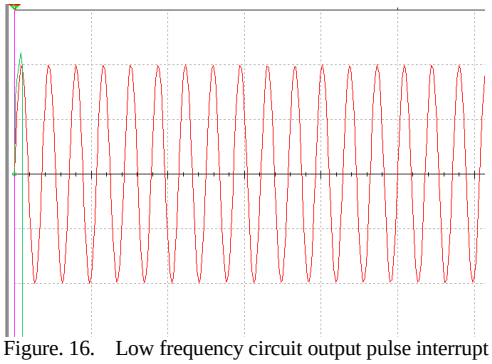


Figure. 16. Low frequency circuit output pulse interrupt

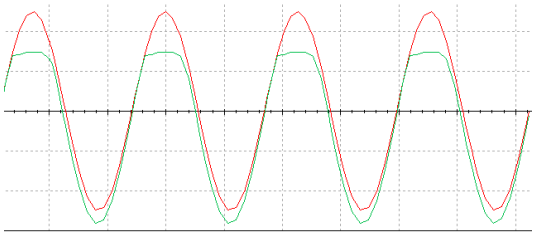


Figure. 17. Low frequency circuit pulse peak missing

In the simulation operation, the simplest way to solve this problem is to lower the input pulse amplitude. Then by adjusting the scale of the channel to flexibly control the view window to observe whether the output pulse is complete, if the output pulse waveform is complete, but the shape near the peak is obviously "decapitated", you need to continue to adjust the input pulse amplitude, until the input and output pulse are intact can continue to complete the experiment.

V. Conclusion

In this paper, the concepts of negative group delay and superluminal phenomenon are introduced. By using the principle and simulation test of RC negative group delay circuit, the superluminal phenomenon under negative group delay can be realized. In this paper, the principle of negative group delay and super-lumen phenomenon is analyzed and elaborated, so as to explore the negative group delay related technology at a higher level. Finally, through the simulation test of electronic circuit, the expected results are obtained. In the follow-up research work, we will continue to carry out theoretical exploration, plan better test methods, try more component parameters to meet more working bandwidth, and further in-depth study on negative group delay and super-lumen. The exploration and research of superluminal and negative group delay is of great significance in today's rapidly developing information age. In recent years, with the development of communication and space exploration, the transmission capacity and channel transmission characteristics of signals are increasingly required. Group delay is a very important parameter in the process of signal transmission. The study of group delay and superluminal speed is of great significance in space testing, navigation system, military positioning and so on.

There are many possibilities in the field of negative group velocity research. It is not a single discipline study, but a multi-discipline study. Interdisciplinary research can bring more vitality and possibilities to the subject. Low band circuits try to use visual LED demonstration to show the results of group speed control. In the future, there is also a lot of research space by deep learning, continuous system simulation and other methods.

Acknowledgment

This research is partially funded by the Project funds in Xi'an University Student Innovation and Entrepreneurship Fund Project(X202210702179).

References

[1] Kitano M. Nakanishi T. Sugiyama K. Negative group delay and superluminal propagation: an electronic circuit approach [J]. IEEE journal of selected topics in quantum electronics: A publication of the IEEE Lasers and Electro-optics Society, 2003, 9(1). [2] Chen Yu. Study on the relationship between group velocity and energy velocity [J]. Science in China (Series A), 1995, 25(3):261 -- 267.

[3] Fei Zhaoyu, Ma Yuhan, Sun Changpu. Phase velocity and group velocity superluminal problem to differentiate [J]. Journal of modern physics, 2022 (01): 3-7, DOI: 10.13405 / j. carol carroll nki XDWZ. 2022.01.016.

[4] Chen Wen-bing. Review of research on negative group delay circuits: Characteristics, present situation and Prospect [J]. Telecommunication Technology, 2017, 57(03):369-376.

[5] Chen Yu. Research on the relationship between group velocity and energy velocity [J]. Science in China (Series A Mathematical Physics Astronomy Technical Sciences), 1995(03):261-267.

[6] ZHONG Caichi. Phase Velocity, Group Velocity and super-luminization [J]. University Physics, 1985(10):6-18. DOI:10.16854/j. cnki.1000-0712.1985.10.005.

[7] ZHANG Junhui. Research on Super-lumen Velocity Problem [D]. Northeastern University, 2010.

[8] Huilin Mao. Studies on Group Delay and Signal Fidelity in Negative Group Delay Circuit [D] Zhejiang University, 2019.

[9] Huiling Mao, Linhua Ye, Li-Gang Wang. High fidelity of electric pulses in normal and anomalous cascaded electronic circuit systems [J]. Results in Physics, 2019, 13.

[10] Brillouin L. Wave propagation and group velocity. New York Academic Press, 1960 [10] M. Kitano, T. Nakanishi, and K. Sugiyama: Negative Group Delay and Superluminal Propagation: An Electronic Circuit Approach.

[11] Han Shen, Zhongsheng Wang. A Circuit Principle and Simulation Test for Negative Group Delay. International Journal of Advanced Network, Monitoring and Controls, 2022, 7(2):46-57.

[12] M. Kitano, T. Nakanishi, and K, Sugiyama. Negative Group Delay and Superluminal Propagation: An Electronic Circuit Approach. IEEE Journal of Selected Topics in Quantum Electronics, 2003, 9 (1):43-51

[13] T. Nakanishi, K. Sugiyama, and M. Kitano. Demonstration of negative group delays in a simple electronic circuit. Am. J. Phys. 70(11), November, 2002:1117-1121

[14] Hua Cao, Arthur Dogariu, and L. J. Wang, Negative Group Delay and Pulse Compression in Superluminal Pulse Propagation. IEEE Journal of Selected Topics in Quantum Electronics, 2003, 9 (1):52-58

[15] Brillouin L. Wave propagation and group velocity. New York: Academic Press, 1960.

[16] Morgan W. Mitchell, Raymond Y. Chiao. Negative group delay and "front" in a causal system: An experiment with very low frequency bandpass amplifiers. Physics Letters A,230(1997):133-138.