Table of Contents

Research on Commodity Mixed Recommendation Algorithm

Chang Hao
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 271203550@qq.com

Yang Shengquan
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 1355593006@qq.com

***Abstract*—With the advent of the era of big data, our lives generate huge amounts of data every day, and the field of e-commerce is no exception. It is particularly important to analyze these data and recommend products. It is reported that through the recommendation algorithm, Amazon has increased its sales by about 30%. Among the recommended algorithms, the collaborative filtering algorithm is currently relatively mature and has achieved very good results in various fields. But the traditional collaborative filtering algorithm is too rough when calculating the similarity and prediction score, and the efficiency is very low. We combine the traditional collaborative filtering algorithm with the decision tree algorithm, and improve the traditional recommendation algorithm, create a collaborative filtering decision tree algorithm to recommend products, and run the new collaborative filtering decision tree algorithm on the Hadoop platform on. Experiments show that the improved algorithm makes the accuracy of recommendation significantly improved.**

Keywords-E-Commerce; Recommendation Algorithm; Decision Tree; Collaborative Filtering

I. INTRODUCTION

With the development of science and technology, Internet technology has also rapidly developed and popularized, so that the data on the network is growing at the level of PB every day, bringing a lot of information resources to users and greatly enriching people's daily lives. However, the problem of rapid expansion of a large number of information resources has also emerged. "Information overload" is also a problem that Internet users are facing. "Information overload" refers to the difficulty for Internet users to accurately and quickly locate the information they need from massive data [1]. The generation of recommendation system has greatly eased the problem of "information overload". The recommendation system is automatic and intelligent to recommend items for users, and it will dynamically adjust the recommended item types according to the changes of user behavior, which truly avoids the "information overload" problem.

Faced with such a huge amount of data, it is necessary to adopt a big data model for analysis. Compared with the traditional data model using random analysis (sampling survey), the big data model analyzes all data and has the characteristics of 4V, Namely Large Volume, High Speed, Variety, Value. Collaborative filtering algorithm is one of the most concise and practical recommendation algorithms. If you use traditional data model for sampling survey, it will inevitably aggravate the sparsity problem of the algorithm itself, so it is of great significance to design a big data model based on collaborative filtering. And necessary. If you want to process big data, a single computer cannot be realized, so the application of distributed architecture is particularly important. So the algorithm model is run under the Hadoop distributed framework. MapReduce is a distributed computing framework under Hadoop [2]. It uses the "divide and conquer" idea to decompose complex tasks or data into several simple tasks for parallel processing. Afterwards, it performs global summarization, which greatly improves the efficiency of the algorithm. This article mainly studies the distributed recommendation algorithm under the Hadoop platform. The recommendation algorithm combines the decision tree and the collaborative filtering algorithm, and improves the traditional collaborative filtering algorithm to improve the timeliness of recommendation.

II. INTRODUCTION TO RELATED TECHNOLOGIES

A. Introduction to the traditional collaborative filtering algorithm

Collaborative filtering algorithm is the most successful information filtering algorithm used in the

DOI: 10.21307/ijanmc-2020-021

current recommendation system. The main method is to extract the historical behaviors generated by users to make recommendations. The traditional collaborative filtering algorithm is mainly divided into item-based collaborative filtering algorithm (ItemCF) and user-based collaborative filtering algorithm (UserCF) [3]. The core process of collaborative filtering algorithm is as follows: Collect user preferences, find similar users or items, and calculate recommendations, the core of which is the calculation of similarity Euclidean distance similarity method (Formula 1), Pearson correlation coefficient similarity method (Formula 2), Salton similarity method (Formula 3) and Cosine similarity method (Formula 4) are several common similarities Calculation method [4].

$$O(x,y) = \sqrt{(\sum(x_i - y_i)^2)} \quad (1)$$

$$P(x,y) = \frac{n\sum x_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}} \quad (2)$$

$$S(x,y) = \frac{|N(u) \cap N(v)|}{\sqrt{|N(u)|} \cdot \sqrt{|N(v)|}} \quad (3)$$

$$COS(x,y) = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}} \quad (4)$$

B. Introduction to Decision Tree

It is a typical classification algorithm. It first processes data, uses inductive algorithms to generate readable rules and decision trees, and then uses decisions to analyze new data. After getting the recommended products in the previous step, extract the features [5]. A data set will extract many features, but which feature is selected as the root node? Which feature is selected as the optimal solution? At this time, we need to introduce a new measure-entropy. Entropy refers to the uncertainty of random variables. The algorithm for calculating their entropy values is shown in Formula 5:

$$H(x) = -\sum pi * \log pi, i = 1,2, \dots, n \quad (5)$$

Where pi represents the probability of each feature, it can be seen from the formula that when the probability is greater and the purity is greater, the entropy value will be smaller, and the smaller the entropy value, the more stable the feature. There are three selection criteria for features: information gain (ID3), information gain rate (C4.5) and Gini index (CART). The development of decision trees is getting faster and faster, and many excellent algorithms have been derived, such as GDBT (Gradient Boosting Decision Tree), RF (Random Forest random forest), etc, the decision trees have considerable advantages in terms of accuracy improvement, they are used frequently in various games, and the effect is very good. At some times, the neural network that has been painstakingly built is not as accurate as the relatively simple random structure. The forest is high, so this article will use random forest to optimize the algorithm.

III. RECOMMENDED ALGORITHM DESIGN

The collaborative filtering algorithm is the most widely used algorithm in the recommendation system. Because of its versatility of the model, it does not require too much expertise in the corresponding data field. The engineering implementation is relatively simple, and the effect is also good. It is widely praised by all walks of life. But collaborative filtering has its own problems, such as "cold start" and "sparseness" has always been a problem of collaborative filtering itself. Therefore, in the context of today's big data era, collaborative filtering may not be suitable for direct recommendation algorithms. To solve this problem, this paper designs a hybrid recommendation algorithm.

A. Random Forest random forest

Because a single decision tree has obvious drawbacks in the recommendation system, a random forest composed of multiple decision trees can effectively solve the problems of a single decision tree. In essence, the decision tree is a special tree because it contains many judgment nodes, and the model of the random forest is composed of multiple decision trees[6].

The main idea of the random forest algorithm is to use some single classifiers to form a large classifier, which mainly includes the following three steps:

Step 1: Randomly sample multiple decision trees that need to be generated. As many decision trees as needed, as many training subsets as there should be. This process involves a statistical sampling method, which is to extract several training subsets from the original training set. The sampling method to be used in this thesis is a sampling method based on Bagging thought. When the training set is extracted, it can be sampled repeatedly to ensure that the chance of the sample being selected is random and the probability is equal.

Step 2: how to build the decision tree. The decision trees constructed from several training subsets selected in the first step are the main elements that constitute a random forest. The construction of these trees is not restricted by any factors, and does not do any pruning operations on the trees. When constructing a decision tree, not all attributes in the data set are selected as indicators for calculation, but are divided into several randomly selected "optimal" feature attributes, Then decompose according to the eigenvalue of k<K. In the decision tree, the C4.5 algorithm is used as the splitting algorithm for attribute selection, and the information gain rate algorithm is given. According to the randomness peculiar to the random forest, first select k attributes as the features of the decision tree. These features all act as classifiers. From the calculation of the training set, we can know the classification standard $h(x,\theta) \in (0,1), x \in R^N$ is a randomly selected training sample. $\theta = (\alpha,\ \varphi)$ represents the parameter of this node, φ represents the matrix, α represents the filtering function, and the surface style of the node feature is determined by α, The Formula 6 represents the calculation of the nonlinear plane, and the calculation formula of the linear plane is Formula 7:

$$h(x,\theta) = \delta(\alpha^T(x)\varphi > 0) \quad (6)$$

$$h(x,\theta) = \delta(\alpha^T(x)\varphi\alpha(x) > 0) \quad (7)$$

Use a recursive method to operate on the data set until the data on a node has all belonged to the same type of feature or the number of data sets on the node has reached the threshold set in advance, then this node will stop continuing to classify, Converted to a leaf node. If the above requirements are not met, the node will continue to randomly search for feature attributes for classification.

Step 3: the formation of the forest. After repeating the first step and the second step several times, the resulting trees can be used to build random forests. First of all, according to the function of these trees, you can classify the training set, integrate the results of the data set processed by the decision tree, and vote. The final output of the algorithm is the result of the classification with the most votes.

B. AHP model

AHP (Analytic Hierarchy Process) is a method for making decisions based on the weight of layers. This method through further exploration and analysis of the root of more complex problems and their influencing factors, further proposed a qualitative method to quantify the problem, so as to provide more detailed quantitative information for decision-making. In this research, we adopt the method of analysing the weight of influencing factors in the analytic hierarchy process, and give corresponding weights to different operation behaviors, and then determine the similarity between different users and the similarity between different brands according to the obtained weights degree. The following is the specific process of calculating the weights:

Step 1: decision-level analysis. First of all, through in-depth study of the problem, in the research process, the factors that have related relationships are analyzed and compared with each other, and each factor is arranged in layers to form a multi-layered hierarchical structure model. Through the analysis, we can know that the following four factors have the most impact on the user's future purchases: the number of user clicks on the product, whether the user has added the product to the shopping cart, whether the user has collected the product, and whether the user has purchased the product. Formed a structural model as shown below:

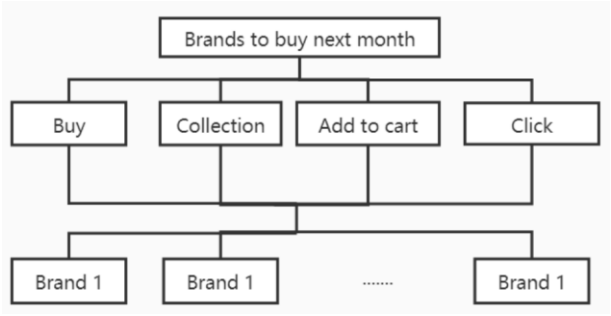


Figure 1. Hierarchical model diagram

Step 2: Construct the judgment matrix. First of all, it is necessary to compare all the influencing factors with each other. In the measurement, the introduction of relative scale is used to reduce the difficulty of comparing two different factors with each other, which further improves the accuracy.

If you want to compare the influence of n elements $A_1, A_2, \ldots, A_n$ on the same goal, you will obtain two factors A_i and A_j each time. c_{ij} represents the ratio of the influence level of A_i and A_j on the goal. All the comparison results are also can be written as a comparison matrix:

$$C = (c_{ij})n * n,\ \ c_{ij} > 0,\ \ c_{ij} = \frac{1}{c_{ij}} \quad (8)$$

The larger the value of c_{ij}, the higher the importance of A_i relative to A_j. In general, the differences between these factors are presented on a scale of 1-9.

Step 3: Solve and test. The elements of W are the ranking weights of the relative importance of the factors at the same level to the factors at the previous level. This process is called hierarchical single sorting. Then, whether the hierarchical single sorting can be confirmed requires a consistency test. The so-called consistency test is Refers to the allowed range of inconsistency determined by the pairwise comparison matrix. The consistency index can be defined as: $CI = \frac{\lambda_{max}-n}{n-1}$. Different CI values represent different meanings, the greater the CI, the more serious and obvious the inconsistency, the consistency ratio is defined as $CR = \frac{CI}{RI}$. RI is defined as a random consistency indicator, its value is shown in the table below:

TABLE I. RI VALUE TABLE

n	1	2	3	4	5	6	7	8	9	10	11
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51

C. Improved collaborative filtering algorithm

Because Random Forest also has its own drawbacks, it can only recommend brands that have been in contact with users, and those brands that have not been in contact with users, Random Forest will not recommend it. Even the brand may meet the needs of users. At this time, AHP improved collaborative filtering algorithm is proposed to solve this problem [7]. Improve through the following steps.

Step 1: use the AHP model to find the weight of user behavior. Because when calculating user similarity and brand similarity, collaborative filtering recommendation algorithm cannot treat all user interactions equally. Therefore, the AHP model is used to assign values to the behaviors of users. With this set of weights, the similarity between the user's image and the brand's image can be calculated, which solves the drawback of collaborative filtering recommendation. This set of weights can be trained using the AHP analysis model introduced in the previous section.

Step 2: The user's rating data for the brand. The operation of scoring products on the e-commerce website, and the size of the rating also represents the user's love for the brand, Therefore, before calculating the user's similarity and brand's similarity, you first need to calculate the user's rating value for the brand. You can obtain the user's rating value for the brand by calculating the user operation type and frequency. The calculation formula is as follows:

$$R_{u,i} = \sum_{c=1}^{C} Op(c)Fp(c) \tag{9}$$

Among them, u represents the user, i represents the brand, then $R_{u,i}$represents the user's rating of the brand, $Op(c)$ represents the weight of the user's operation type c, and $Fp(c)$ represents the user The frequency of operations on the brand. From this we can get the matrix that $R_{u,i}$ can be composed of: $\begin{bmatrix} R_{11} & R_{12} & \dots & R_{1n} \\ R_{21} & R_{22} & \dots & R_{2n} \\ \dots & \dots & \dots & \dots \\ R_{m1} & R_{m2} & \dots & R_{mn} \end{bmatrix}$.

This is a matrix reflecting u and i interacting with each other. Based on these, the similarity between users and brands can be calculated.

Step 3: Calculation of similarity. This is the most important link in the collaborative filtering recommendation algorithm. This article uses an attribute-based similarity algorithm. The similarity is to process the information of the user and its nearest neighbors, so the core part of the algorithm is that if the nearest neighbor is obtained, we conduct relevant research and analysis on it, and obtain the following method of the user's comprehensive similarity.

The following is the formula for the user's rating similarity $YSD1(u, i)$:

$$YSD1 = \frac{\sum_{t\in I_{u,i}}(R_{u,t}-A_u)*(R_{i,t}-A_i)}{\sqrt{\sum_{t\in I_u}(R_{u,t}-A_u)^2 * \sum_{t\in I_i}(R_{i,t}-A_i)^2}} \tag{10}$$

This formula compares the similarity of users u and i. $R_{u,t}$ and$R_{i,t}$each represent the rating value of users u and i on brand t, and the set of brands that user u has rated respectively represent For I_u. Similarly, all brands rated by user i. are denoted as I_i, and the intersection of those users who have shared a rating is $I_{u,i}$; the average value of user u's rating in I_u is set to A_u, similarly, in I_i The average value of user i score is set to A_i.

The following formula represents the user's feature similarity YSD2 algorithm:

$$D(u,i) = \sqrt{\sum_{k=1}^{n}(u_k - i_k)^2} \quad (11)$$

$$YSD2(u,i) = \frac{1}{1+D(u,i)} \quad (12)$$

Formula 12 shows the weighted Euclidean Metric of users u and i, that is, the Euclidean distance, n represents the feature dimension of the user, and the k-th eigenvalues of users u and i will be represented by u_k and i_k, respectively. YSD2 illustrates the calculated similarity of features of users u and i.

Based on the above calculation formula to obtain the user's similarity score YSD1 for the product and the user characteristic similarity YSD2 , the following formula is used to calculate the user's comprehensive similarity YSD:

$$YSD(u,i) = w_1 * YSD1(u,i) + (1 - w_1) * YSD2(u,1) \quad (13)$$

In these indicators, w_1 is the combined weight of the user's comprehensive similarity. The actual value of the combined weight of the user's comprehensive similarity is determined by the degree of influence of the score similarity and the feature similarity on the user's comprehensive similarity.

The calculation of brand similarity is the same as the calculation of user similarity. You only need to change the parameters, which will not be described in detail here.

Step 4: the selection of the nearest neighbor. In order to achieve the purpose of accurate recommendation, it is necessary to accurately target the other neighbor users that match the user's interests, so the selection of the nearest neighbor is very important. Select the user's nearest neighbor and the brand's nearest neighbor, this article uses the Top-N method. The first step of this method is to calculate the similarity between other users, brands and target users, brands, and then sort the calculated similarity values.

Step 5: Generate recommendations. Using the method described in step 3, we can obtain the nearest neighbor set NU[u] of the target user, and then we recommend the user according to the formula listed below:

$$PU(t,u) = A_u + \frac{\sum_{i=1}^{c}(R_{i,t} - A_i) * YSD(u,i)}{\sum_{i=1}^{c} YSD(u,i)} \quad (14)$$

A_u is the average score of user u for all brands in the dataset, the value of c is the number of users in the nearest neighbor of user u., $R_{i,t}$ is the nearest neighbor user i. of user u, For the rating made by the brand t, $YSD(u,i)$ represents the user's overall similarity. PU means that the recommendation degree of this brand is recommended to the user u based on the result of the user recommendation.

The brand-based recommendation idea is similar to the above, so I won't elaborate more.

D. Fusion of the two algorithms

After introducing the basic characteristics of the random forest recommendation algorithm and the collaborative filtering recommendation algorithm, the parameters and the calculation process required, we use the characteristics of each other. For example, the random forest model can recommend brands that users have interacted with before, while the collaborative filtering algorithm recommends brands that have not interacted with users. Therefore, if the two models can be perfectly combined, the common advantages of the two algorithms can be combined, which plays a role in promoting strengths and avoiding weaknesses. All the information required for user recommendation can be included. Naturally, better recommendation results will be obtained. The model the accuracy is also more improved, as shown in detail in Figure 2 describes the fusion strategy of the two models:

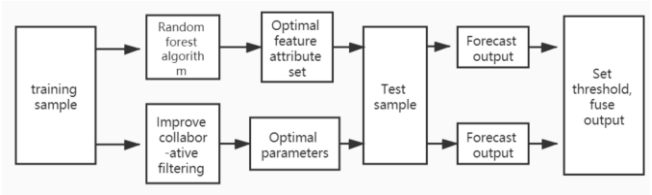


Figure 2. Schematic diagram of algorithm fusion process

IV. EXPERIMENT AND RESULT ANALYSIS

A. Experimental data and experimental environment

In order to verify the efficiency of the improved algorithm model, we selected an e-commerce company's internal data set for testing, and analyzed and evaluated its performance. The data set contains all the behaviors of about 100,000 random users with behaviors within one week of the e-commerce company. Among them, user behaviors include clicks, purchases, favorites, and purchases. The data set contains users 99799, the number of goods 416202, the number of all actions 10015080, such a huge data set, if

you use traditional stand-alone operations, the time consumption is immeasurable, so using the Hadoop distributed system in a big data environment to perform distributed calculations on its data greatly improves the efficiency of the operation.

The experimental operating environment is a Hadoop cluster. The cluster has one Master node and three Slave nodes, and all machines in the cluster have the same configuration. The cluster installation is configured under the CentOS-6.7 operating system, and the JDK1.8 environment is configured for both CentOS and Windows, and the code is compiled into the IDEA compiler on the Windows side.

B. Algorithm evaluation index

1) Recall rate

The number of items in the recommended list calculated by the algorithm is what consumers like, which is the recall rate of the algorithm. Formula 15 describes how to calculate the recall rate:

$$R(L_u) = \frac{L_u \cap B_u}{B_u} \quad (15)$$

Among them, the item that user u likes is L_u, and the recommendation algorithm lists the product recommended by the consumer as B_u.

2) Accuracy

The accuracy of the algorithm tests the ratio of the items in the recommended list given by the system to the items that consumers actually like. Formula 16 describes how to calculate the accuracy:

$$P(L_u) = \frac{L_u \cap B_u}{L_u} \quad (16)$$

Among them, the item that user u likes is L_u, and the recommendation algorithm is denoted as B_u by the consumer's recommended product list.

3) F measure

As shown in Formula 17:

$$F = \frac{2*P*R}{P+R} \quad (17)$$

Since there is a negative correlation between the accuracy rate and the recall rate, it is necessary to fit the F measure, and the F score will prevail. As can be seen from the Formula 18, the prediction result hopes to cover more users and brands while ensuring accuracy.

C. Random forest model experiment results

There are two parameters that can be controlled by the random forest model, one is the number of random forest decision trees, and the other is the number of feature attributes that are randomly extracted to build the decision tree. The number of decision trees in the random forest model is K_{RF}, and the selected parameters are K_{RF}=100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, and the following experimental results can be obtained:

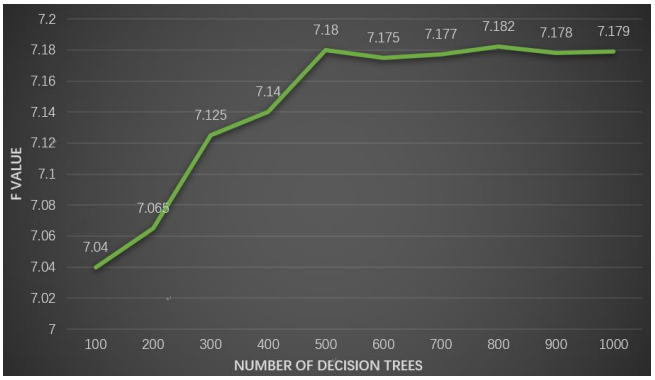


Figure 3. The relationship between decision tree and result in random forest model

It can be seen from the experimental result graph that the F value rises with the increase of the decision tree. After rising to 500, it basically tends to be gentle, so K_{RF} is taken as 500.

Also for the eigenvalues of random forests, not as many as possible, in order to ensure the stability of the model, through experiments, the following results are obtained:

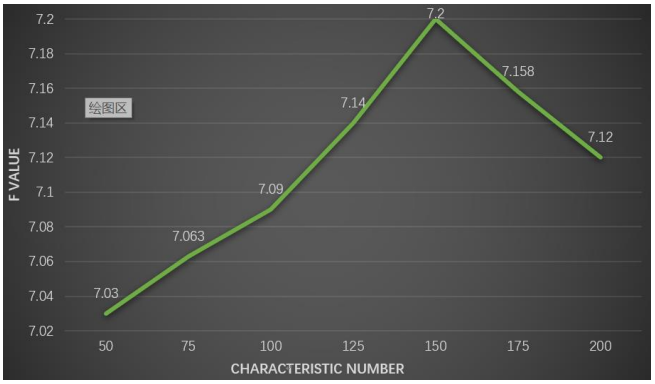


Figure 4. The relationship between the number of features and the results in the random forest model

It can be seen from the above experiment that when the number of decision trees is 500 and the number of features is 150, the random forest model has the best effect, that is, the F value is the highest.

The final experimental results of the random forest model are shown in the table below.

TABLE II. RANDOM FOREST MODEL FINAL EXPERIMENTAL RESULTS

Accuracy	Recall rate	F measure
7.19	7.21	7.20

D. Experimental results of the hybrid algorithm

The weight of user behavior in the improved collaborative filtering algorithm can be calculated according to AHP, as shown in the following table：

TABLE III. USER BEHAVIOR WEIGHT

Interaction type	weight
Click	0.08
Collection	0.12
Add to cart	0.30
Buy	0.50

The number n of brands recommended by the collaborative filtering algorithm is used as the parameter for the experiment. We take n = 600, 800, 1000, 1200, 1400, and 1600 to conduct the experiment. The experimental results are shown in the figure.

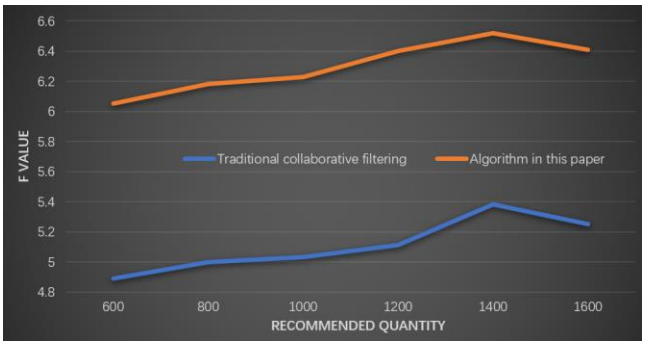


Figure 5. The relationship between the recommended number of collaborative filtering algorithms and the result

It can be seen from the experimental results that with the increase in the number of recommendations, the algorithm's recommendation performance is improving. When the number of brand recommendations is close to 1400, the recommendation effect is the best.

The final experimental results of the fusion of the final collaborative filtering recommendation algorithm and the random forest algorithm are shown in the following table:

TABLE IV. EXPERIMENTAL RESULTS OF FUSION ALGORITHM

Accuracy	Recall rate	F measure
7.33	7.42	7.35

Due to the large number of data sets in this experiment and the time-consuming operation in a single machine, the above operations are performed under the distributed system Hadoop. In order to compare the advantages of Hadoop in data processing speed, the collaborative filtering and the design The algorithm and the processing time of the algorithm designed in this paper are compared in the Hadoop environment, as shown below:

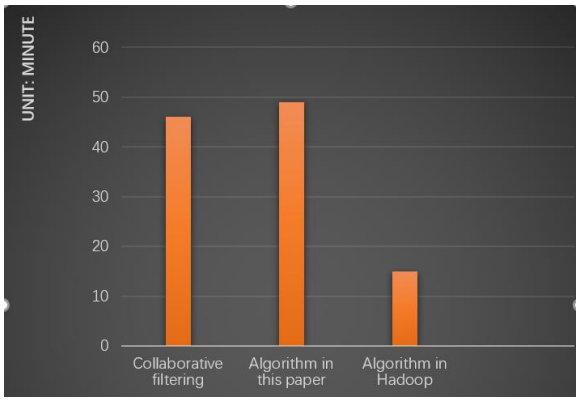


Figure 6. Time comparison chart

Because the algorithm in this paper is more accurate than the traditional collaborative filtering algorithm and the calculation is relatively complicated, the time efficiency is slightly insufficient. However, if it is run in a Hadoop distributed cluster environment, the time efficiency is effectively improved by nearly 3 times. It can be seen that the current big data In a large environment, it is necessary to use Hadoop distributed clusters to process data.

V. CONCLUSION

This paper attempts a model that uses different prediction and recommendation methods for prediction and recommendation, namely the random forest model, and gives a detailed introduction and further analysis of this model. This article also gives a detailed

introduction to the basic principles of the traditional filtering algorithm of collaborative filtering, and thoroughly analyzes the advantages and disadvantages of the algorithm. For example, this traditional collaborative filtering algorithm lacks the ability to calculate the brand score of users. Personalized investigation, treat all user behavior as the same. In view of the limitation of this traditional algorithm, this paper made some necessary improvements, and proposed the optimization of collaborative filtering similarity based on the weight of AHP. In addition, from two perspectives, user interaction and brand interaction, this article randomly integrates the collaborative filtering model with the random forest model. This will result in more accurate recommendation results, and the recall rate will naturally increase. Finally, the data is analyzed through real data cases to obtain reliable experimental results. The results show that the combination of this analytic hierarchy process and collaborative filtering algorithm makes the recommendation performance better than a single collaborative filtering algorithm. And after being fused with the random forest model, compared with the single random forest algorithm or collaborative filtering algorithm, the performance has been greatly improved.

REFERENCES

[1] Lu Xiaocui. The application of big data analysis technology in cross-border e-commerce[J]. Electronic Technology and Software Engineering, 2020 (01): 141-142.

[2] Tian Bin. Big data machine learning under the framework of distributed computing[J]. Electronic Technology and Software Engineering, 2019(20):168-169.

[3] Yang Wu, Tang Rui, Lu Ling. News recommendation method combining content-based recommendation and collaborative filtering[J]. Computer Applications, 2016, 36(02):414-418.

[4] Yang Hailong. Power recommendation system based on item-based collaborative filtering algorithm[D]. Lanzhou Jiaotong University, 2019.

[5] Sheng Wenshun, Sun Yanwen. An improved ID3 decision algorithm and its application[J]. Computer and Digital Engineering, 2019, 47(12):2943-2945+3094.

[6] Wang Jingna. Research on used car valuation model based on random forest algorithm[D]. Beijing Jiaotong University, 2019.

[7] Cui Yan, Qi Wei, Pang Hailong, Zhao Hui. A recommendation algorithm combining collaborative filtering and XGBoost[J]. Computer Application Research, 2020, 37(01):62-65.

Rheological Properties of Pullulan and Aloe Vera Nanofiber Solutions

Haifa El-sadi
Wentworth institute of technology
Boston, USA
E-mail: elsadih@wit.edu

Alex Bossart
Stevens institute of technology,
Hoboken, NJ

Sally Shady
Stevens institute of technology,
Hoboken, NJ
E-mail: sshady@stevens.edu

Dilhan Kalyon
Stevens institute of technology,
Hoboken, NJ

***Abstract*—Nanofibers have been used with increasing success for drug delivery and various biomedical engineering applications. The mixture of 13% pullulan-aloe vera can be pumped with a minor effects on the flow behavior. The fiber is then characterized by SEM, shear viscosity, storage and loss moduli. The oscillatory shear data is employed. The mixture has a complex rheological properties that includes extreme shear-thinning as well as viscoelastic properties and yield stress. The rheology of the pullulan- aloe vera nanofber was characterized, the measurement method may influence the results, it is unclear how the behavior near walls influence the measurement method, Parallel-plate rheometer is used to measure rheological properties. The data and rheological parameters should facilitate a better understanding of the process-ability characteristics of the mixture. The cross model provides a simple way of quantifying the viscosity/shear rate profile for a shear thinning mixture. SEM images are carried out.**

Keywords-Nanofiber; Rheology; Pulluan; Aloevera; Yield Stress

I. INTRODUCTION

Aloe vera is a therapeutic medicinal plant that has been used for many centuries to treat a wide range of ailments. There are over 250 species of the Aloe vera plant that have been studied. The most popular commercially grown are Aloe barbadensis miller and Aloe arborescens [1]. Many of the health benefits are associated with the structure and internal composition of the plant. Studies have shown that the active ingredients in the aloe vera plant have demonstrated wound healing, antifungal, anti-inflammatory, anticancer, immunomodulatory and gastroprotective properties [2-4].

Despite the exceptional healing properties of the aloe vera plant, there are several limitations with keeping these active ingredients stable. The bioactivity of the plant decreases approximately 6 hours after harvesting [1,5]. Exposure to light, humidity and temperature can also diminish these significant properties. Current techniques used to overcome these challenges have been focused on the usage of antioxidants and stabilizing agents. Such methods introduce other chemicals that would limit the scalability and purity of using non-toxic ingredients.

Pullulan is a linear polysaccharide produced by the polymorphic fungus Aureobasidium pullulans. Pullulan has a wide range of commercial and industrial applications in many fields such as food science, health care, pharmacy and even in lithography [6-8]. Pullulan is a water-soluble, low moisture permeable polysaccharide with excellent mechanical properties[9]. Due to these properties, pullulan is often used as a drug tablet coating. As a result, pullulan has been used in many drug delivery applications including nanotechnology [8, 10-12].

Nanofibers have been used with increasing success for drug delivery due to their extremely high surface to volume ratio, low density, and high porosity [reference]. In order to prevent aloe vera from

degrading, it can be electrospun into nanofibers to encapsulate, preserve and administer the unique properties of aloe vera. Pullulan, is a biodegradable polymer with adhesive properties that can has been shown success in the electrospinning [10]. In this study, we investigate the rheological properties of pullulan-aloe vera nanofibers, which will provide insight into the material properties in a scaled-up manufacturing process.

II. Materials and Method

A. Nanofiber preparation

A mixture of polymer Pullulan Pf-20 food grade (Nagase, Japan), Molecular weight 200,000 g/ mol and fresh aloe vera were used to create nanofibers. To prepare the nanofibers, first aloe vera leaves were first washed and their outer green rinds were removed to obtain the gel fillet. The gel was then broken down by using a conventional blender (Ninja BL490T) at an agitator speed of 21000 rpm. The procedure was performed within 60s to avoid enzymatic decomposition [13]. A homogenous gel mixture was then formed using 13% pullulan. The solution was then electrospun using a Harvard apparatus PHD 4400 syringe pump at 28 and 30 kV with a distance of 20 cm and a flowrate of 7.2 µl/mine..

B. Scanning Electron Microscopy (SEM)

A Zeiss Auriga Focused ION FE SEM (Oberkochen, Germany). All materials were coated with gold nanoparticles for one minute using a Leica EM MED020 (Wetzlar, Germany) under vacuum. Analysis was carried out at 2.00 kV EHT, and estimation of the dimensions of the fibers was performed using ImageJ and DiameterJ.

C. Rheological Properties

Prepared mixture was loaded between two 50mm-diameter parallel plates with a gap of 0.5mm in rheometer (ARES). High frequency and strain were chosen to make sure that enough torque was generated for the transducer. Time sweep test with frequency of 50rps and strain of 10% was performed to analyze the evaporation effect of the mixture to the result. Prior to dynamic property measurements, strain sweep test at constant frequency of 50rps determined the linear viscoelastic regions. Frequency range of 1-100rps and strain of 10% (within linear viscoelastic region) was performed in frequency sweep test at room temperature to obtain dynamic properties (G',G'') of the mixtures. G', the storage modulus which describes the elastic properties and G'', the loss modulus, which describes the viscous properties, were calculated by a computer program of the rheometer by deconvolution the torque (shear stress) versus time data.

III. Discussion

The homogenous solution of 13% pullulan in aloe vera were prepared to study the rheological properties of these nanofibers as shown in Figure 1. These properties will demonstrate the capabilities of the material solution in manufacturing processes such as 3D printing and extrusion. To our knowledge, this is the first study to prepare pullulan-aloe vera nanofibers from a fresh aloe vera plant. With the medicinal properties of aloe vera, pullulan is used as a carrier to help stabilize the bioactivity.

SEM images confirmed that pullulan- aloe vera nanofibers can be successfully developed as shown in Figure 2 and 3. The 28kV group produced 82 + 13 nm while the 30kV group was 99 + 21nm [10]. Fibers prepared at the 28kV setting produced nanofibers that were more consistent.

The storage (elastic) and loss (viscous) modulus of the mixture (pullulan and Aloe Vera) at 19oC for 3 s are increasing with the frequency as shown in Figure 4. The storage and loss modulus are almost constant at low frequency and increase significantly at higher frequencies which is related to the increase in the interaction between mixture components. It is observed that the viscous modulus is significantly larger than the storage modulus. It is very obvious that viscous modulus is higher than elastic modulus. Figure 5 shows the viscoelastic behavior of the mixture of pullulan and Aloe Vera, the complex modulus increases with increasing the frequency.

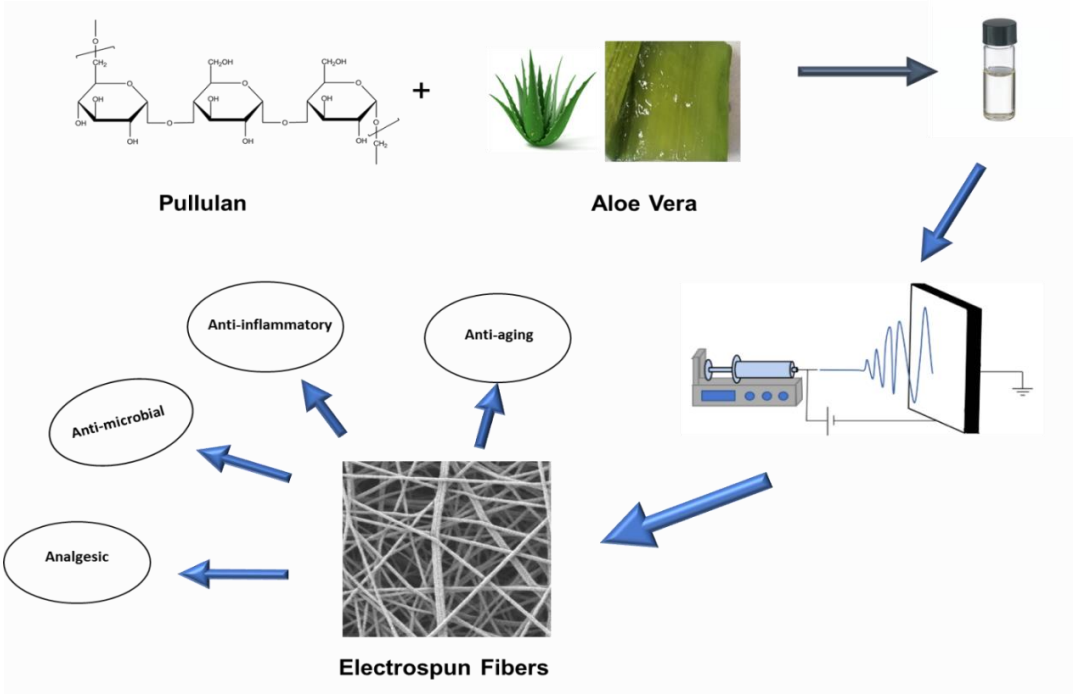


Figure 1. The overall process to develop the nanofiber demonstrating the multiple applications.

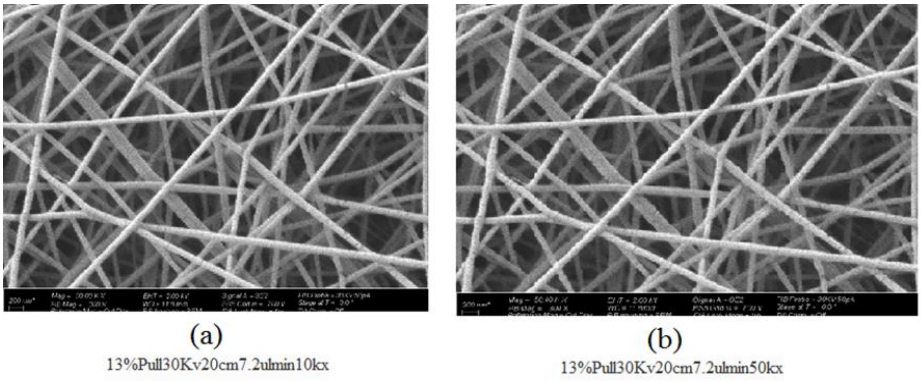


Figure 2. Nanofiber using PHD 4400 Syringe Pump at 30 kV (a) 10,000 times magnification, (b) 50,000 times magnification

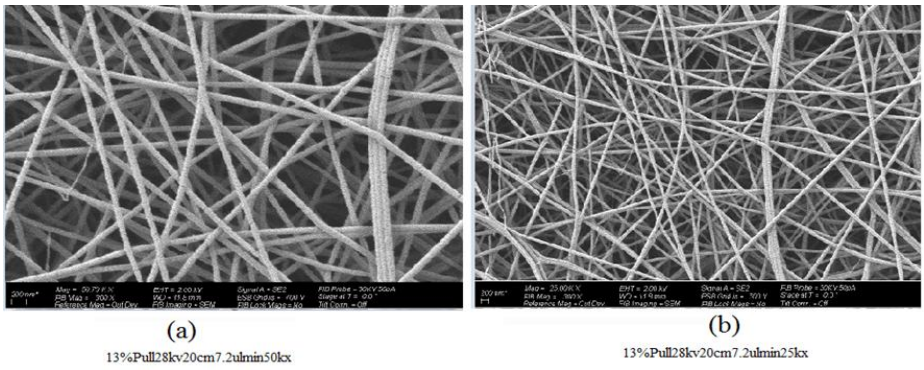


Figure 3. Nanofiber using PHD 4400 Syringe Pump at 28 kV (a) 50,000 times magnification, (b) 25,000 times magnification

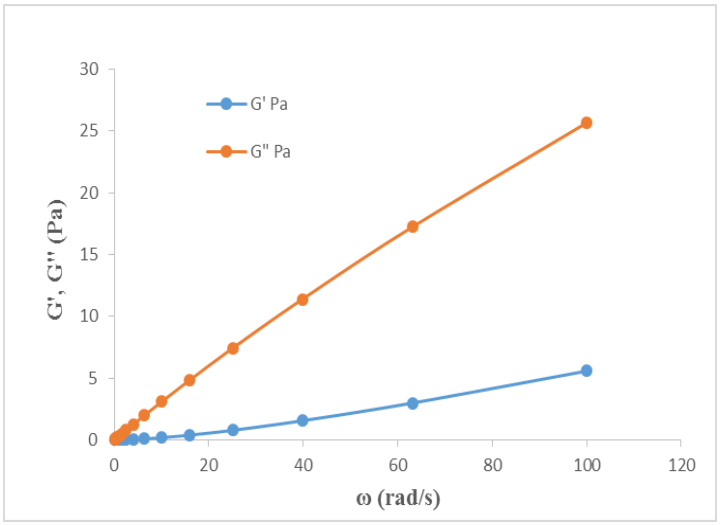


Figure 4. Elastic and viscous modulus of a mixture as a function of frequency at 19 oC.

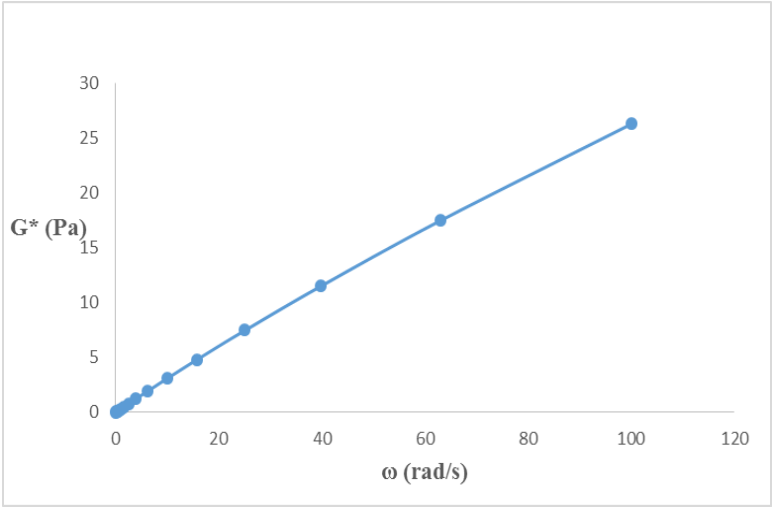
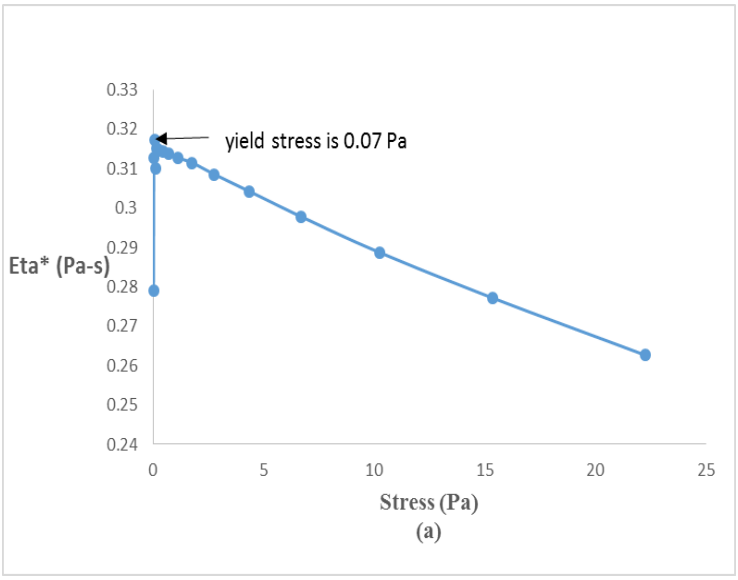


Figure 5. Complex modulus as a function of frequency at 19 oC

Figure 6 describes the changes in viscosity of the mixture of pullulan and aloe vera with respect to stress. The figure shows that the viscosity is decreasing with increasing the stress. Such viscosity decrease is directly related to the decrease of zeta potential. The yield stress is a key parameter in production, determining the force required to fill a product into its container or to flow. The stress at the viscosity maximum, is a measure for the yield stress. The yield stress is 0.07 pa which it shows that small amount of force is required to flow.



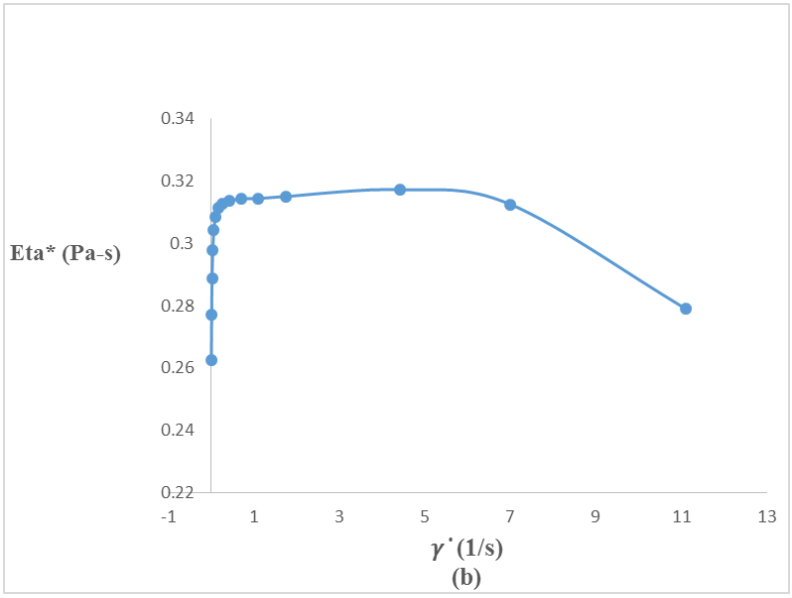


Figure 6. (a) Viscosity vs. stress, (b) Viscosity vs. shear rate

Cross rheological model shear thinning model gives a good agreement with measured data. $\dot{\gamma}$ is the shear rate, η_0, the zero shear viscosity and it is zero. m is the cross rate constant. It is dimensionless and is a measure of the degree of dependence of viscosity on shear rate in the shear thinning region. A value of zero for m shows Newtonian behavior and if m =1, it is due to increasing shear thinning behavior

$$\eta = \eta_\infty + \frac{\eta_0 - \eta_\infty}{1 + (c\dot{\gamma})^m} \quad (1)$$

$$\eta = \eta_\infty - \frac{\eta_\infty}{1 + (c\dot{\gamma})^m} \quad (2)$$

C is the cross time constant and has time unit. m and 1/C are related to the texture of the mixture, pumping and the characteristics of mixing and pouring of the flow. Figure 7 indicates that at C=50, the cross model fits the experimental data at shear rate above than 1.76 1/s. this might be due to the experimental error. Above C=50 the cross model isn't applicable for this type of mixture.

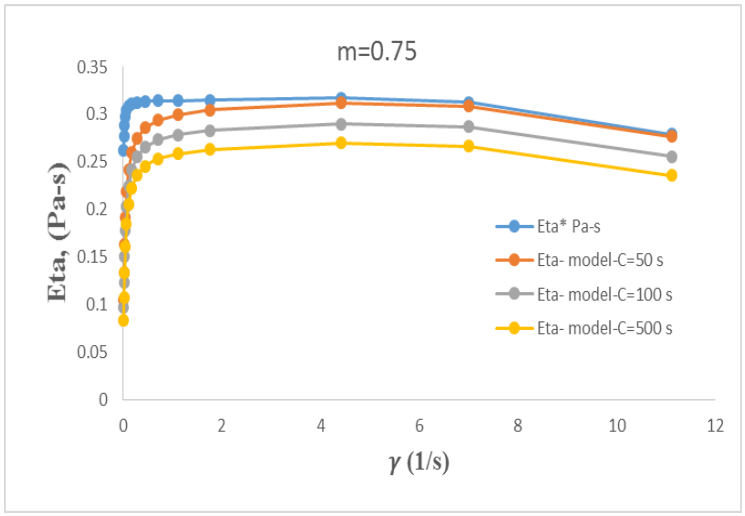


Figure 7. Validation of cross model with experimental data.

The rheometer was again used for stress relaxation experiments. The mixture of pullulan and Aloe vera is stable under temperature 19 oC as shown in Figure 8. However viscous modulus is higher than elastic modulus of the mixture.

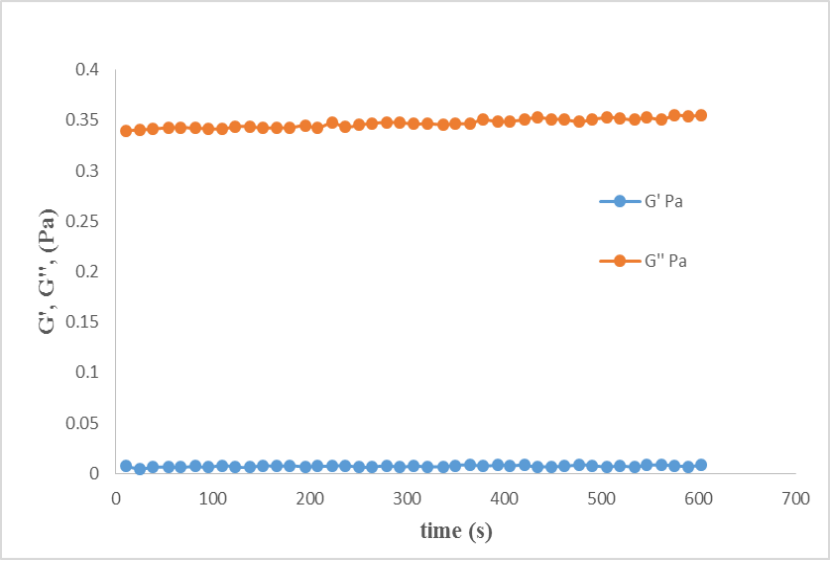


Figure 8. Storage and loss modulus vs. time

IV. Conclusion

The objective of this research is to study the flow behavior of the mixture of pullulan and aloe vera. Different samples of the mixture have been carried out. It is concluded that the mixture of 13% pullulan-aloe vera can be pumped with a minor effects on the flow behavior. The rheological properties of nanofiber at 50,000 times magnification using PHD 4400 Syringe Pump at 28 kV with a distance 20 cm were carried out. The yield stress is small since it is a key parameter in production, the mixture requires small force to flow. The viscosity decreases with increasing the shear rate. The cross model gives a good agreement with measured data at cross time constant 50 s. However, elastic modulus doesn't change with increasing the frequency. In the future, more rheological tests will be studied to indicate better predications of the fluids behavior.

REFERENCES

[1] Thorat, K. A review of the stability of aloe vera gel. Research J Topic and Cosmetic Sci.

[2] Torres- Giner, S. Nanoencapsulation of Aloe vera in synthetic and Naturally occurring polymers by Electrohydrodynamic processing of Interest in Food Technology and Bioactive Packaging

[3] Aloe vera Incorporated bioimetic nanofibrous scaffold: a regenerative approach for skin tissue engineering, Suganya, S.

[4] Tarameshloo, M., Nourouzian, M., Zarein-Dolab, S. Aloe Vera gel and thyroid hormone cream may improve wound healing in Wistar rats

[5] Pellicconi, M., Molinari, G.P. Lucini, L. Stability of the main aloe fractions and aloe-based commercial products under different storage conditions. Agrochimica, Vol. LV-N5, September 2011.

[6] Kuan-Chen Cheng, Ali Demirci, Jeffrey M. Catchmark, "Pullulan: biosynthesis, production, and applications, Applied Microbiology and Biotechnology, July 2011.

[7] Ram S.SinghaGaganpreet K.SainiaJohn F.Kennedy, "Pullulan: Microbial sources, production and applications", Carbohydrate Polymers, Volume 73, Issue 4, 5 September 2008, Pages 515-531.

[8] Shady, S, Gaines, P. Garhwal, R. Leahy, C. Ellis, E. et al. Synthesis and Characterization of Pullulan-Polycaprolactone Core–Shell Nanospheres Encapsulated with Ciprofloxacin. J of Biomed Nanotech Vol. 9, 1–12, 2013.

[9] J.K. Park, T. Khan Other microbial polysaccharides: pullulan, scleroglucan, elsinan, levan, alternant, dextran., in Handbook of Hydrocolloids (Second Edition), 2009.

[10] A Bossart, S. Shady, and D. Kalyon, "Electrospun Pullulan Based Nanofibers for Medical Applications", 42nd Society for 2019 - abstracts.biomaterials.org

[11] Burger,Christian, et al. "Nanofibrous Materials and Their Applications", doi:10.1146/annurev.matsci. 36.011205.

123537.

[12] Dias, J R, et al. "Advances in Electrospun Skin Substitutes." doi:10.1016/j.pmatsci.2016.09.006.

[13] Coats B.C. Method of Processing Stabilized aloe Vera gel Obtained from the Whole Aloe Vera Leaf. 5,356,811. U.S. Patent. 1994 Oct 18.

Improvement of Architecture Based on Hierarchical Message Bus

Zhengwei Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 1328217689@qq.com

Kun Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 644790698@qq.com

Jinhui Li
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 715319152 @qq.com

***Abstract*—The traditional architecture based on the hierarchical message bus has defects such as strict component requirements and lack of support for reuse. The article introduces message-message modem and interface-message modem construction middleware to propose a new hierarchical message bus. The structure new hierarchical message bus solves the problem that the hierarchical message bus architecture requires too strict system components, and the components that fail to provide corresponding interfaces lack support for reuse. Firstly, the paper introduces the traditional hierarchical message bus architecture and its existing defects. Secondly, the article proposes new hierarchical message bus. based on the hierarchical message bus architecture. Finally, the article summarizes the architecture new hierarchical message bus proposed in this paper.**

Keywords-Hierarchical Message Bus; Middleware; Modem; New Hierarchical Message Bus

I. INTRODUCTION

One of the core questions of software architecture design is whether it is possible to use repetitive architecture patterns, that is, whether it can achieve software reuse of the architecture [1]. The structure of the traditional architecture is mostly based on the C/S or hierarchical mode. Although the C/S and hierarchical modes are easy to understand and have high efficiency, they are highly targeted and expand some large systems in specific fields. And the reuse of components is not very friendly. In reality, there is often a need for an architecture with both a stable static structure and sufficient component reuse. Based on the traditional hierarchical message bus structure, this paper further expands and improves the hierarchical message bus for different types of components, and develops intermediate components with wider application scope and better reusability to maximize software reuse. The effect is to achieve greater efficiency of component reuse.

DOI: 10.21307/ijanmc-2020-023

II. KEY TECHNOLOGY

Based on the traditional hierarchical message bus structure (HMB), this paper constructs two convenient middleware message-message modem (MMM) and interface-message modem (IMM) for conversion and adjustment of different standard types of components, so as to achieve greater efficiency of component reuse. This article developed two general middleware:

1) Message-Message Modem (MMM), a non-standard component based on the message interconnect interface, uses MMM middleware to send and receive and parse messages between the component and the message bus.

2) Interface-Message Modem (IMM), calling non-standard components of the interface, using IMM middleware to complete the conversion and transmission between the interface and the message.

III. ARCHITECTURE DESIGN

A. Traditional hierarchical message bus architecture

The Hierarchy Message Bus (HMB) architecture is a software architecture proposed by Yang Fuqing et al.[2]. This architecture is based on a hierarchical message bus and supports the distribution and concurrency of components.

All components pass through the message bus Communicate as shown in Figure 1.

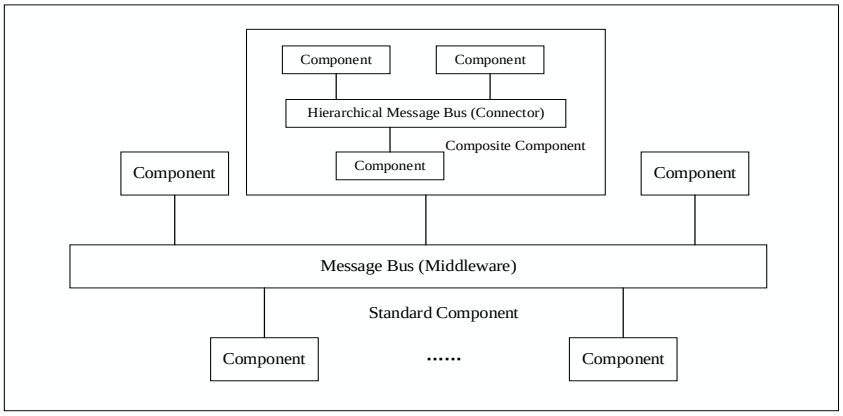


Figure 1. Hierarchical message bus architecture diagram

The connection of the system is the message bus, which is responsible for the dispatch, delivery, filtering and return of the processing results of the message [4]. The components in the system are all connected to the message bus and register the type of messages that interest them to the message bus. The component sends out the corresponding message according to the needs of the system, and the message bus is responsible for dispatching the message to all components interested in the message in the system. In the HMB architecture, the message is the only way for all components in the system to communicate. When the component receives the message from the message bus, it responds to the message according to its own state, and then returns the corresponding processing result to the target component through the message bus [5].

Since all the components in the system are connected through the message bus, it is not required that each component has the same address space or be limited to one machine, so this style can better describe the distributed concurrent system.

The system and the components that make up the system are usually more complicated, and it is difficult to obtain a complete understanding of them from one perspective, so a good software engineering method often models the system from multiple perspectives, generally including the static structure and dynamic behavior of the system And functions [6].

1) Component model:

The component model of HMB includes three parts: component interface, static structure and dynamic behavior, as shown in Figure 2 [3] .

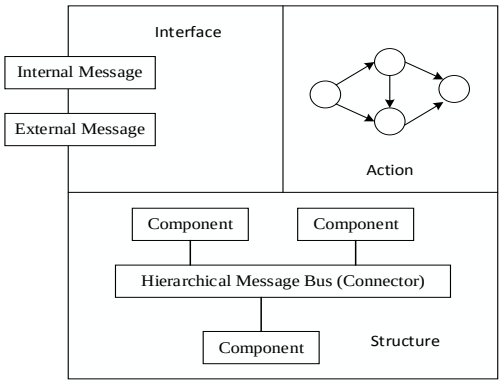


Figure 2. Component model of HMB architecture

The interface part of the component represents the set of interfaces supported by the component. Common ones are internal standard message interface, external message standard interface and external interface, and each interface defines a set of input and output messages. The behavior part of the component is represented by the finite state automaton with output. When the component receives the message, the finite state automaton responds to the received message according to the current state, resulting in a state transition. The structural part of the component represents the internal structure of the composite component. The composite component is composed of an atomic component and a local message bus. Several local components can be connected to the local message bus. The local message bus is a simple representation of the message bus. The role of the message bus it shows that it provides a unified integration mechanism for the entire system and components at all levels.

2) Component interface:

Generally speaking, the component interface represents the interaction path between the component and the external environment. In the HMB architecture, the component interface is a message-based interconnection interface, where the interface represents the set of messages sent and received by the component, and the components communicate with each other through messages, which together constitute the component interface. In the general system defined by the interconnection interface, the connection between components is usually a fixed match between the requester of the service and the service provider, while in the system defined by the HMB architecture, the component has external messages After the response, it will cause a change of state. Therefore, it can be found that, after receiving the same message, a component may have different responses at different times and in different states.

3) Message bus:

The message bus is the connection of the system. In the system, all components form a component-message response registration form by registering the types of messages of interest to the message bus. The message bus locates the corresponding message based on the type of message it receives and the information in the component-message response registration table, passes the message to the corresponding responder, and is responsible for returning the processing result. When necessary, the message bus also needs to filter and block specific messages, as shown in Figure 3.

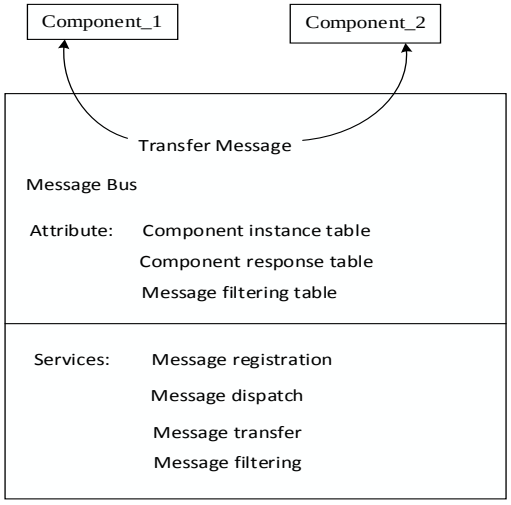


Figure 3. Message bus structure diagram

Message registration: The component registers the response message set with the message bus.

Message dispatching and delivery: Messages are passed between components through the message bus. The message bus dispatches messages to components interested in the message according to the component-message response registration form, and returns the processing results [8].

Message filtering: The message bus implements message filtering between components through message conversion or message blocking, to avoid message conflicts and message mismatch problems during component integration.

Through the above analysis, it can be found that the HMB architecture improves the degree of cohesion between the various components, while reducing the degree of coupling between the components, but there are still certain defects:

First, if there is an existing well-functioning component, but due to the different form of the interface, it will not be used in this system, and it is necessary to re-develop the same functional component that fully complies with the interface standard, which is in line with the architectural style to improve software reuse Development efficiency is contradictory.

Second, although the message filtering mechanism of the message bus has a certain message matching function, it can only process components that interact through messages, but not components that interact through interface calls, and it may be necessary to match different messages. Modifying the design of the message bus has caused some instability to the system.

Third, the architectural style of HMB has too strict requirements on the components that make up the system. Only strictly qualified components can be used, and there is a lack of reuse support for components that fail to provide corresponding interfaces.

Therefore, it is necessary to improve and expand the architecture based on the hierarchical message bus.

B. New hierarchical message bus architecture

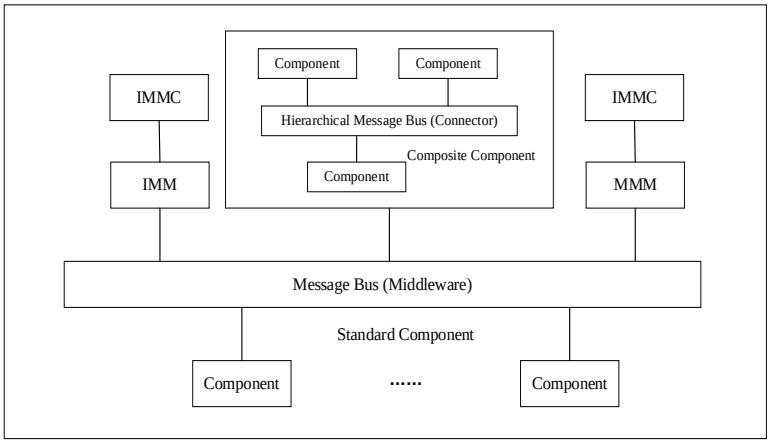


Figure 4. Schematic diagram of the NHMB architecture

This paper improves the traditional HMB architecture and becomes a new HMB architecture style (NHMB, New Hierarchical Message Bus), as shown in Figure 4 below.

Compared with the traditional HMB architecture, the NHMB architecture adds non-standard components of interface-message modem and message-message modem and message interface. Among them, the message-message modem (MMM), a non-standard

component based on the message interconnection interface, uses MMM middleware to send and receive and parse messages between the component and the message bus. The interface-message modem (IMM) calls the non-standard components of the interface and uses IMM middleware to complete the conversion and transmission between the interface and the message.

Complex components in the system can be decomposed into relatively low-level subcomponents, which are connected through a local message bus [7]. Such complex components are called composite components. If the sub-components are still relatively complex, they can be further decomposed, and the decomposition continues in this way, and the entire system forms a tree-like topology. The end nodes of the tree structure are called leaf nodes. They are atomic components in the system and no longer contain subcomponents. The entire system can be used as a component to integrate into a larger system through a higher-level message bus. Therefore, the entire system and the individual components that make up the system can be characterized in a unified manner.

The improved NHMB architecture model in this paper is similar to the HMB structure, which has three parts: interface, static structure and dynamic behavior.

Interface: The interface in NHMB can be either the originally supported interface with message interaction function or an interface that only provides formal calls.

Static structure: The composite components in NHMB are also based on NHMB's architectural style, not the original HMB's architectural style.

Based on the above two changes, the architecture can greatly reduce the requirements for the specificity of components, so that existing components can be reused to a greater extent.

1) Component model:

The improved NHMB architecture in this paper is similar to the HMB architecture, which has three parts: interface, static structure and dynamic behavior.

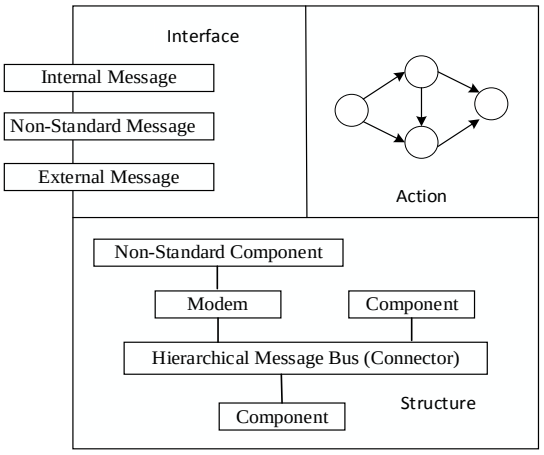


Figure 5. Component model of NHMB architecture

2) Component interface:

The NHMB architecture achieves the indirect use of such components by adding an interface-message modem or a message-message modem between the type of component and the message bus. From the overall system perspective, although a module is added to the system, it has no effect on other modules of the system. In addition, there is no need to modify the message bus. In a sense, the NHMB architecture can achieve the effect of supporting more types of components by adding corresponding middleware modules, thereby greatly improving the system development efficiency and system stability.

3) Modem middleware:

In order to support non-standard components, this article adds interface-message modem or message-message modem middleware to complete the reception and analysis of messages in the system in the NHMB architecture. So as to better adapt to the development of software systems, and greatly reduce the complexity of software system development, maintaining the stability of the system. The working principle of the modem is shown in Figure 6 below.

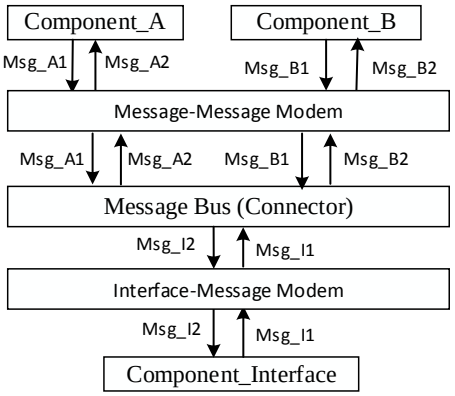


Figure 6. Message modem middleware

Component A, Component B, and Interface component send messages to the message bus together. Component A and Component B are standard components. Component Interface is an interface component. Two different types of components first send their own characteristic messages Msg_A1, Msg_B1, Msg_I1 to MMM and IMMM, then MMM and IMMM parse the message Msg_A1, Msg_B1, Msg_I1 from the component according to the rules prescribed in advance with the standard component and interface component, and then repackage the message into Msg_A2 according to the format that the message bus can handle. Msg_B2, Msg_I2, and forwarded to the corresponding message bus, so as to achieve indirect communication between component A, component B and interface component Interface to the message bus, the process of returning messages is similar to the process of sending messages.

4) Structural characteristics of NHMB system:

High robustness: The message-message modem/interface modem added in this paper has a strong resolution and modulation mechanism, and has strong fault handling and recovery capabilities for non-standard components based on the message interconnect interface.

Higher development efficiency: the system can use mature functional components, even if its interface does not meet the needs of the system, it can be easily integrated into the system by adding a simple modem.

Support the distributed operation of each functional module: since all components in the system interact through the message bus, the system can pass the messages of each module to each other without affecting the distributed concurrent operation.

More secure: The NHMB architecture designed in this paper introduces a modem. In order to increase the security of the system and prevent the communication of the message bus from being maliciously destroyed, authentication and encryption are adopted in the NHMB architecture style to ensure the security of message transmission. The introduction of modem not only reduces the requirements for components, increases the range of component selection, improves the scalability of the system, but also provides a convenient way for the system to expand the security mechanism, that is, the modem developed at this time not only contains message interface matching It also provides special functions for encryption and decryption.

Easy to expand and maintain: Each component in the system is independent of each other, so when expanding or modifying system functions, you can only operate the specific modules in it, without affecting other modules.

IV. CONCLUSION

The NHMB architecture proposed in this paper has the advantages of high robustness, higher system development efficiency, easy expansion and easy maintenance. By adding two intermediate components of message-message modem and interface-message modem on the basis of the traditional HMB architecture The operation greatly reduces the complexity of software system development, maintains the stability of the system, and solves the problem that components with good functions cannot be used because of different interfaces, and the same functional components and HMB system that meet the interface standards need to be redeveloped The structure has too strict requirements on the components constituting the

system, and lacks reuse support for components that fail to provide corresponding interfaces.

REFERENCES

[1] Zhang Yousheng. The third article of software architecture series-style of software architecture [J]. Programmer, 2002(08): 45-47.

[2] Zhang Shikun, Wang Lifu, Chang Xin, Yang Fuqing. Software Architecture Description Language Based on Hierarchical Message Bus [J]. Chinese Journal of Electronics, 2001(05): 581-584.

[3] Zhang Shikun, Wang Lifu, Yang Fuqing. Software architecture style based on hierarchical message bus [J]. Chinese Science Series E: Technology Science, 2002(03):393-400.

[4] Zhang Yousheng, software architecture [M]. Version 2. Beijing: Tsinghua University Press, 2006.

[5] Zhang Shikun, Wang Lifu, Chang Xin, Yang Fuqing. Software Architecture Description Language Based on Hierarchical Message Bus [J]. Acta Electronica Sinica, 2001(05):581-584.

[6] Yang Fuqing, Mei Hong, Li Keqin.Software reuse and software component technology [J]. Journal of Electronics, 1999(02):69-76+52.

[7] Qin Guorong, Zhang Shikun. Research on Dynamic Simulation and Evolution of Software Architecture Based on Hierarchical Message Bus [J]. Computer Science, 2001(03):75-77.

[8] Wang You, Rong Mei, Zhang Guangquan. Application system model based on hierarchical message [J]. Computer Science, 2006(11): 285-288+292.

A Comparative Study of Face Recognition Classification Algorithms

Wang Changyuan
School of computer science and engineering
Xi'an Technological University
Xi'an, China
E-mail: cyw901@163.com

Xue Pengxiang
School of computer science and engineering
Xi'an Technological University
Xi'an, China
E-mail: xuepx@xatu.edu.cn

Li Guang
Northwest Institutes of Advanced Technology
Xi'an Technological University
Xi'an, China
E-mail: 865413666@qq.com

Wu Qiyou
School of computer science and engineering
Xi'an Technological University
Xi'an, China
E-mail: 314650592@qq.com

***Abstract*— Due to the different classification effects and accuracy of different classification algorithms in machine learning, it is inconvenient for scientific researchers to choose which classification algorithm to use. This paper uses the face data published by Cambridge University as an experiment. The experiment first reduces the dimensionality of the data through the principal component analysis (PCA) algorithm, extracts the main features of the data, and then respectively through linear logic classification, linear discrimination LDA, nearest neighbor algorithm KNN, support vector machine SVM and the integrated algorithm Adaboost are used for classification. By comparing the advantages and disadvantages of the classification performance and complexity of different algorithms, the final review reviews accuracy, recall, f1-score, and AUC as evaluation indicators.**

Keywords-Classification Algorithm; Machine Learning; Face Recognition; Model Evaluation

I. INTRODUCTION

With the rise of artificial intelligence and machine learning, face recognition technology is widely used in life, such as station security, time and attendance punching, and secure payment [1-3], but different face recognition devices use different algorithms. Therefore, this paper analyzes and compares the commonly used classification algorithms in face recognition. The data set in this paper uses the ORL face data set published by Cambridge University in the United Kingdom. The methods used involve linear logistic regression, linear discriminant analysis (LDA), K-Nearest Neighbor (KNN), support vector machine (SVM), Naïve Bayes (NB) and other methods. The definition and advantages and disadvantages of the act are briefly explained. Finally, the five methods are compared and analyzed according to the evaluation indicators such as the accuracy rate, recall rate, F1-score, and AUC area commonly used in machine learning.

DOI: 10.21307/ijanmc-2020-024

II. Related Work

A. *Principal component analysis PCA data dimensionality reduction*

The data set contains a total of 400 photos. We use the machine learning library Scikit-learn provided by python to process the data, and display part of the data set pictures as shown in Figure1.

Figure 1. ORL partial face image

The experimental data has 4096 features per picture. Since the number of features is much greater than the number of samples, it is easy to regenerate overfitting during training. Therefore, a principal component analysis algorithm is required to reduce the dimensions of the features and select K main features as the input of the data set. . The main idea of PCA [4] uses the covariance matrix to calculate the degree of dispersion of samples in different directions, and selects the direction with the largest variance as the main direction of the sample set. Processing process:

a) Data preprocessing normalizes and scales the data first. Normalization makes the mean value of data features 0, and scaling is to solve the case where feature values are different by an order of magnitude.

b) Calculate the covariance matrix and eigenvectors of the processed data. The eigenvectors can be obtained by singular value decomposition.

c) Retain the feature vector and feature value corresponding to the largest first K feature values to form an orthogonal basis.

d) Project the sample into a low-dimensional space, and the acquired dimensionality-reduced data can represent the original sample approximately, but with a certain degree of distortion.

This paper use formula (1) to calculate the distortion of PCA.

$$X = \frac{\frac{1}{m}\sum_{i-1}^{m}\left\|X^{(i)} - X_{approx}^{(i)}\right\|^2}{\frac{1}{m}\sum_{1}^{m}\left\|X^{(i)}\right\|^2} \quad (1)$$

Through scikit-learn processing, the reduction rate after dimensionality reduction is obtained from the PCA model diagram is shown in Figure 2. It can be seen from the figure that the larger the value of k, the smaller the distortion rate. As k continues to increase, the data reduction rate will approach 1 Using this rule, choose between 10 and 300, and perform sampling calculation every 15th. Under the k features of all samples, the reduction rate is obtained after processing by the PCA algorithm. We select the reduction ratios at 98%, 90%, 80%, and 70%, and the corresponding k values are 195, 75, 32, and 20. The corresponding pictures after PCA processing are displayed. The first line of the image is the original image, and then each column corresponds to the It is a picture at different reduction rates. The lower the reduction rate, the more blurred the image is shown in Figure 3.

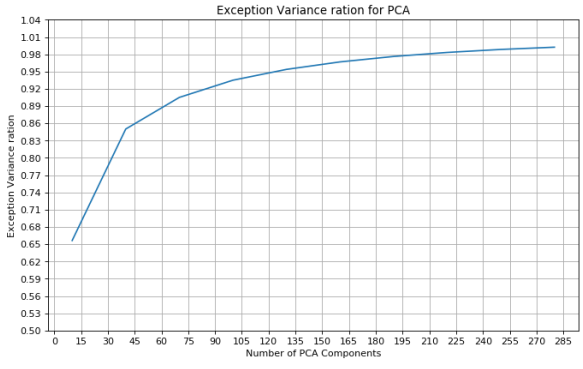


Figure 2. Relationship between reduction rate and K characteristics

Figure 3. Relationship between reduction rate and K characteristics

B. Research on classification method

1) Logistic Regression

Supervised learning is the most widely used branch of machine learning in industry. Classification and regression are the main methods in supervised learning. Linear classifier is the most basic and commonly used machine learning model. This paper uses linear logistic regression to classify and recognize faces.

The prediction function of linear regression is:

$$h_\theta(x) = [\theta_0, \theta_1 ..... \theta_n] \begin{bmatrix} X_0 \\ X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{bmatrix} = \theta^T X \qquad (2)$$

Where is the prediction function, x is the feature vector, To handle the classification problem, this paper hope that the value of the function is [0,1], This paper introduce the Sigmoid function:

$$g(z) = \frac{1}{1+e^{-z}} \qquad (3)$$

Combined with linear regression prediction function:

$$h_\theta(x) = g(z) = g(\theta^T X) = \frac{1}{1+e^{-\theta^T X}} \qquad (4)$$

If there is a simple binary classification of class A or class B, then this paper can use Sigmoid as the probability density function of the sample, and the result of the input classification can be expressed by probability:

$$P(y=1 \mid x, \theta) + P(y=0 \mid x, \theta) = 1 \qquad (5)$$

The cost function is a function that describes the difference between the predicted value and the true value of the model. If there are multiple data samples, the average value of the replacement price function is obtained. It is expressed by J(θ). Close to, based on the maximum likelihood estimate available cost function J (θ).

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} (y^{(i)} \log(h_\theta(x^{(i)}) + (1-y^{(i)}) \log(1-h_\theta(x^{(i)}))\right] \qquad (6)$$

2) Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) [5, 6], also known as Fisher Linear Discriminant (FLD), was introduced into the field of machine learning by Belhumeur. LDA is a dimensionality reduction technique in supervised learning, which can not only reduce dimensionality but also classify, and mainly project data features from high latitude to low latitude space. The core idea is that after projection, the projection points of the same category of data should be as close as possible, and the distance between the category centers of different categories of data should be increased as much as possible [5]. If the data set has two data sets, for the center of the two classes

Then within-class scatter matrix (within-class scatter matrix):

$$S_w = \sum 1 + \sum 2 = \sum_{x \in x_1} (x-u_1)(x-u_1)^T + \sum_{x \in x_2} (x-u_2)(x-u_2)^T \quad (7)$$

Between-class scatter matrix:

$$S_b = (u_1 - u2)(u_1 - u2)^T \quad (8)$$

3) KNN

Among N training samples, find the k nearest neighbors of the test sample x. Suppose there are m training samples in the data set, and there are c categories, namely $\{\varpi_1, .. \varpi_c\}$, and the test sample is x. Then the KNN algorithm can be described as: Find k neighbors of x in m training samples, among which the number of samples belonging to category w_i in k neighbors of x are $k_1, k_2, .. kc$, then the discriminant function is

$$g_i(x) = k_i, i = 1,2,..c \quad (9)$$

The core idea of K-nearest neighbor algorithm [7] is to calculate the distance between unlabeled data samples and each sample in the data set, take the K nearest samples, and then K neighbors vote to decide the type of unlabeled samples.

KNN classification steps:

a) Prepare the training sample set X, which contains n training samples, and select an appropriate distance measurement method according to specific requirements. This paper use dis(xa,xb) to represent the distance between ax and bx in the sample set.

b) For the test sample x, use the distance measurement formula to calculate the distance between the test sample x and n samples to obtain the distance set Dis, where

$$Dis = \{dis(x, x_1), dis(x, x_2), \ldots, dis(x, x_n)\}.$$

c) Sort the distance set, select the smallest k elements from it, and get k samples corresponding to k elements.

d) Count the categories of these k samples, and obtain the final classification results by voting.

Assuming that x_test is an unlabeled sample and x_train is a labeled sample, the algorithm is as follows:

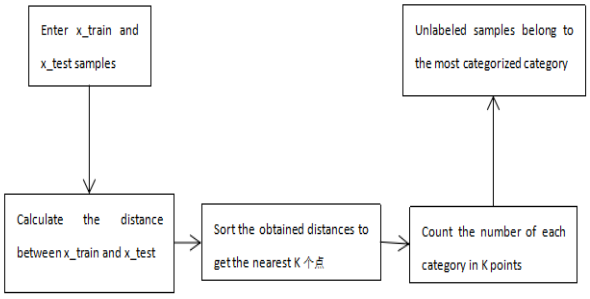


Figure 4. K nearest neighbor algorithm

For distance measurement, Euclidean distance, Manhattan distance, Chebyshev distance, etc. are usually used. Generally, Euclidean distance is mostly used, such as the distance between two points and two points in N-dimensional Euclidean space.

$$d_{ab} = \sqrt{\sum_{i=1}^{N} (x_{1i} - x_{2i})} \quad (10)$$

4) SVM

Support Vector Machine (SVM)[7] for short is a very important and extensive machine learning algorithm. Its starting point is to find the optimal decision boundary as far as possible, which is the farthest from the two types of data points. Furthermore, is the farthest from the boundary of the two types of data points, so the data point closest to the boundary is defined as a support vector. Finally, our goal becomes to find such a straight line (multi-dimensional called hyperplane), which has the largest distance from the support vector. Make the generalization ability of the model as good as possible, so SVM prediction of future data is also more accurate, as shown in Figure 5 below. Find the best Dahua margin.

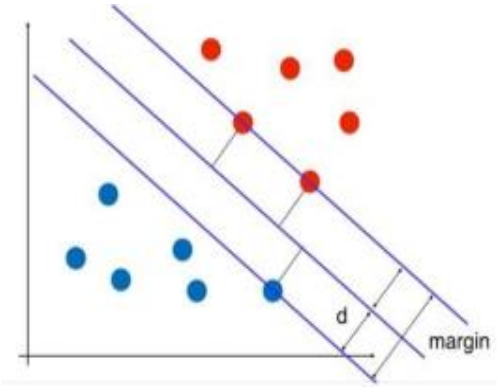


Figure 5. Classification model diagram

Let this plane be represented by g(x)=0, its normal vector is represented by w, the actual distance between a point and the plane is r point, and the distance from the plane can be measured by the absolute value of g(x) (called the function interval) .

$$\begin{aligned} &\min_{w,b} \frac{1}{2} W^T W \\ &y_i (W^T X_i + b) \geq 1, \\ &i = 1 \ldots l \end{aligned} \tag{11}$$

The penalty function is also called the penalty function, which is a kind of restriction function. For constrained nonlinear programming, its constraint function is called a penalty function, and the coefficient is called a penalty factor. The objective function of SVM (soft interval support vector machine) with penalty factor C is:

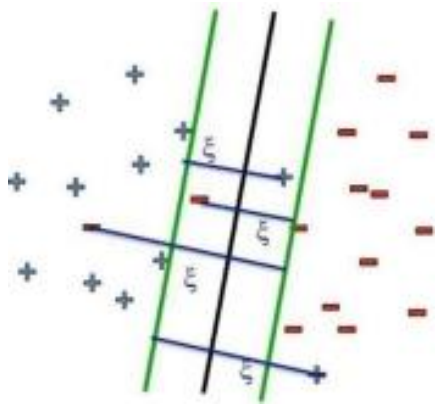


Figure 6. Introduction of C classification model diagram

$$\min_{w,b,\ \xi} \frac{1}{2} W^T W + C \sum_{i=1}^{l} \xi \tag{12}$$

Advantages of SVM:

a) Non-linear mapping is the theoretical basis of SVM method, SVM uses inner product kernel function to replace the nonlinear mapping to high-dimensional space;

b) The optimal hyperplane to divide the feature space is the goal of SVM, and the idea of maximizing the classification margin is the core of the SVM method;

c) Support vector is the training result of SVM. It is the support vector that plays a decisive role in SVM classification decision.

d) SVM is a novel small sample learning method with a solid theoretical foundation. It basically does not involve probability measurement and the law of large numbers, so it is different from the existing statistical methods. In essence, it avoids the traditional process from induction to deduction, realizes efficient "transduction inference" from training samples to forecast samples, and greatly simplifies the usual classification and regression problems.

5) Naive Bayes

Naive Bayes [8, 9] is a conditional independence assumption, and there is no correlation between leave and leave. Suppose there is a labeled data set, where the data set has a total of categories, and for new samples, this paper predict the value. This paper use statistical methods to deal with this problem, so this paper can understand it as the probability of the category to which the sample belongs. The conditional probability formula is:

$$P(C_k \mid X) \tag{13}$$

Therefor $C_k \in [C_1, C_2 \ldots, C_m]$, this paper only require the probabilities of m categories, the category

belongs to the largest value, and use Bayes' theorem to solve:

$$P(C_k \mid X) = \frac{P(C_k)P(X \mid C_k)}{P(X)} \tag{14}$$

And because x has n feature vectors, it is in the determined data set， C_k , P(x) are fixed values, according to the joint probability formula:

$$P(C_k)P(x|C_k) = P(C_k, x) \tag{15}$$

III. EXPERIMENTAL RESULTS

The data set in this paper uses the ORL face data set of Cambridge University, a total of 400 photos. After experimental analysis, this paper chose PCA to reduce the information rate after dimensionality reduction to 98%, and select 195 as the main features of each picture as data input. In order to make the experimental results more generalized, 80% of the data set is used as the training set and 20% is used as the test set. The 10-fold cross-validation method is used during model training. In order to ensure that the experimental results run the same data every time, a fixed random seed is set Is 7. For the test standard, this paper selected the accuracy rate (P), recall rate (R), F value, and AUC area. Compared with our prediction results, the accuracy rate indicates how many true positive samples are in the positive prediction samples. There are two possibilities for the prediction results, that is, the positive prediction is positive (TP), and the negative prediction is positive (FP), the formula is:

$$P = \frac{TP}{TP + FP} \tag{16}$$

The recall rate refers to our original sample, indicating how many positive examples in the sample were predicted correctly. There are also two possibilities, that is, the original positive class is predicted as a positive class (TP), and the other is to predict the original positive class as a negative class (FN).

$$R = \frac{TP}{TP + FN} \tag{17}$$

F1 combines the results of precision rate and recall rate. When F1 is higher, it means that the verification method is more effective.

$$F1 = \frac{2PR}{P + R} \tag{18}$$

TABLE I. THE COMPARISON OF EXPERIMENTAL RESULTS OF FIVE CLASSIFICATION ALGORITHMS

	PCA+LR	PCA+LDA	PCA+SVM	PCA+KNN	PCA+NB
Precision (%)	99	98	94	59	91
Recall (%)	99	96	91	45	85
Fl-score (%)	99	96	91	44	85

IV. CONCLUSION

Face recognition technology [10] has become one of the most popular research directions in computer vision and has made great achievements. With the development of computer technology, more and more classification methods will appear. This thesis is only through several common mainstream classification methods for experimental analysis and comparison. Through experiments, it is found that the PCA + linear logic classification method has obvious advantages in accuracy rate and recall rate. However, specific analysis should be combined with specific issues. Then I am ready to do experiments on different data sets, understand more classification algorithms, and constantly improve my results.

REFERENCES

[1] Wu Xiaotian. Research on Face Recognition Algorithm in Subway Security Inspection [D]. Dalian Jiaotong University, 2017.

[2] Chen Fuqiang. Research on invalid face filtering method in video attendance [D]. Southwest Jiaotong University, 2018.

[3] Ma Yukun. Research on key technologies of face-based secure identity authentication [D]. Beijing University of Technology, 2018.

[4] Pattern Analysis; New Pattern Analysis Findings from King Saud University Discussed (Pcapool: Unsupervised Feature Learning for Face Recognition Using Pca, Lbp, and Pyramid Pooling) [J]. Journal of Robotics & Machine Learning, 2020.

[5] Zhang Yuting, Chen Junhua, Yang Xinkai, Zhang Liyan. An improved PCA+LDA face recognition algorithm [J]. Computer Knowledge and Technology, 2020, 16(03): 221-222.

[6] Guan‐Hua Huang, Chih‐Hsuan Lin, Yu‐Ren Cai, Tai‐Been Chen, Shih‐Yen Hsu, Nan‐Han Lu, Huei‐Yung Chen, Yi‐Chen Wu. Multiclass machine learning classification of functional brain images for Parkinson's disease stage prediction [J]. Statistical Analysis and Data Mining: The ASA Data Science Journal, 2020, 13(5).

[7] Ou Lisong. Design and improvement of face recognition system based on SVM [J]. Network Security Technology and Application, 2019(12): 58-60. [8]Liu Jie, Song Bo. Naive Bayesian Classifier Based on Genetic Simulated Annealing Algorithm [J]. Procedia Engineering, 2011, 23.

[8] Tie Fuzhen. Application of face recognition system in hotel industry [J]. Computer Products and Circulation, 2020(07): 81+97.

[9] Jiang Ajuan, Zhang Wenjuan. A summary of face recognition [J]. Computer Knowledge and Technology, 2019, 15(02): 173-174+190.

[10] Youqiang Zhang, Guo Cao, Bisheng Wang, Xuesong Li. A novel ensemble method for k -nearest neighbor [J]. Pattern Recognition, 2019, 85.

Design Heat Exchanger: Optimization and Efficiency

Haifa El-Sadi*, Joe Aitken, Jason Ganley, David Ruyffelaert, Cam Sweeney
Wentworth Institute of Technology
550 Huntington, Boston, MA, USA
02115
*Corresponding author: elsadih@wit.edu

***Abstract*—Modern commercial and residential buildings procure HVAC systems, to provide heating and cooling for designed open and enclosed spaces to dissipate throughout the accustomed zones. HVAC (heating, ventilating, and air conditioning) systems have become a required industry standard for the construction of new buildings. The objective is the optimization of a heat exchanger model by resolving common system concerns; efficiency, fouling, leakage, dead zones, and vibration. These issues are prevalent in the HVAC industry, which are critical to the under-performing heat exchangers. The heat exchanger was tested at only three different wind speeds (20, 40%, 60%) to take the temperature readings every 5 minutes to allow for maximal heat transfer. The efficiency of heat exchanger at the specified speeds was determined to be .7413 at 20%, .6463 at 40% and .6351 at 60%.**

Keywords-Heat Exchanger; Efficiency; Buildings; Design

I. INTRODUCTION

Mankind is about to install their 700 millionth air conditioner. This baffling number creates great opportunity for engineers because behind each air conditioner there is a heat exchanger doing the heavy lifting to bring comfort to billions of people. Heat exchangers have sought be an engineering challenge and still after 700 million there are still problems that can be resolved. Common problems that heat exchangers endure are fouling, dead zones, and leakage and tube vibrations. Fouling is the reduction of performance due to the build of undesired material. This can be the result of corrosion within the heat exchanger and the lack of filtration prior to fluid entering the heat exchanger. Dead zones can lead to significant fouling and are sections of the heat exchanger that the flow is notably less compared to the rest of the heat exchanger. In most cases dead zones are the result of baffles that most heat exchangers use. But in most cases the baffles are essential for the heat transfer so eliminating them cannot be discussed. Leakage occurs when there is a loss of fluid from the heat exchanger, this results in reduced efficiency. Typically, leakage is the result of faulty connections from poor welding or stressed joints. Tube vibrations can cause the most significant damage to heat exchanger. This can be the result of very high velocity axial and perpendicular flow applications. In addition to resolving common heat exchanger problems the end goal is to improve the efficiency of the heat exchanger. R.L Webb et al. focused on [1] four design cases: (1) reduced heat exchanger material; (2) increased heat duty; (3) reduced log-mean temperature difference; and (4) reduced pumping power. The novel method was presented by B. Linnhoff et al. [2], the method is the first to combine sufficient simplicity to be used by hand with near certainty to identify "best" designs, even for large problems. "Best" designs feature the highest degree of energy recovery possible with a given number of capital items. Previous work [3] showed that three rough tube applications are presented: 1. to obtain increased heat exchange capacity; 2. to reduce the friction power; and 3. to permit a reduction of heat-transfer surface area. Adrian [4] examined the coupling between losses due to heat transfer across the stream-to stream ΔT and losses caused by fluid friction using the concept of heat-exchanger irreversibility.

II. DESIGN AND ANALYSIS

Solid works was used to design different models of heat exchanger, Equation Engineering solver (EES) was used for calculations and analysis, the efficient prototype heat exchanger is shown in Figure 1. Table 1 shows the dimensions of the inner tube and the properties of the fluid (water).

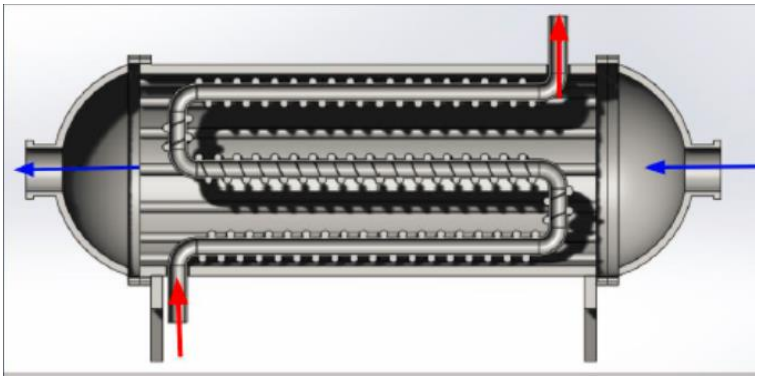


Figure 1. Sectioned SolidWorks model of shell and tube heat exchanger with copper coil and finned inner wall.

TABLE I. INTERNAL FLOW CALCULATIONS

Temperature Inside Pipe	Length	Inner Diameter	Volume Flow Rate	Temperature of Pipe	Average Temperature Between the Surface of the Pipe and Liquid
T_{hi} (C)	L (m)	D_i (m)	$\dot{V}$ $(\frac{m^3}{s})$	T_s (C)	T_{ave} (C)
75	1.397	.013843	.0000335	70	72.5

Properties of Water: Properties were found at the average temperature between the surface of the pipe and liquid and a pressure of 101.3 N/m^2, using Equation Engineering solver (EES) to write the code and solve for the amount of heat transfer.

TABLE II. EXTERNAL FLOW CALCULATIONS

Temp. of Air	Outer Diameter of the Pipe	Temp. of Outer Pipe	Velocity of Air	Vertical Distance Between Pipes	Horizontal Distance Between Pipes	Length the Air Travels	Average Temp. Between the Surface of the Pipe and Liquid
T_{ci} (C)	D_o (m)	T_{so} (C)	V_o $(\frac{m^3}{s})$	S_t (m)	S_l (m)	L_o (m)	T_{aveo} (C)
20	.015875	45	.222	.057785	.0127	1.9304	32.5

Initially the shell of the heat exchanger was going to be formed using sheet metal. With concerns of burn-through while welding the shell together it was determined that buying HVAC duct would reduce manufacturing time and eliminate concerns of sealing the shell together. The second difference between the prototype and the model is the use of fans to send air through the shell. Initially the plan was to incorporate fans within the shell that would send air through and out the other side. The main concern with this idea was if we were going to be able to vary the fan speed for calculation purposes. It was determined that a 4" flexible duct was attached from the heat exchanger to the wind turbine in Wentworth's Fluid Dynamics Lab where we would be able to vary wind speed during testing. Not only did this decision save money it also eliminated doubt from our calculations. The third difference is the overall length of the prototype. This work initially planned to make the shell 18" long with 6" diameter. While we kept the diameter the same, the heat exchanger is now 24" long due to the need for an additional connection piece for the HVAC duct that

was not initially incorporated into the SolidWorks model.

III. MANUFACTURING PROTOTYPE PARTS

The prototype consists of ten major part that will be assembled into a heat exchanger. The parts consist of the controller, pump, piping, valves, stand, reservoir, coil, slides, shell and caps. Out of these nine parts, the stand, reservoir and slides needed to be manufactured in the projects lab. This work manufactured all these parts on the Wentworth Campus in the Manufacturing Lab and Projects Lab.

The stand was manufactured in three major steps. First, we bout ¼ inch plywood and drew the dimensions that would sufficiently hold the tube and reservoir tank. The two vertical pieces were measure to the dimensions of 1' wide and 1.5' tall with a 6" diameter half circle cut at the top to hold the tube in place. The base piece was also dimensioned to be 1' wide and 2' long to hold the tube at the two ends. A small shelf to hold the pump and valves was dimensions to be 1' wide and 6" long. The second steps consisted of cutting all the pieces. This work used two tools to accurately cut the plywood, a circular and jig saw. The circular saw was used to cut the straight angles while the jig saw was used to cut the 6" diameter half circles. After the pieces were cut the stand could be assembled. The third step consisted of using 15 angle brackets, 60 screws and the three-stand pieces to assemble the stand. Three angle brackets were screwed into each inside portion of the base and vertical pieces with two angle brackets on the outside to ensure stability. The shelf was fastened in a similar manner to the outside of the vertical piece.

A. Reservoir

This work manufactured the reservoir using copper plates, angle brackets, screws, JB weld, and flex seal. First, it the precut copper plates and angled them against each other at 90 degrees to construct an open box, and screed the brackets in place at 3 points per corner. Next, with duct taped the outside of the box to ensure a tight seal between the two pieces and applied the JB weld to all the seams. After an hour of curing we applied flex seal to the entire surface of the inside. This sealed between the screws ensuing a waterproof reservoir.

B. Slides

The slides were manufactured using a Solidworks model, which was then converted over to a CAM file. The milling machines would have taken about 3 hours per slide, so we opted to use the LPM. The process with 26 minutes per part, which was a noticeable efficiency difference. The center hole was a quarter of an inch and was pressed out. The plate was then clamped down into the LPM and the larger holes began to mill. Then the circles were cut out from the existing material, with a helix pattern. A hole was drilled at both ends of the shell to allow the ends of the coil to protrude out so that we could connect the fluidics tubing. Once that was all done the end caps were placed on the ends of the shell which shrank the shell diameter from 6” to 4".

Figure 2. The heat exchanger prototype

After that insulation was put on the whole shell assembly to prevent any possible heat from escaping. With the whole shell finished it was time to put the rest of the manufactured parts together. The shell assembly was placed into the 6" diameter half circles in the wooden stand so that it could be supported horizontally. The copper reservoir was then placed underneath the shell assembly. Fluidics tubing was then connected from one of the ends of the copper coil from the shell to one side of the reservoir tank. This where the water will go after going through the heat exchanger. An overflow hole and tube were made in the reservoir to be directed back towards the sink. Electronic valve A was used to pull water from the hot sink water that was stored in a trash car and heated by a Bunsen burner. Piping from the trash can to Electronic Valve A, then the valve to the pump was made. Electronic Valve B was used to pull water from the reservoir tank. Piping from the reservoir tank to Electronic Valve B, then the valve to the pump was made. The final piping from the pump to the inlet part of the coil completed the piping system. Figure 2 shows the heat exchanger prototype.

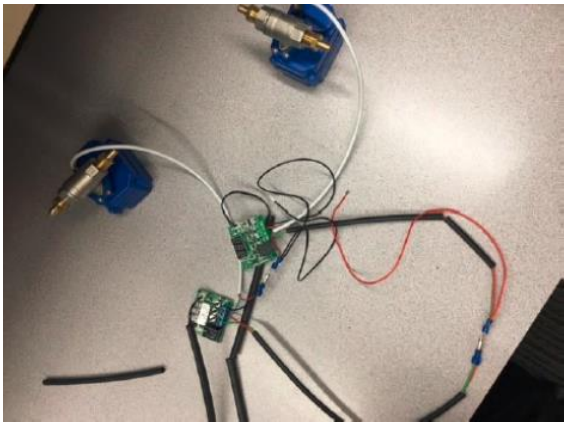


Figure 3. The electronic part of the heat exchanger

Two temperature controllers as shown in figure 3 were used to regulate the opening and closing of the valves. The wires from Electronic Valve A are connected to Controller A. The wires from Electronic Valve B are connected to Controller B. The source connection for the two controllers are to be wired together and connected to a DC source controller supplied by the Electrical Lab. Figure 4 shows testing the heat exchanger. Figures 5, 6 and 7 show the effect of time on the air outlet temperature.

Figure 4. Testing the heat exchanger

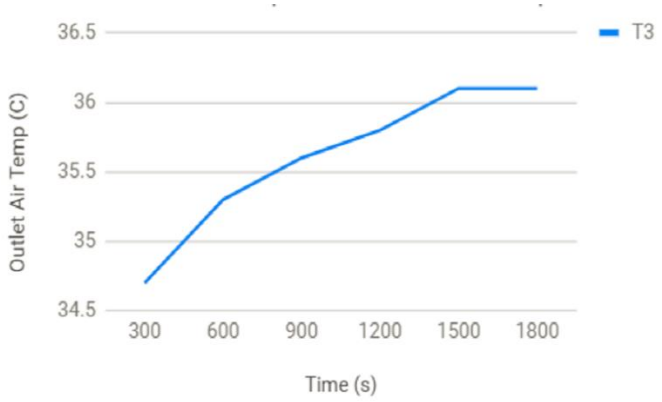


Figure 5. Time vs. outlet air temperature at 20% wind speed

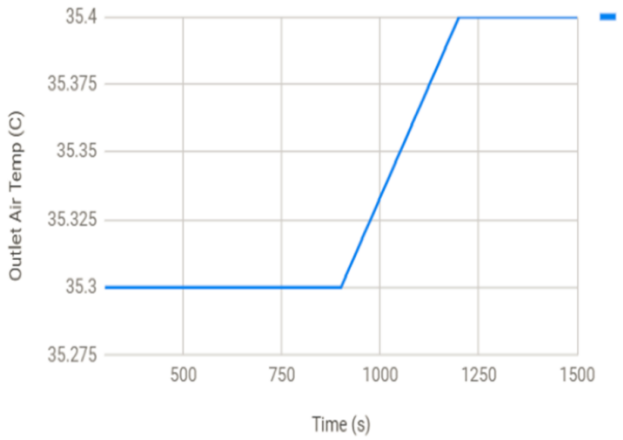


Figure 6. Time vs. outlet air temperature at 40% wind speed

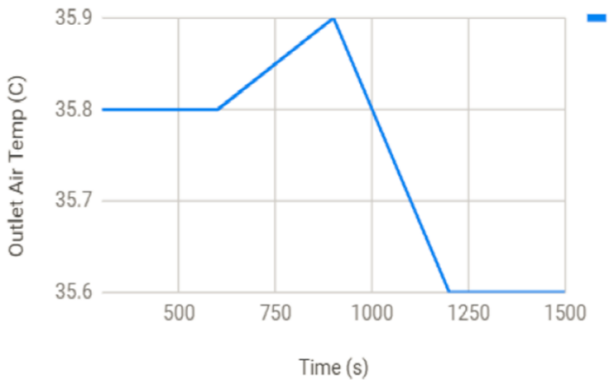


Figure 7. Time vs. outlet air temperature at 60% wind speed

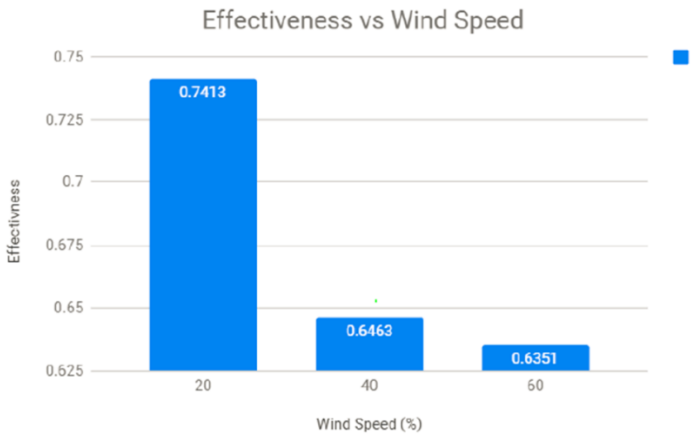


Figure 8. Effectiveness vs wind speed

Initially the heat exchanger was tested with fan speeds starting at 20% increasing up to 100% in increments of 10. However, it was determined that by increasing the fan speed percentage so rapidly there would not be enough time for maximal heat transfer. Knowing this, it was determined that the heat

exchanger would be best to test at only three different wind speeds (20, 40%, 60%) and to take the temperature readings every 5 minutes to allow for maximal heat transfer. Before testing the heat exchanger for the second trial it was hypothesized that since slower fan speeds would allow the air to stay in contact with the inside of the heat exchanger longer the efficiency would be higher. Looking at the graphs above this hypothesis was proven to be correct. Every 5 minutes the inlet and outlet temperatures of both the air and the water were recorded based on the readings given from the thermocouples. While running the heat exchanger the pump sent the water through the copper coils at a constant velocity of .0667 m/s. For the three different fan speed percentages of 20, 40, and 60, the air velocities were measured at 2.2 m/s, 3.87 m/s, and 6.5 m/s. At each fan speed 5 temperature readings were taken, the temperature change between the 1st and 5th reading was then used to determine the efficiency of the heat exchanger at the specified air and water velocities with respect to the initial temperatures of both the water and air. By using the same equations that were used to analyse the initial four designs the effectiveness of the heat exchanger at the specified speeds was determined to be .7413 at 20%, .6463 at 40% and .6351 at 60% as shown in figure 8.

IV. CONCLUSION

After months of research, designing, building and testing we concluded that the heat exchanger provides an effectiveness that is below average compared to common industrial designs. Some variables that may have affected our results are the following; since it is the middle of the summer the outside temperature is quite high making it difficult to determine the true change in temperature since the inlet and outlet temperatures are quite close to one another. If this test was done in the winter are values could potentially be much better. Also, the reservoir tank had a hard time maintaining a warm enough temperature so that could start using the water from it rather than main water source. This may be due to the fact that we used Flex Seal to make sure the tank was water tight. The thermal conductivity of Flex Seal is much lower than copper which is what the tank is made out of. By spraying it everywhere the thermal conductivity of the tank as a whole might not be as high as thought. Overall heat exchanger did what wanted it to and by making these and other modifications the work believe can perform better.

ACKNOWLEDGEMENTS

This project very much to thank Wentworth Project lab and Professor Jackson for Their Support.

REFERENCES

[1] R.L Webb, “Performance evaluation criteria for use of enhanced heat transfer surfaces in heat exchanger design”, international journal of heat and mass transfer, volume 24, pages 715-726.

[2] B. Linnhoff and E. Hindmarsh, “The pinch design method for heat exchanger networks”, chemical engineering science, volume 38, pages 745-763.

[3] R.L Webb and E.R.G Eckert, “Application of rough surfaces to heat exchanger design”, international journal of heat and mass transfer Volume 15, pages 1647-1658.

[4] Adrian Bejan, “General criterion for rating heat exchanger performance”, international journal heat and mass transfer, volume 21, pages 665-658.

A Method to Access a Decimal Network (IPV9) Resource

Guangzhou Liu, Fuya Yu
Xi'an Decimal Network Technology Co. LTD
Xi'an V9 Network Research Institute Co. LTD
Email: 5571200@qq.com

***Abstract*—Network security is highly valued by world leaders. The current Internet technology core is IPv4, IPv6, completely controlled by the United States. On December 14, 2017, the US Federal Communications Commission (FCC) formally abolished the net neutrality law. At that time, the Internet took on an obvious political color and posed a serious threat to Internet applications in various countries. China's economy is already highly dependent on the Internet, and if the network is disrupted, the whole country will suffer heavy losses. The decimal Network Standard working Group of The Ministry of Industry and Information Technology of China and The Decimal Network Information Technology Co., LTD of Shanghai have been researching on the future network for more than 20 years. Developed a complete set of decimal network framework system, completed the future network series research and development with China's independent intellectual property rights, and built the second Internet network system besides the United States. The technology has been fully tested in many places and achieved good results, truly achieving the goal of "autonomy, safety, high speed and compatibility". This paper will introduce the method of accessing decimal network resources in the current network environment.**

Keywords-Decimal Network; CHN; Domain Name; Network Resources

Decimal network is a complete independent intellectual property rights based overall decimal digital code, the establishment of 2^{256} times of cyberspace sovereignty. It includes 13 root domain name servers from the parent root, the primary root, and the zero-trust security mechanism for communication after verification. Compatible with current Internet systems, it has a future Internet architecture that overlaps geographical location and IP address space. Most Internet applications today are based on IPv4 environments. In the context of the existing Internet network, the IPV9 .chn domain name network can be accessed by setting up the existing computer or terminal. Most current computer browsers and mobile browsers support access. For example, Firefox, Google Chrome, Microsoft Edge, 360 speed browser and so on are common on computers. Safari and Baidu browser commonly used on mobile phones need to set the network DNS and point to the IPV9 DNS server before using the browser to open the website. The addresses are: 202.170.218.93 and 61.244.5.162. Once set up, you can access the resources of the decimal network in the current Internet environment.

Before visiting, a few typical IPV9 sites are recommended, as shown in Table 1. Here are the steps to accessing the .C web site on your PC and mobile phone.

DOI: 10.21307/ijanmc-2020-026

TABLE I. TYPICAL CHN DOMAIN NAME WEBSITES

Website domain name	Web resources	Resource management	Resources to address
http://www.v9.chn	.chn portal website	Decimal Network Standard Working Group	Shanghai
http://em777.chn	Decimal technology introduction website	Shanghai Decimal Network Information Technology Co. LTD	Shanghai
http://www.xav9.chn	Xi 'an Decimal System portal	Xi 'an Decimal Network Technology Co. LTD	Xi 'an
http://www.xa.chn	V9 Research Institute portal	Xi 'an Weijiu Research Institute Co. LTD	Xi 'an
http://www.hqq.chn/	The red Flag Canal craftsman	Xi 'an Decimal Network Technology Co. LTD	Xi 'an
http://www.zjsjz.chn	Zhejiang Decimal System portal website	Zhejiang Decimal Network Co. LTD	Hangzhou
http://www.zjbdth.chn	Beidou day draw	Beidou Tianhua Information Technology Co. LTD	Hangzhou

I. COMPUTER ACCESS. CHN WEBSITE SETTINGS

Introduce with Windows10 system settings (PC).

1) First click the "Network" icon on the desktop and select the "Properties" option. The interface appears as shown in Figure 1.

Figure 1. Network and share Center setup interface

2) Click the "Connection: Ethernet" option in the network and Sharing Center setting interface. The interface appears as shown in Figure 2.

Figure 2. Ethernet status interface

3) In the Ethernet status interface, click the "Properties" button. The dialog box appears as shown in Figure 3.

Figure 3. Ethernet property interface

4) In the Ethernet property interface, double-click the option "Internet Protocol Version 4 (TCP/IPv4)". The dialog box appears as shown in Figure 4.

Setting the preferred DNS and alternate DNS and finished setup.

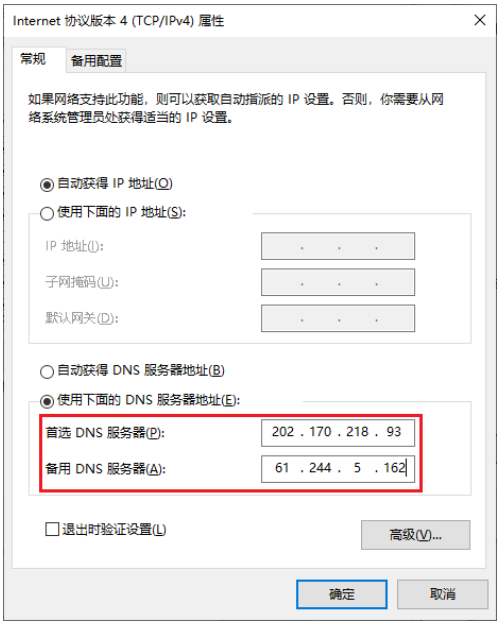


Figure 4. Internet Protocol version 4 (TCP/IPv4) properties

5) Open a browser. Firefox or Google Chrome is recommended. Enter http://www.hqq.chn in the browser address bar to access the IPV9 site, as shown in Figure 5.

II. MOBILE ACCESS .CHN WEBSITE

At present, there are many types of mobile phones, but the setting method is similar. Android mobile phone can download the plug-in (download address: https://www.dtgty.com/HomeSearch) by flow direct access. But in most cases, access to .chn resources will be more convenient over local Wi-Fi. It can also be accessed through mobile hotspots, with the same Settings as Wi-Fi and mobile hotspots. Take Huawei (Android system) mobile phone and iPhone (iOS system) mobile phone as an example to introduce the setting method of mobile DNS.

A. Huawei Mobile Phone setting

The phone type is HUAWEI Mate 20, Android 10 and EMUI 10.1.0.

1) Click "Settings" on the desktop of the mobile phone to display the setting interface, as shown in Figure 6.

Figure 5. Access the IPV9 site

Figure 6. Mobile phone Setting Interface

Figure 7. Wireless connection setting interface

2) Click "Wireless LAN" in the interface, and the interface appears as shown in Figure 7.

3) Press on the connected network name for a while, and additional menu options appear, as shown in Figure 8. Click "Modify Network" menu, the interface of network parameter setting appears, and select "Display Advanced Options", as shown in Figure 9. Select the "Static" option, as shown in Figure 10.

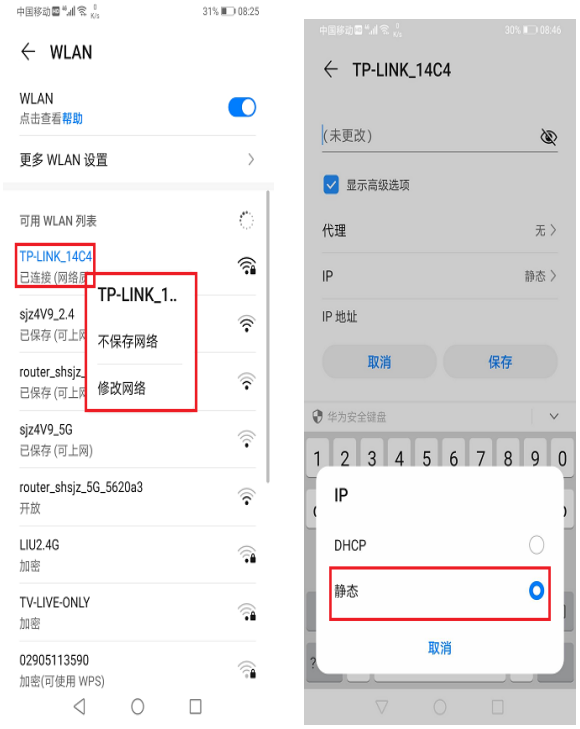


Figure 8. Modification of network Interface

Figure 9. Parameter setting interface

4) Modify DNS according to the parameters in the figure. After modification, click "Save" button to complete the setting.

Figure 10. Modification of network Interface

Figure 11. Parameter setting interface

5) Return to the main interface of the mobile phone and enter http://www.xand.chn in the browser

(Firefox or Google Chrome) to browse the overseas study service website for testing, as shown in Figure 11.

The rest are Xiaomi phones, Vivo phones and so on. You can access IPV9 network resources by simply setting the DNS Settings for the connection network.

B. iPhone parameter setting

Mobile phone model: iPhone XR, system: IOS13.5.

1) Click "Settings" on the desktop of the mobile phone to appear the setting interface. Click "Wireless LAN" in the interface. The interface appears as shown in Figure 12.

2) Click the icon ⓘ on the right of the connected WLAN, and the network setting interface appears, as shown in Figure 13.

Figure 12. Interface of wireless LAN Figure

Figure 13. Interface of wireless connection parameters

Figure 14. DNS Setting Interface

3) In the setting interface, select "Configure DNS" and the DNS setting interface appears, as shown in Figure 14. Select the Add Server option and enter the DNS address shown in the figure. Click the "Save" command in the upper right corner of the interface to complete the setup.

4) Open the browser. Enter http://www.xav9.chn in the address bar to open the main interface of Xi 'an Future Network, as shown in Figure 15.

III. METHOD OF ACCESSING IPV9 WEBSITE WITH CHINESE DOMAIN NAME

In addition to accessing network resources through character domain names, the decimal network system can also use Chinese domain names to access, in the format: http:// Chinese.*****, but before access to the following Settings. Take the Firefox browser, for example.

1) Open the Firefox browser and click the menu button in the upper right corner to open the browser Settings menu, as shown in Figure 16.

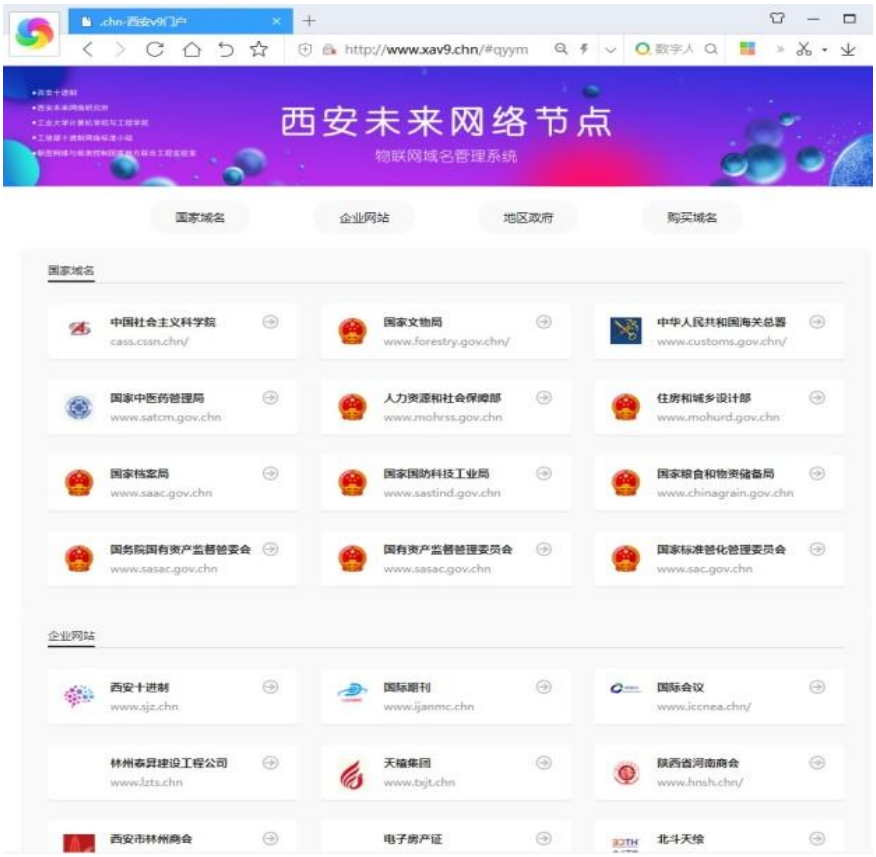


Figure 15. Xi 'an Future Network main interface

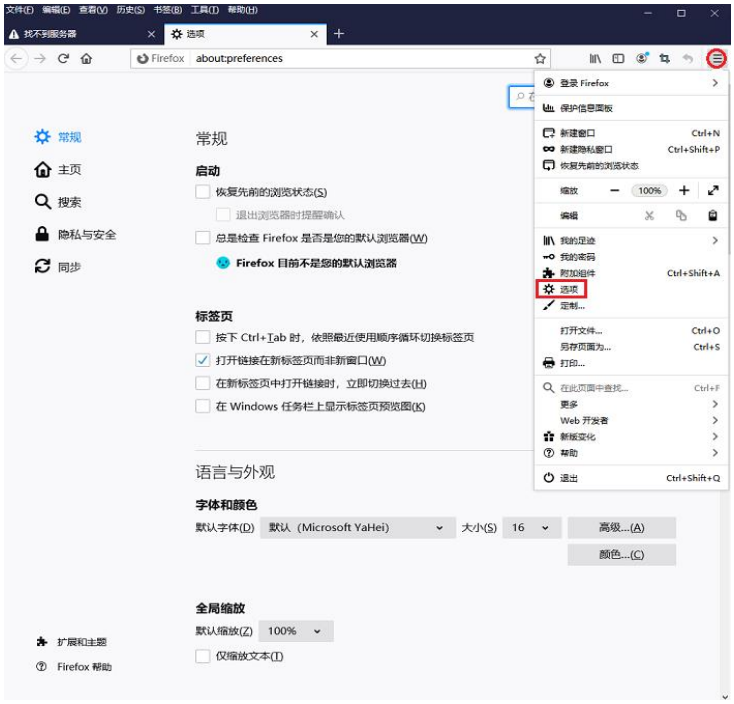


Figure 16. Firefox menu Settings screen

2) Click the "Options" command, drag the right scroll bar to the bottom of the page, and network Settings appear, as shown in Figure 17.

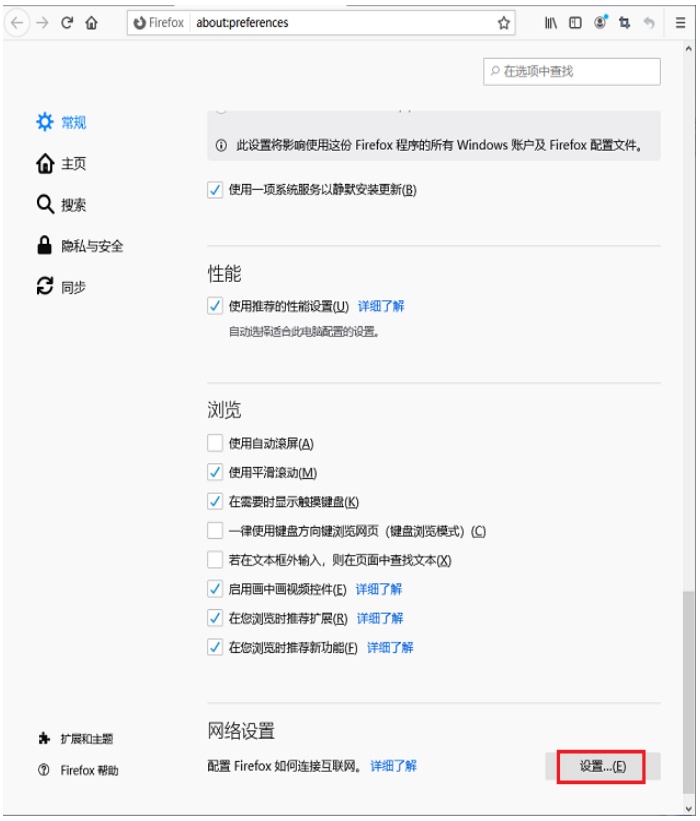


Figure 17. Firefox menu options screen

3) Click the "Settings" button in network Settings, and the "Connection Settings" dialog box appears, as shown in Figure 18. In the Configure Proxy Server to Access the Internet option, select Do not Use proxy Server (Y), and then select Enable HTTPS DNS at the bottom of the screen. Finally enter https://doh.zsw9.cn/dns.query in the "custom" edit box.

4) After setting, click "OK" button to complete setting. Enter the Chinese domain name "China Micro Nine Research Institute" into the Firefox browser to access Chinese website resources. This is shown in Figure 19.

To facilitate test access, several typical IPV9 sites are recommended, as shown in Table 2.

Figure 18. Firefox connection Settings screen

Figure 19. Website of Xi 'an V9 Research Institute

TABLE II. TYPICAL CHINESE DOMAIN NAME WEBSITES

Character of the domain name	Web resources	Chinese domain name	Resource management
http://www.ijanmc.chn	New online international journals	http:// in China. New network and detection control	Xi'an Technological University
http://www.iccnea.chn	ICCNEA International Conference Website	http:// in China. The international conference on	Xi'an Technological University
http://www.xa.chn	.chn portal website	http:// in China. Micro Nine Research Institute	Xi 'an Decimal Network Company
http://www.xav9.chn	Xi 'an Decimal System portal	http:// in China. Xi 'an Future Network Portal	Xi 'an Decimal Network Company
http://www.xand.chn	Xi 'an NORTON Study Abroad website	http:// in China. Xi 'an NORTON Study Abroad	Xi 'an Decimal Network Company
http://www.hqq.chn	The red Flag Canal craftsman	http:// in China. The red Flag Canal craftsman	Xi 'an Decimal Network Company
http://www.xazn.chn	The website of Zhengnuo Conference Company	The website of Zhengnuo Conference Company	Xi 'an Decimal Network Company

In addition to accessing network resources through character domain names and Chinese characters, the decimal address can also be used to access resources. A website corresponds to a decimal address. At the same time, we can also realize a decimal address corresponding to multiple network resources in the way of subdirectory structure. Since decimal address access is bound to the computer in the background, setup is cumbersome, and only a presentation interface is provided here, as shown in Figure 20.

Figure 20. Red Flag Canal Craftsman website

At present the decimal network is in the experimental application stage, although the network resources are less, but the original resources running on the Internet can be completely translated to the decimal network system. With the introduction of national policy, the decimal network of resources will be more and more. The decimal network application of China's independent intellectual property rights is bound to enter thousands of households.

IV. CONCLUSION

This paper introduces the method of using browser to access decimal network resources through personal computer terminal or personal mobile phone under the current Internet environment. A simple DNS setup is required to point to the decimal server to complete resource access. The setup is very simple, which lays the foundation for a wide range of network applications.

REFERENCE

[1] Xie Jianping. A method for assigning addresses to networked computers using full decimal algorithm, Chinese patent No. : ZL00135182.6, 2004.2.6.

[2] Xie Jianping. A method for assigning addresses to networked computers using a full decimal algorithm, U.S. Patent No. :US: 8082365, [4] RFC - Internet Standard. Internet Protocol, DARPA INTERNET PROGRAM PROTOCOL SPECIFICATION, RFC 791, 1981.09.

[3] S. Deering, R. Hinden, Network Working Group. Internet Protocol, Version 6 (IPv6)-Specification, RFC-1883, 1995.12.

[4] M. Crawford. Network Working Group. Transmission of IPv6 Packets over Ethernet Networks. RFC-2464, 1998.12.

[5] J. Onions, Network Working Group. A Historical Perspective on the usage of IP version 9. RFC1606. 1994.04.

[6] V. Cerf, Network Working Group. A VIEW FROM THE 21ST CENTURY, RFC1607. 1994.04.

Demand Forecast of Weapon Equipment Spare Parts Based on Improved Gray-Markov Model

Li Ou
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 18792584953@163.com

Li Chenhao
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 894708601@qq.com

Liu Bailin
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 498194312@qq.com

Gao Dan
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 1366935590@qq.com

***Abstract*—The demand for spare parts of weapons and equipment is time-varying and random. It is difficult to predict the demand for spare parts. Therefore, on the basis of gray GM(1,1), a state transition probability matrix based on improved state division is used to establish a demand forecast model for weapon equipment and spare parts. The model not only considers the characteristics of the GM(1,1) model's strong handling of monotonic sequences, but also extracts the characteristics of random fluctuation response of data through the transformation of the state transition probability matrix, avoiding the phenomenon of the worst prediction results when the maximum probability state is not the actual state. It is proved through experiments that the prediction result based on the improved gray-Markov model is superior to the traditional model and the classic gray-Markov prediction model, and the accuracy of the improved model is about 1.46 times higher than that of the gray model.**

Keywords-Grey Theory; Markov Model; Spare Parts Forecast

I. INTRODUCTION

Weaponry is composed of many parts, and these parts need to be repaired and replaced during use. In order to shorten the short interval between the repairs of weaponry and equipment, increase normal working hours, reserve a reasonable number of spare parts for replacement at any time. Weaponry spare parts are effective measures to improve the availability of weaponry and equipment, reduce life cycle costs, and ensure the effectiveness of combat effectiveness on time [1]. In the issue of spare parts support, the prediction of spare parts demand is an important means to promote "precision support", and is also one of the key and difficult points in the research of equipment comprehensive support.

There are many reasons for the occurrence of spare parts demand. In addition to the reliability rules of weapons and equipment itself, the use of equipment, repair strategies and maintenance methods will affect the number and time of spare parts demand. Spare parts demand shows randomness and fluctuation Therefore, equipment managers urgently need to find a method to predict the random demand for repair spare parts.

At present, there are many technologies for demand planning, forecasting and decision-making, mainly including time series forecasting models, regression analysis methods, support vector machines, neural networks, gray forecasting and decision making, Markov forecasting, Combined optimization decision-making, and their Between each other. Among them, the method based on the time series prediction model requires a large amount of historical data, and the data must not have periodic changes or mutations. Neural network-based spare parts demand prediction method requires a large number of statistical samples, and the

DOI: 10.21307/ijanmc-2020-027

prediction results are highly subjective and random. These factors greatly limit the application of artificial intelligence methods [2]. Therefore, how to improve the prediction accuracy of the demand for spare parts for weapons and equipment is an important link for effectively guaranteeing the supply of spare parts for our military equipment.

The gray prediction model is suitable for solving the problems of small samples, poor information and uncertainty. It has less calculation and is convenient and practical. It has a good prediction effect for short-term prediction, but it has a poor fit for long-term prediction and volatile data series. The Markov theory describes the influence of random factors and the internal laws of transition between states through the state transition probability, which can effectively make up for the deficiencies of the gray model [3]. To this end, this paper proposes to predict the demand for spare parts for weapons and equipment based on an improved gray Markov model, in order to effectively improve the prediction accuracy of random volatility data and broaden the application range of gray theory.

II. RESEARCH STATUS OF SPARE PARTS DEMAND MANAGEMENT

With the development and progress of economic globalization and science and technology, people have higher requirements for product quality and production efficiency, and there are more and more complex equipment in industrial production. In the process of using complex equipment, failures will inevitably occur due to factors such as maintenance damage, wear, corrosion, and expiration of life. In order to restore the normal operation of the equipment in time, minimize the economic loss caused by equipment failure or shutdown. [4] Companies generally purchase and store a certain number of accessories in advance, which are called spare parts. Spare parts, as support materials for the daily maintenance and emergency handling of equipment, are an important factor in ensuring the normal operation of complex equipment. Accurate and timely spare parts supply can ensure the continuity of production operations of the enterprise. Spare parts are an important material basis for equipment support work. Spare parts management Work has become an important part of equipment support work. Spare parts planning is affected by spare parts demand, so an effective spare parts demand forecasting model will provide an important basis for spare parts management decision making, and it is also the basis for quickly responding to changes in customer needs and improving corporate service levels.

The demand for spare parts is very special. In most cases, the demand for spare parts occurs in uncertain and irregular time intervals, and the quantity is also unstable and changeable. Strictly speaking, demand is usually divided into: slow-moving spare parts, intermittent demand spare parts. The consumption of spare parts is extremely special. Some spare parts consume a large amount, while some spare parts consume a small amount, and have not even been consumed in a few years. This greatly increases the difficulty of accurately predicting the consumption of spare parts. In fact, in addition to conventional methods for forecasting spare parts, it is more important to study forecasting methods for uncertain or intermittent demand. Spare parts demand forecasting is a very important part of equipment management, and it is the basis of inventory management. Accurately planning the supply of spare parts can reduce the huge budget spent on spare parts, supply the required spare parts in time, improve the availability of equipment and the completeness of weapons and equipment, ensure that the equipment can complete production tasks and normal operations on time, and guarantee weapons and equipment In military exercises and battles, fighters will not be missed. On the other hand, accurate demand forecasting will have a very important impact on the formulation of spare parts inventory strategies and the construction of inventory models [5].

In many complex equipment companies, spare parts inventory management has not attracted enough attention. In general: On the one hand, there are generally backward spare parts inventory management methods, improper inventory management methods, difficult spare parts search, and excessive backlog of a large number of unimportant spare parts, resulting in high inventory costs of enterprises; on the other hand, some spare parts Inventory will not be able to meet equipment maintenance or customer demand changes. The shortage of key spare parts may cause equipment delays in maintenance or even shutdown accidents occasionally. Inventory issues increasingly become a bottleneck restricting the survival and development of complex equipment companies. For companies, the most important issue is how to use inventory management strategy and inventory management system optimization to greatly reduce the amount of inventory funds occupied by spare parts, and then increase Capital turnover and corporate economic benefits.

III. THEORETICAL OVERVIEW OF GREY SYSTEM AND MARKOV CHAIN

A. Basic concepts of gray system

The naming of the gray system is different from the naming methods of other systems, it is named according to the degree of mastery of the information. The completely unknown information is represented by "black", this system is called the black system, and the information that is clearly grasped is represented by "white", this system is called the white system. According to this law, it is clear that "gray" is in the middle ground, and our grasp of this part of information is in a state of "ambiguity". This system is called the gray system.

The most widely used grey system in the field of forecasting is the GM(1,1) model. Due to its small sample data, simple calculation and other advantages, it has been widely used in various fields such as society, economy, ecology, agriculture, etc., especially in the case of small samples, poor information and uncertain systems and lack of data, it has also been successful. Application, which determines that the GM(1,1) model in the gray system occupies an important position in the fields of prediction and decision-making. In order to expand the scope of application and prediction accuracy of the GM(1,1) model, many scholars have done a lot of theoretical research. These studies mainly focus on: processing the original data, constructing the background value in the GM(1,1) model, discussing the method of determining the initial value in the GM(1,1) model, combining the gray system theory and other theoretical models Combine. Among them, there are two types of concepts in optimized combination forecasting. One is a forecasting method that selects appropriate weights for the weighted average of the forecast results obtained from several forecasting methods. The key is to determine the weighting coefficient of each individual forecasting method. Compare in several prediction methods, choose the prediction model with the best fit or the smallest standard deviation as the best model for prediction. Combination forecasting is to play its role when a single forecasting model cannot completely describe the changing law of the forecast.

B. Basic principles of the gray system

Using the information currently available to explore and predict unknown information is the most important theory of the gray system, which is the process from "grey" to "white" until the research purpose is achieved. The basic principles of the gray system are:

1) Principle of difference information

"Difference" is information, and all information must be different. Saying "thing A is different from thing B" means that A has special information that B does not, which is the so-called difference. The basic way people understand the world is to observe the differences of different things in the objective world.

2) The principle of non-uniqueness of solution

When the information is not fully grasped, the solution obtained for it is uncertain and non-unique. This is caused by the incompleteness of the gray system information and cannot be avoided.

3) Principle of least information

How to use the least information that has been mastered to maximize the effect is an important feature of the gray system to study unknown information.

4) Cognitive basis principle

Information is the basis of cognition. All cognition must be based on the information you have.

5) New information priority principle

The new information has priority in cognition, and its use value is greater than the old information.

6) The principle of immortality

Incomplete information and uncertainty are universal. Information is completely relative and temporary. The completeness of information is specific to a specific environment, and the objective environment is constantly changing, which will also be accompanied by the emergence of new uncertainties.

C. Grey prediction model

The gray prediction model uses GM (1,1) modeling on the known information to find the simulated value, and then predicts the unknown information. This determines the change range of the gray forecast object, and its range must be bounded and related to time. As the core model of grey prediction, GM(1,1) model is widely used in practical research. The establishment of the model is based on the data in the time series. Through this model, the law of data change is analyzed, and the internal relationship is analyzed from the external relationship of various factors to find the hidden law, thereby generating the corresponding data sequence, and performing differential equation modeling. Forecast the development trend of things.

Predicting the gray system is to mine and discover the development laws of the system by processing the

original data and establishing the gray model, so as to predict the future state of the system. Since the gray system contains both known information and unknown information, its prediction is the prediction of the gray process that is related to time and changes in a certain direction. Although the laws shown by some gray processes are random and disorderly, their essence is bounded, orderly, and potentially regular. There are four main types of gray system forecasts:

a) Sequence prediction: refers to the prediction of the behavior of the system variables in the future. After the qualitative analysis of the data series, the appropriate sequence operator is selected, and the gray model is established based on the number sequence after the operation of the operator. The accuracy of the model It can be used to predict after inspection.

b) Catastrophe prediction: refers to the prediction of the abnormal value of the gray system, that is, the prediction of the time when a given gray number occurs, so as to provide guidance for the work of relevant departments.

c) System prediction: refers to the prediction of multiple variables in the system together, which is mainly used to predict the relationship between system variables and reveal the development and changes of the system.

d) Topological prediction: also called waveform prediction. When the original data fluctuates greatly, the gray model is used to predict the waveform of future behavior data.

D. Markov process

Markov process has great significance in stochastic process and is widely used in biology, physics and other sciences. The definition of Markov chain was first proposed in the early nineteenth century, and then people began to study and explore a random process with no aftereffect. Ineffectiveness means that the state of the future is only related to the present and not related to the past. Markov has made great contributions to probability theory, number theory, and differential equations[6]. During 1906-1912, he proposed and studied Markov chains. His research results are of great help to probability theory. The random process he studied is called Markov process.

Markov processes are continuous or discrete according to their state and time parameters, and are divided into three categories

1) Time and state are both discrete Markov processes, called Markov chains.

2) The Markov process with continuous time and discrete state is called continuous-time Markov chain.

3) Time and state are continuous, which is called Markov process.

The research object of Markov model is mainly the state of the system and the transition between states. The main purpose of establishing Markov model for calculation analysis is to predict the possible state of variables in the future by analyzing the current state and change trend of system variables, and then provide corresponding theoretical basis for decision-making.

The steps of using Markov process to predict include:

Step1 Reasonably divide the state. When the sample has less data, the number of states should be as small as possible, so that each state contains more data, which can more objectively reflect the transfer law between each state;

Step2 Calculate the transition probability between each state in the Markov process of the system, and determine the corresponding state transition matrix;

Step3 According to the initial state of the system and the transition probability matrix to predict the state in the future.

IV. IMPROVED GRAY MARKOV MODEL

A. Gray model

Grey system theory was first proposed by our famous scholar Professor Deng Julong in 1982. It is a theory of gray system modeling, prediction, analysis, control and decision-making. Grey prediction method is a new type of nonlinear prediction technology in the late 1990s, which is a method system based on sequence operators and gray sequence production. The gray prediction model uses accumulation, accumulation or step ratio generation techniques, and uses differential fitting to directly convert the time series into a first-order univariate constant coefficient differential equation. According to the principle of continuity of prediction, the gray theory can be used to get more Accurate prediction results [7]. For data with relatively short data series, less information, and less regularity, data that meets the characteristics of poor information systems, the gray model can show superiority and advanced. The gray model is usually divided into a first-order single-variable model and a first-order Combined model. The model is used in this paper.

1) The original data sequence required for the design of spare parts for weapons and equipment is:

$$X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \cdots, X^{(0)}(n)\} \quad (1)$$

among them: $x^{(0)}(k) > 0, k = 1,2,\cdots,n$。

2) Accumulate the original sequence once:

$$X^{(1)} = \left\{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)\right\} \quad (2)$$

among them: $$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)(k = 1,2\cdots,n)$$

3) Generate a sequence of immediate means:

$$Z^{(1)} = \left\{z^{(1)}(2), z^{(1)}(3), \cdots, z^{(1)}(n)\right\} \quad (3)$$

among them: $$z^{(1)}(k) = \frac{1}{2}\left[x^{(1)}(k) - x^{(1)}(k-1)\right]$$

4) Establish model whitening equation:

$$\frac{dz^{(1)}}{dx} + az^{(1)} = b \quad (4)$$

Among them:a and b are undetermined parameters.

5) The gray equation of the model is:

$$x^{(1)}(k) + az^{(1)}(k) = b \quad (5)$$

6) The corresponding time response equation is:

$$x^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right]e^{-at} + \frac{b}{a} \quad (6)$$

7) The least square estimates of parameters a and b are:

$$A = \begin{bmatrix} a \\ b \end{bmatrix} = \left(B^T B\right)^{-1} B^T Y \quad (7)$$

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & 1 \\ -z^{(1)}(n) & 1 \end{bmatrix} \qquad Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$$

8) Reducing the prediction results to the prediction sequence:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \quad (8)$$

among them: $k = 1,2,\cdots,n-1$

It can be seen from the above that the model predicts the original sequence after one accumulation. Because the one-time accumulation sequence is monotonic, the model is suitable for predicting the data of the exponential change law, and the prediction error is large for random volatility data.

The GM(1,1) forecasting model occupies an important position in the grey system theory. The starting point of the model research is to explore valuable information in its own time series and explore the law of the research content without considering the impact of the research content. The gray system prediction model takes a small amount of data information as the research object. Its research characteristics are simple calculation, simple principle and high prediction accuracy. The model has good prediction accuracy for modeling small amounts of information, but the prediction accuracy for irregularly changed and fluctuating data will be greatly reduced, so the gray prediction model does not have high prediction results for any data. The Markov chain prediction model is different from the gray system model. It makes up for the shortcomings of the gray prediction model and can study data with large volatility. Its model requires the prediction object to have Markov properties. In this chapter, the two models are reasonably combined to form a gray Markov prediction model. The GM(1,1) model is used to characterize the trend of the original data sequence, and the Markov prediction model is applied to the simulated data obtained by the GM(1,1) model. The length of the ruler is selected and the shortness of the inch is compensated to improve the accuracy.

B. Grey Markov model

According to the gray system theory, the gray mean model and the Markov model are fused, and the gray Markov model is used to predict the demand for

weapon equipment and spare parts. Specifically, the gray mean model is used to predict the future demand for weapon equipment and spare parts [8]. The state transition matrix determines the possible state of future spare parts, and finally corrects the prediction result based on the ratio between the predicted value and the actual value, so as to realize accurate prediction of weapon equipment spare parts. After determining the state division and state transition matrix, the general gray Markov model uses the maximum value of the current transition probability as the next transition value [9]~[11]. This method ignores the possibility of other transition probabilities and based on the last state. For prediction, it is easy to be affected by randomness, which makes the prediction accuracy low, so the gray Markov model is improved.

The modeling idea of the gray-Markov combination model is to first establish a gray prediction model to obtain a prediction sequence, and then use the relative difference sequence of the prediction sequence and the actual sequence to divide the state space, find the interval of the predicted value, and follow the prediction [12]. The interval modifies the results of the model prediction, increases the credibility of the prediction, calculates the transition probability matrix from the point where the original data sequence falls into each state, and estimates the future change trend based on the state transition probability matrix.

Calculate the state transition probability matrix and obtain the predicted value according to the gray model $\hat{x}^{(0)}(k)(k=1,2,\cdots n)$ .Using the curve as a reference, it is divided into several bar regions parallel to the trend curve, and each region constitutes a state. In this way, a non-stationary random sequence that matches the characteristic points of the Markov chain is divided into n states. Any state is

$$\otimes_i = [\otimes_{1i}, \otimes_{2i}] \tag{9}$$

among them: $\otimes_{1i} = \hat{x}^{(0)}(k) + A_i$, $\otimes_{2i} = \hat{x}^{(0)}(k) + B_i$

The state transition probability is:

$$p_{ij}(k) = \frac{M_{ij(k)}}{M_i}, i = 1,2\cdots, n \tag{10}$$

In the formula: $M_{ij}(k)$ is the number of original data that state $\otimes_i$ transfers to state l after $\otimes_j$ steps; M_i is the number of original data in state $\otimes_i$.

The specific steps based on the improved gray Markov prediction model are:

Step1: According to the original data $x(k)$, find the coefficient a、b in the gray model, and get the fitting value $\hat{x}(k)$ corresponding to the actual data.

Step2: The range of the relative residual value sequence is obtained from the relative residual value formula.

Step3: According to the actual situation, use the residual formula to divide the relative residual value range into states: $\otimes_1 、\otimes_2 \cdots, \otimes_L$

Step4: Based on the improved Markov chain state transition probability determination method, the one-step state transition probability matrix is calculated:

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1L} \\ p_{21} & p_{22} & \cdots & p_{2L} \\ \vdots & \vdots & & \vdots \\ p_{L1} & p_{L2} & \cdots & p_{LL} \end{pmatrix} \tag{11}$$

among them: $p_{ij} \geq 0, \sum_{j=1}^{L} p_{ij} = 1$

Suppose a prediction system has $\otimes_1 、\otimes_2 、\otimes_3 、\otimes_4$ four states, the last data in actual data $x(k)$ is in state $\otimes_4$, then the initial distribution is $I^{(0)} = (0 \quad 0 \quad 0 \quad 1)$, and the predicted value of the next data is $I^{(1)} = I^{(0)} \cdot P$, based on which the state interval in which the predicted data is located can be obtained.

The gray Markov model is a model that combines the gray model and the Markov model. Its prediction process is: firstly generate gray according to the original data sequence, determine the parameters of the gray model according to the generated sequence; then use the established gray model Carry out simulation prediction, test the accuracy of the established model according to the predicted value and actual value; then appropriately divide the gray prediction result into

several states, establish a Markov model, and predict and modify the state of the gray prediction result. According to the combination of gray model and Markov model, traditional gray Markov model prediction methods mainly include the following two:

Use the Markov process to predict the sign of the gray prediction result residual model. After processing the given time series, a gray model is established, and the residual data series can be obtained by comparing the predicted result of the gray model with the actual value. The absolute value of the residual data sequence is used as the original data, and the gray prediction is performed to obtain the gray residual model. In order to improve the prediction accuracy, the Markov process is used to predict the sign of the residual model at the future time.

Use Markov process to predict the relative error of gray prediction results. Establish a gray model based on the original data series, divide the state of the relative error of the gray prediction result, and predict the state of the relative error of the gray prediction result at the future time according to the current state transition, thereby improving the accuracy of the gray prediction result Sex. The essence of the first combination forecasting method is to correct the gray residual model to improve the accuracy of the forecast; the essence of the second combination forecasting method is to directly correct the gray forecast results, so it is better than the first combination forecasting method. The calculation process is simple.

When the Markov process is used to predict the relative error of the gray results, there are two ways: one is to directly predict the state of the system in the future based on the initial state of the system and the n-step state transition matrix; the other is It is based on the initial state of the system and the one-step state transition matrix to predict the state of the next moment, and then based on the state of the next moment and the one-step state transition matrix to predict the state of the system at a later time. The advantage of the first method is that when predicting the state of the system at any time, the initial state used is known and accurate, and the various states of the system at a certain time are considered, and The probability of overestimation and underestimation is small; the disadvantage is: when predicting the state of the system at a distant time, a multi-step state transition matrix is needed. The calculation process of the multi-step state transition matrix is more complicated, and when the number of steps is large The accuracy of the corresponding matrix cannot be guaranteed. The advantage of the second method is that only one step of the state transition matrix is required for Markov prediction, the calculation process is simple, and when the predicted state is accurate, the prediction result is very ideal; the disadvantage is: when the state of the system at a certain moment When the forecast is inaccurate, the forecast result will have a big deviation, and it will affect the forecast accuracy at subsequent moments. In the following articles, the model corresponding to the first calculation method is called the first gray Markov model, and the model corresponding to the second calculation method is called the second gray Markov model.

The improved gray Markov model is a reasonable combination of the first and second gray Markov models, that is, two calculation methods are used to predict the corresponding values of the system parameters at the future time, and then the two results are calculated. Calculate the average value. The improved gray Markov model is optimized for Markov process prediction, so the gray modeling part and The GM(1,1) model is consistent, but the difference is the correction effect of the Markov process on the gray prediction results. The specific steps for the establishment of the improved gray Markov model mainly include: firstly generate gray for the prediction sequence, obtain the parameters of the gray model according to the generated data sequence, determine the gray model and make predictions; then divide the relative error of the gray prediction results State, determine the state transition of the Markov process; then according to the calculation method of the first gray Markov model and the second gray Markov model to obtain the corresponding correction value of the gray prediction result, so as to obtain the improved gray The predicted value of the Markov model.

V. CASE ANALYSIS

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. Experiment Verification

Taking the consumption of a spare part of a certain type of self-propelled artillery of the Military Machinery Maintenance Institute as an example, Table I. is the historical data of the consumption of a certain spare part during the continuous service of a certain type of self-propelled artillery of 2016~2019 for 3 years, with the serial number of the spare part as the horizontal axis, and the demand for spare parts For the vertical axis, the improved gray Markov model is used to predict the future demand for spare parts.

TABLE I. HISTORICAL DATA OF SPARE PARTS CONSUMPTION OF A CERTAIN EQUIPMENT

Serial number	1	2	3	4	5	6	7	8
Number of spare parts	20	14	18	9	10	11	6	5

The following data is used as a sample, and the improved gray-Markov model and other models are used to predict the number of spare parts for the eighth time, and a comparative analysis is made.

Using the improved gray-Markov model to predict the demand for spare parts:

From the above data can be obtained

$$X^{(0)} = \{20,14,18,9,10,11,6,5\}$$

It can be obtained by smoothing the historical data of the demand for spare parts of weapons and equipment

$$\lambda_X = \{1.4268,0.7778,2.0000,0.9000,0.9091,$$
$$1.8333,1.2000\}$$

Therefore, $X^{(0)}$ needs to be exchanged for translation. Take $c = 32.5$, after calculation, we can get

$$Y^{(0)} = \{52.5,46.5,50.5,41.5,42.5,43.5,38.5,37.9\}$$

$$\lambda_Y = \{1.1290,0.9208,1.2169,0.9765,0.9770,$$
$$1.1299,1.0267\}$$

So, the $Y^{(0)}$ sequence can be operated with the gray model, and finally subtract c to obtain $\hat{X}^{(0)}$

Using Equation 2 to do an accumulation generator for we can get

$$Y^{(1)} = \{52.5,99,149.5,191,233.5,277.5,315.5,353\}$$

The sequence is generated by using the immediate mean for ,which can be obtained according to Equation 3.

$$Z^{(1)} = \{75.75,124.25,170.25,212.25,255.25,$$
$$296.25,234.25\}$$

Establish prediction equations according to formulas 4, 5, and 6

$$\hat{y}^{(1)}(k) = -1216.14 \times e^{-0.0405412(k-1)} + 1268.64$$

Find the reduction value according to equations 7 and 8

$$\hat{Y}^{(0)} = \{52.5,48.3176,46.398,44.5545,42.7844,$$
$$41.0845,39.4522,37.8848\}$$

$$X^{(0)} = \{20.0000,15.8176,13.8980,12.0545,$$
$$10.2844,6.9522,5.3848\}$$

Through the prediction of the gray model, you can get the fitted figure of the simulated value and the actual value after data processing as shown in Figure 1.

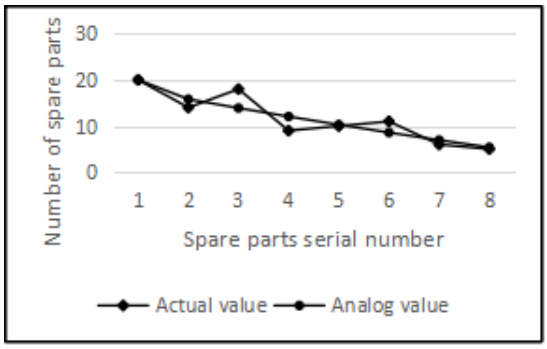


Figure 1. Fitting diagram of simulated value and actual value

Obtain the relative residual sequence according to Equation 9

$$\varepsilon^{(0)} = \{-12.9831,22.7892,-33.9394,-2.8437,$$
$$21.9588,-15.8705,-7.6955\}$$

To divide the state, the upper and lower limits of the state are

$$\begin{cases} \Theta_{\min} = \min(\varepsilon^{(0)}) - 0.05 \times \left|\max(\varepsilon^{(0)}) - \min(\varepsilon^{(0)})\right| \\ \Theta_{\max} = \max(\varepsilon^{(0)}) - 0.05 \times \left|\max(\varepsilon^{(0)}) - \min(\varepsilon^{(0)})\right| \end{cases}$$

The range of the residual error is expanded by 5% above and below, calculated

$$\Theta = (-36.7758,25.6256)$$

The interval between the cells in the state is taken to be equal, and the number of states is sequentially selected from 2 to be calculated. The maximum value is the sequence length. From this, 7 sets of trial calculation results can be obtained.

Calculate the one-step state transition probability matrix according to Equation 12

$$P=\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0.08 & 0.42 & 0 & 0.5 \\ 0 & 0 & 0.3249 & 0.6751 \\ 0.5 & 0.5 & 0 & 0 \end{bmatrix}$$

(1) Use the state transition probability matrix to obtain the prediction results, as shown in Table II

TABLE II. FORECAST RESULTS

Raw data	Model calculation data	Raw data	Model calculation data
20.0000	20.0000	10.0000	10.5184
14.0000	15.8176	11.0000	10.4467
18.0000	16.9127	6.0000	6.1321
9.0000	10.6324	5.0000	6.5528
Predict three numbers	3.4219	2.4896	1.2732

The state transition probability matrix is used to predict the consumption of spare parts for a certain equipment spare part in the future time period, but its accuracy and accuracy need to be calculated using error analysis formulas, and compared with other gray models or classic gray-Markov models.

B. Model comparison

Average error calculation：

$$\bar{E}=\frac{\sum_{k=2}^{n}\left|e(k)\right|}{n-1} \tag{12}$$

First, according to the average error calculation formula, calculate the average error of the model when different states are divided. As shown in Table Ⅲ, it can be seen that when the number of states is 4, the accuracy of the model is the highest.

TABLE III. MODEL AVERAGE ERROR UNDER DIFFERENT STATE DIVISION

Number of states	Model mean error	Number of states	Model mean error
2	15.9536%	6	Undesirable
3	13.7272%	7	Undesirable
4	11.5191%	8	Undesirable
5	Undesirable	—	—

Second, the classic gray-Markov model is used for prediction, and the prediction accuracy is compared, as listed in Table Ⅳ.

TABLE IV. FORECAST ACCURACY OBTAINED BY USING DIFFERENT MODELS

GM(1,1)model	Number of states	Classic model	Improve the model
16.8686%	2	15.9536%	15.9536%
	3	**13.7272%**	13.7272%
	4	Undesirable	11.5191%
	5	Undesirable	Undesirable
	6	Undesirable	Undesirable

When the number of state divisions is 2, 3, the accuracy of the improved model is about 1.06 times and 1.23 times higher than that of the model, which is the same as the accuracy of the classic gray Markov prediction model. When the number of state divisions is 4, the accuracy of the improved model is about 1.46 times higher than that of the model, and the classical Markov prediction model cannot achieve the prediction effect because it cannot build a successful Markov chain state transition probability matrix. As the number of state divisions increases, the average accuracy of the improved gray Markov model is on the rise. The improved model supports more number of state divisions, so it is better than the classic gray Markov prediction model. The application environment of the gray prediction method is relatively loose. It does not need to determine whether the data changes follow the same type of distribution. It does not require large sample statistics to make predictions. It has certain application prospects.

VI. CONCLUSION

In the field of spare parts support for weapons and equipment, forecasting has always been a relatively difficult problem. Due to the time-varying and random characteristics of weapon equipment spare parts demand, this paper proposes to predict the demand of weapon equipment spare parts based on the improved gray-Markov prediction method. By comparing the prediction results, the modified model is superior to the traditional gray model and classic the gray-Markov model has high prediction accuracy and provides a reliable scientific basis for the evaluation of weapon equipment reliability and the maintenance of its equipment.

ACKNOWLEDGMENT

This work is supported by the Shanxi Natural Science Basic Research Project (2019JM-603) for its fund support.

REFERENCES

[1] Shao Yanjun. Research on Preventive Maintenance Strategy of Weapon Equipment Based on Fault Prediction [D]. North University of China, 2013.

[2] Yan Xiaoyao, Yan Li. Application of neural network in the forecast of aviation weapon equipment spare parts demand [J]. Today Keyuan, 2008(20): 56-57.

[3] Zhou Hao, Huang Shanzhong. Equipment consumption prediction based on GM(1,1) and grey Markov model[J]. Journal of Wuhan University of Technology (Transportation Science and Engineering Edition). 2015, 39(06):1166- 1168+1174.

[4] Qiu Chundong, Yang Yuzhi, Wang Chunyun. Prediction of CT tube failure interval based on gray Markov chain model [J]. Modern Instruments and Medical, 2019, 25(06): 86-88+66.

[5] Xu Ning, Dang Yaoguo, Ding Song. GM(1,1) model background value optimization method based on error minimization[J]. Control and Decision. 2015,30(02):283-288.

[6] Lv Z, Hou H, Yi Z, Singh HP, editors. Grey prediction model application in early warning of ice-coating lines[C]. Recent Developments in Control, Automation and Power Engineering (RDCAPE), 2015 International Conference on; 2015 12-13 March 2015.

[7] Ma Liangli, Li Gang, Tao Daoqiang. Fault prediction method based on gray GM(1,1) model[J]. Computer Applications and Software. 2013, 30(04):198-200.

[8] Shi Pei, Gao Shan, Su Yanqin. An improved GM(1,1) model used in equipment fault prediction [J]. Computer Measurement and Control. 2012, (05): 1176-1178.

[9] Zhang Wenjie, Yuan Hongping. Research on the failure prediction of energy-saving equipment based on grey Markov model[J]. System Science and Mathematics, 2019, 39(01): 65-75.

[10] Li Li, Xiong Wei, He Jie, Yuan Xufeng, Zou Xiaosong. Equipment failure rate prediction based on improved grey Markov model [J]. Electric Power Science and Engineering, 2015, 31(08): 20-24.

[11] Hong Zhanpeng, Chen Zhenghua. Gray Markov Model for Metro Car Door Failure Prediction [J]. Equipment Manufacturing Technology, 2017(03):223-226.

[12] Ma Chunmao, Shao Yanjun, Pan Hongxia, Liu Yongjiang. Research on prediction of equipment failure interval based on gray Markov model [J]. Journal of Ordnance Engineering, 2013, 34(09): 1193-1196.

Analysis and Forecast of Urban Air Quality Based on BP Neural Network

Wenjing Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 1908644938@qq.com

Shengquan Yang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: xaitysq@163.com

***Abstract*—The rapid economic development has led to the declining quality of the atmospheric environment. At present, my country is facing a very serious problem of atmospheric environmental pollution. Accurate prediction of air quality plays a vital role in the realization of air pollution control by environmental protection departments. Based on the historical air pollution concentration data, this paper establishes a BP neural network model to learn the statistical law of air pollutant values to realize the prediction of air quality in the future. Through the analysis of the target of air quality prediction, the design of an air quality prediction method based on BP neural network is designed. This method includes four stages: air pollutant concentration data collection, data processing, air quality index calculation, and prediction network construction. The experimental results show that the air quality prediction method based on BP neural network designed and implemented in this paper, combined with the developed air quality prediction system, can effectively predict the recent changes in air quality and various air pollutant concentrations. By collecting the concentration data of air pollutants and learning the changes of air pollutants to achieve air quality prediction, it provides a quantitative reference for government environmental protection departments to achieve air pollution control.**

Keywords-AQI; Air quality Prediction; BP Neural Network

I. INTRODUCTION

Air quality prediction, as the name suggests, is based on the historical emission concentration values of various pollutant items in the air to predict the concentration values of various pollutants in the air pollution in the future and the air environment quality[1]. As China's rapid economic development has led to serious atmospheric environmental pollution problems, the state and the public have paid more and more attention to the treatment and prevention of air pollution. The government environmental protection department hopes to keep abreast of the details of local air pollution and the recent changes in air pollution. The public also hopes to be able to understand the impact of air quality around them on their health in time. In recent years, the state has increased its plans for ecological environmental protection, and the plan clearly clarified that atmospheric pollution control is one of the key contents. The environmental protection departments of local governments strengthen air pollution control work, hoping to understand the changes in air quality in a timely manner by establishing an air quality prediction model.

DOI: 10.21307/ijanmc-2020-028

Xu Dahai proposed an atmospheric advection diffusion box model in 1999, in which the concept of the air pollution potential index was clearly determined, which effectively improved the accuracy of potential prediction on the basis of existing research [2]. In 2002, Liu Shi proposed a statistical model for potential prediction based on the air pollution of Changchun City. The model achieved a certain prediction effect [3]. But generally speaking, the accuracy of the potential prediction is very low, so it needs to be used together with other prediction methods, and cannot be used alone. The chemical model for high resolution of the troposphere in the atmosphere established by Lei Xiaoen is a typical numerical prediction model. Using this numerical prediction model can realize the prediction of the changing process of air pollutants in the atmosphere [4]. Due to its own characteristics, numerical prediction requires detailed geographic, meteorological, and pollution sources to realize the air quality prediction process. Collecting these data in actual situations requires huge costs and is difficult to obtain. In addition, numerical prediction models require high the amount of hardware computing resources is used to calculate the change trend of air pollution at high speed. The calculation complexity is high and it takes a long time, so the current numerical prediction model is not popular in small and medium-sized cities. Taiwan's Pai uses a gray model to achieve air quality prediction. The final actual results show that this method can achieve good results in achieving air quality prediction [5].

The time series analysis method and multiple regression model method in the statistical prediction method simplifies many change factors that affect air quality in the process of achieving air quality prediction, and makes many assumptions in the training process to achieve prediction, and finally achieves air quality the accuracy of the prediction needs to be further improved. The neural network has a good approximation effect in air quality prediction. It can continuously update the newly acquired air pollutant information to the neural network, update the prediction model in time, and improve the prediction accuracy. The neural network has a strong performance in air quality prediction. Dynamic adaptability and fault tolerance. In his research, Wang Jian pointed out that the BP neural network has advantages that other methods do not have in problems such as air quality prediction [6]. This paper uses air quality prediction based on BP neural network, and builds a neural network model to achieve air quality prediction, providing government environmental protection departments with air pollution trends.

II. AIR QUALITY RELATED FACTORS

AQI is the abbreviation of Air Quality Index. AQI does not refer to the value of a specific pollutant project, but reduces the concentration of the six air pollutant projects S02, N02, O3, CO, PM2.5 and PM10 to a single concept. Sex index form, used to represent the overall situation of air quality [7]. According to the size of the AQI value, the air pollution situation can be divided into different levels, and different air quality levels indicate the overall air quality in the local area over a period of time. The goal of this research is to make a short-term prediction of AQI in Xi'an, select the six main pollutant concentrations of AQI as features, build an air quality prediction model, improve the prediction accuracy and efficiency of the air quality prediction model, and provide environmental monitoring and governance Provide accurate air quality information.

In terms of data set acquisition, the air quality pollutant concentration data comes from the weather post website. Using web crawler technology to crawl the data of the website's air quality data module, the data from October 2013 to December 2019 can be obtained. Relevant feature data, after preprocessing the feature data to form an experimental data set. The original data does not necessarily meet the needs of the prediction model. The original data often needs to be processed before the training model is constructed, so

that the collected original data meets the needs of the model. This paper studies the air quality prediction method, and the construction of the air quality prediction model mainly needs to consider the lack of data Processing, data outlier processing, and data normalization processing.

In this paper, the mean value filling method is used to deal with missing values. The mean value filling method is to replace the missing values with the average value of historical data. This method is simple to implement and suitable for models with high accuracy requirements. Data anomaly refers to an unreasonable value in a data set. For example, taking air pollutant concentration data as an example, if the actually collected concentration data value is a negative number, the value is determined to be an abnormal value. In the research method of this paper, the outliers are regarded as missing values, and the outliers are dealt with in the way of missing values. In order to avoid the overflow of the weight of the neural network is too large or too small, to eliminate the possible impact of different variables of the input vector due to different dimensions or too large difference in value, the input vector of the neural network needs to be processed. Normalized data processing is performed on the collected original data set, so that each index of each element data of the vector is at the same order of magnitude, which is suitable for training model for learning. This article uses the Z-score standardized method, the calculation method is:

$$x^* = \frac{x - \mu}{\delta} \tag{1}$$

Among them, μ is the mean value of all sample data, and δ is the standard deviation of all sample data.

III. AIR QUALITY PREDICTION MODEL

A. BP neural network

BP neural network is an error back propagation neural network. Rumerlhar proposed an error back propagation algorithm in the study of forward neural network, referred to as BP neural network algorithm. The network of each layer of the BP neural contains many neuron nodes. There is no connection between the neurons in the layer, and all the neuron nodes between adjacent layers are fully connected [8]. The input layer is used to accept network input information. Each neuron will generate the corresponding link weight according to the input information of the obtained network. The function of the hidden layer in the BP neural network is information detection. According to Tambe's global approximation theory, even if a neural network contains only one hidden layer, as long as there are enough neuron nodes and the appropriate connection function and weight are selected, it can be arbitrary. Approximate the input and output vector of a measurable function [9]. The BP neural network can obtain information and continuously update it to the network, and constantly adjust its structure to meet the characteristics of the model, and has strong self-adaptability and fault tolerance.

The BP neural network learning process is that after receiving the initial input and the given target output, the information forward propagation learning process is performed. This process first calculates and calculates each neural unit of the input layer and each neural unit of the hidden layer. Obtain the output of each neural unit of the hidden layer, and then use the same method to calculate the output of each neural unit of the output layer to determine the error between the actual output of the output layer and the target output. If the error value is within the user's acceptable range, then Fix the weight and threshold, and end the training, otherwise it enters the second stage. The second stage is the error signal back-propagation stage. In this stage, the partial derivative of the error is first calculated

using the output of the output layer, and then the partial derivative obtained by calculation is weighted and summed with the previous hidden layer. Input layer, and finally use the partial derivative calculated by each neural unit to update the weight [10]. Repeat these two stages until the error between the actual output and the target output is reduced to an acceptable range. Figure 1 is the learning flowchart of the BP neural network:

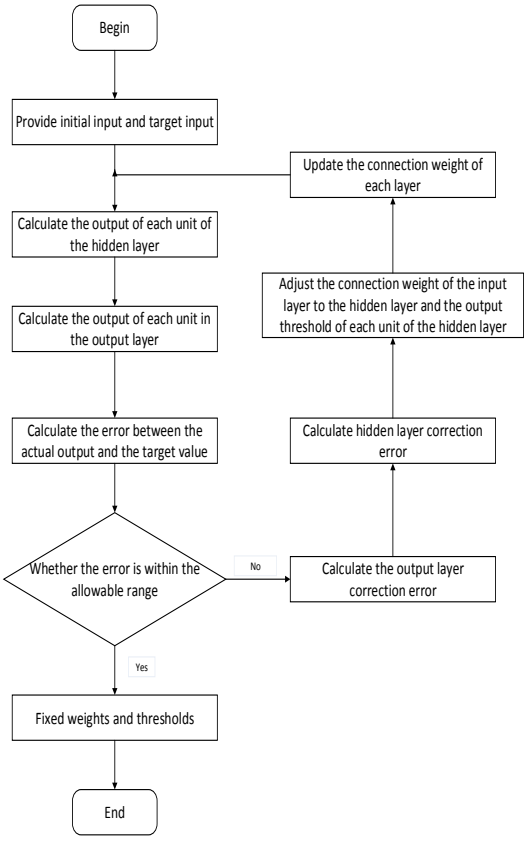


Figure 1. BP neural network learning process

From the above, the algorithm flow of BP neural network can be divided into two processes, as follows:

1) Forward propagation sub-process

It is now defined that the input value of the input layer node is X_{i}, the weight value between the input layer and the hidden layer node is W_{ih}; the threshold value of the hidden layer node is b_h and the value between the hidden layer and output layer node is W_{ho}; output The threshold of the layer node is b_{o}, the network activation function is f, the output value of the output layer node is $\mathrm{y}o_{\mathrm{o}}$, and the expected output value is y_{o}.

The forward propagation process of the BP neural network is to solve the output layer output value X_{i} from the input layer input value $\mathrm{y}o_{\mathrm{o}}$. Specific steps are as follows:

a) Calculate the input and output values of the hidden layer

Hidden layer input value:

$$\mathrm{hi}_h = \sum_{i=1}^{m} W_{ih} X_i + b_h \quad (\mathrm{h}=1,2,...,\mathrm{n}) \qquad (2)$$

Hidden layer output value:

$$\mathrm{ho}_h = f(hi_h) \quad (\mathrm{h}=1,2,...,\mathrm{n}) \qquad (3)$$

b) Calculate the input value and output value of the output layer

Input value of output layer:

$$\mathrm{yi}_o = \sum_{o=1}^{k} W_{ho} ho_h + b_o \quad (\mathrm{o}=1,2,...,\mathrm{k}) \qquad (4)$$

Output value of output layer:

$$yo_o = f(yi_o) \quad (\mathrm{o}=1,2,...,\mathrm{k}) \qquad (5)$$

2) Back propagation sub-process

The back propagation process of BP neural network is based on Widrow-Hoff learning rules. The error function is as follows:

$$E（W，\ \mathrm{b}）=\frac{1}{2}\sum_{o=1}^{k}(y_0-yo_o)^2 \qquad (6)$$

The main goal of the BP neural network algorithm is to iteratively modify the weights and thresholds between layers so as to minimize the value of the error function. According to the Widrow-Hoff learning rule, along the direction of the steepest descent of the sum of squared errors, the weights and thresholds are constantly adjusted. According to the gradient descent method, the amount of weight change is proportional to the gradient of the error function at the current position, as shown in equation (6)：

$$\Delta W=-\eta_1\frac{\partial E(W,\mathrm{b})}{\partial W} \qquad (7)$$

Also for thresholds are:

$$\Delta \mathrm{b}=-\eta_2\frac{\partial E(W,\mathrm{b})}{\partial b} \qquad (8)$$

In the formula: η_1, η_2 is the learning speed, and its value range is (0,1).

The specific steps of the BP neural network back propagation process are as follows:

a) Calculate the weight between the hidden layer and the output layer and adjust the threshold of the output layer

For W_{ho}, according to formula (6), we can get:

$$\Delta W_{\mathrm{ho}}=-\eta_1\frac{\partial E(W,b)}{\partial W_{ho}}=-\eta_1\frac{\partial E}{\partial yi_o}\frac{\partial yi_o}{\partial W_{ho}} \qquad (9)$$

From formulas (3), (4), and (5), we can get:

$$\frac{\partial \mathrm{y}\mathrm{i}_o}{\partial W_{ho}}=ho_h \qquad (10)$$

$$\frac{\partial E}{\partial \mathrm{y}\mathrm{i}_o}(y_0-yo_o)\,f'(yi_o)=\delta_{\mathrm{o}} \qquad (11)$$

From formulas (8), (9), (10), we can get:

$$\Delta W_{\mathrm{ho}}=-\eta_1\delta_{\mathrm{o}}ho_h \qquad (12)$$

Similarly, we can get:

$$\Delta \mathrm{b}_o=-\eta_2\delta_{\mathrm{o}} \qquad (13)$$

b) Calculate the weight between the input layer and the hidden layer and the adjustment amount of the hidden layer threshold

For W_{ih}, according to equation (6):

$$\Delta W_{\mathrm{ih}}=-\eta_1\frac{\partial E(W,b)}{\partial W_{ih}}=-\eta_1\frac{\partial E}{\partial hi_h}\frac{\partial hi_h}{\partial W_{ih}} \qquad (14)$$

Since hi_h affects all output layers, there are:

$$\frac{\partial E}{\partial \mathrm{hi}_h}=\sum_{i=1}^{k}\frac{\partial E}{\partial yi_o}\frac{\partial yi_o}{\partial hi_h} \qquad (15)$$

From formulas (2) and (3), we can get:

$$\frac{\partial \mathrm{yi}_o}{\partial hi_h} = W_{ho} f^{'}(hi_h) \tag{16}$$

From formula (10)、(15)、(16), we can get:

$$\frac{\partial E}{\partial \mathrm{hi}_o} = f^{'}(hi_h)\sum_{o=1}^{k}\delta_o W_{ho} = \delta_h \tag{17}$$

From equations (13), (14) and (17), we can get:

$$\Delta W_{\mathrm{ih}} = -\eta_1 \delta_{\mathrm{h}} \frac{\partial \mathrm{hi}_h}{\partial W_{ih}} \tag{18}$$

Similarly, we can get:

$$\Delta \mathrm{b}_{\mathrm{h}} = -\eta_2 \delta_{\mathrm{h}} \tag{19}$$

c) Update the weights and thresholds of the BP neural network

From (12), (13), the updated weight and output layer threshold between the hidden layer and the output layer are:

$$W_{\mathrm{ho}}^{N+1} = W_{ho}^{N} - \eta_1 \delta_o ho_h \tag{20}$$

$$\mathrm{b}_o^{N+1} = b_o^N - \eta_2 \delta_o \tag{21}$$

From equations (19) and (20), the updated weights and hidden layer thresholds between the input layer and the hidden layer can be obtained:

$$W_{\mathrm{ih}}^{N+1} = W_{ih}^{N} - \eta_1 \delta_h \frac{\partial hi_h}{\partial W_{ih}} \tag{22}$$

$$\mathrm{b}_h^{N+1} = b_h^N - \eta_2 \delta_h \tag{23}$$

B. Design of air quality prediction model

The core algorithm used in this paper is the BP neural network algorithm. According to the characteristics of the BP neural network, this topic needs to determine the number of neuron nodes in each layer of the network, and select the network activation function and initial parameters. The determination of the number of input layer nodes of the BP neural network is very important. Too many or too few selections will affect the prediction accuracy of the model. Therefore, the number of input layer nodes should be determined according to the actual application needs. This subject designs the input layer and output layer of the network based on the collected data. The number of input layer nodes is 6, which are the data of the concentration values of six pollutants such as PM2.5, PM10, SO2, NO2, C0, and O3 in a day. The number of nodes in the output layer is one, that is, the AQI value of the next day. The structure of the BP neural network in this subject is divided into three layers, with only one hidden layer. There is no theoretical guidance for determining the number of hidden layer nodes, and it is usually based on specific practical experience. The empirical formula for selecting the number of hidden layers is:

$$\mathrm{p} = \sqrt{n+q} + \alpha \tag{24}$$

In the formula, n and q represent the number of neurons in the input layer and output layer, respectively, generally take an integer between 1-10. The number of hidden layer nodes in this subject is first determined as 7.

The network activation function is an important factor that affects the performance of the BP neural network algorithm, which makes the network have nonlinear processing capabilities. There are three activation functions of BP neural network: log-sigmiod function, tanh function and ReLU function. According to the characteristics of the research data and the

characteristics of the three activation functions, in this paper, the hidden layer of the BP network selects the log-sigmiod function as the activation function, and the output layer selects the ReLU function as the activation function. Since the sample data is normalized, the value interval between the initial weight and the threshold is between [-1, 1], and they should be a set of random numbers that are not exactly equal.

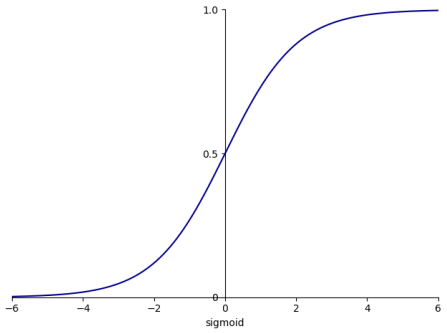


Figure 2. Log-sigmiod function

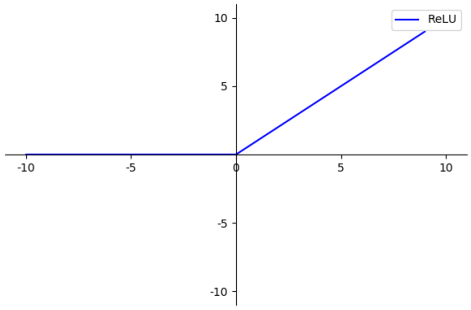


Figure 3. ReLU function

IV. EXPERIMENT

The experimental simulation platform used in this article is the Python programming language. The air data object used in the experiment is the air quality data of Xi'an from October 2013 to December 2019. All the experimental data are sorted in a continuous time series. Take the data for 30 consecutive days as the test data set, and the other as the training data set. For the evaluation of the advantages and disadvantages of the model, this paper uses the average error and the root mean square error to evaluate. The calculation formulas are shown in equations (25) and (26):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y^*_i}{y_i}\right| \quad (25)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y^*_i)^2} \quad (26)$$

Experimental results show that the prediction model established in this paper has high accuracy and high efficiency for PM2.5 concentration prediction. The simulation prediction results are shown in Figure 4 the measured values and predicted values of the first 6 groups are compared to obtain Table 1.

TABLE I. COMPARISON BETWEEN MEASURED AND PREDICTED

AQI	AQI prediction	PM2.5	PM 10	So2	No2	Co	O3
45	56.21	27	40	4	24	0.71	75
49	60.51	23	47	6	36	0.63	84
55	64.30	33	57	6	35	0.62	84
57	68.54	39	60	5	40	0.71	56
68	80.22	40	67	5	39	0.77	85
61	74.35	33	69	6	47	0.70	71

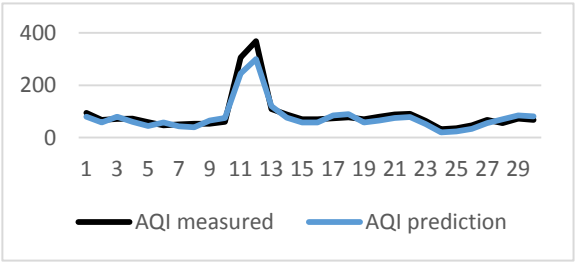


Figure 4. Comparison of sample prediction results with real values

The experimental results show that, after analyzing the prediction results, it is concluded that under the experimental conditions given in this paper, the average error of the experimental results is 0.074 and the root mean square error is 13.41. As can be seen from Figure 4-1 and Table 4-1, the BP neural network established in this paper has lower prediction error when the air quality fluctuates greatly.

This research also has some shortcomings at present. It only considers the relevant factors that can be quantitatively analyzed, and does not take into account some unexpected emergencies. For example, natural disasters, human factors, etc. Due to the unpredictable and unquantifiable characteristics of these factors, they have not been considered in the article. In the future research work, we hope to analyze these factors.

V. CONCLUSION

This article aims at the current situation of severe air pollution problems facing China. The traditional air pollutant online monitoring system cannot effectively use historical air pollutant data to provide quantitative reference for air pollution control and various control measures. The environmental protection department urgently needs to establish an air quality prediction system to realize the supervision and control of local air pollution. This paper studies a method to achieve air quality prediction based on BP neural network. By studying the change law of historical air pollutant project concentration data, it predicts the future air quality change trend for a period of time, and helps government environmental protection departments formulate air pollution control policies to provide quantification Indicators and references.

This article first explains the research background and significance of this topic, and analyzes the necessity of establishing an air quality prediction system for air pollution control. Based on the analysis of the domestic research results of air quality prediction, combined with the regional characteristics and actual conditions of prefectural and municipal government departments, a framework model for air quality prediction based on statistical prediction is proposed. Then, an air quality prediction method model based on BP neural network is established, and the realization of the method includes three stages of air pollutant project concentration data collection, data processing, and prediction algorithm network model construction. This paper uses BP neural network to predict the air quality in Xi'an. Through the analysis of experimental results, BP neural network has a significant effect in dealing with such nonlinear problems, especially in the place where the AQI fluctuation is relatively large. The research is conducive to the prediction and prevention of air pollution problems. The government can also make appropriate measures and decisions based on the prediction results, such as closing schools or reducing outdoor sports, thereby reducing the damage caused by pollution. It can also provide new ways and methods for forecasting research in other fields.

ACKNOWLEDGMENT

The Research is supported by the new network and detection control national and local joint engineering laboratory. (Financing projects No. GSYSJ2016014).

REFERENCES

[1] Ren Wanhui, Su Zongzong, Zhao Hongde. Advances in the study of urban environmental air pollution forecasting [J]. Environmental Protection Science, 2010, 36(03):9-11.

[2] Xu Dahai, Zhu Rong. Popularization and application of urban air pollution forecasting model [J]. Annual Report of CAMS, 1999(00):33.

[3] Liu Shi, Wang Ning, Zhu Qiwen, Wang Xinguo, Hu Zhongming, Chen Changsheng. Research on the Statistical Model of Air Pollution Potential Forecast in Changchun City [J]. Meteorology, 2002(01):8-12.

[4] Han Zhiwei, Du Shiyong, Lei Xiaoen, Ju Lixia, Wang Qingeng. Urban air pollution numerical prediction model system and its application [J]. Chinese Environmental Science, 2002(03): 11-15.

[5] Tzu‐Yi Pai,Keisuke Hanaki,Ren‐Jie Chiou. Forecasting Hourly Roadside Particulate Matter in Taipei County of Taiwan Based on First‐Order and One‐Variable Grey Model [J]. John Wiley & Sons, Ltd, 2013, 41(8).

[6] Wang Jian, Hu Xiaomin, Zheng Longxi, Liu Zhenshan. Research on air pollution forecasting method based on BP model [J]. Environmental Science Research, 2002(05):62-64.

[7] Wang Qingeng, Xia Sijia, Wan Yixue, Jin Longshan. Problems and new ideas in current urban air pollution forecasting methods [J]. Environmental Science and Technology, 2009, 32(03):189-192.

[8] A. Elkamel, S. Abdul-Wahab, W. Bouhamra, E. Alper. Measurement and prediction of ozone levels around a heavily industrialized area: a neural network approach [J]. Advances in Environmental Research, 2001, 5(1).

[9] Jaakko Kukkonen, Leena Partanen, Ari Karppinen, Juhani Ruuskanen, Heikki Junninen, Mikko Kolehmainen, Harri Niska, Stephen Dorling, Tim Chatterton, Rob Foxall, Gavin Cawley. Extensive evaluation of neural network models for the prediction of NO 2 and PM 10 concentrations, compared with a deterministic modelling system and measurements in central Helsinki [J]. Atmospheric Environment, 2003, 37(32).

[10] Hunt K.J., Sbarbaro D., Żbikowski R., Gawthrop P. J. Neural networks for control systems â€ ”A survey [J]. Pergamon, 1992, 28(6).

Hierarchical Image Object Search Based on Deep Reinforcement Learning

Wei Zhang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: weivanity@ gmail.com

Yuxing Tan
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 842061340@qq.com

Hongge Yao
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 835092445@qq.com

***Abstract*—Object detection technology occupies a pivotal position in the field of modern computer vision research，its purpose is to accurately locate the object human beings are looking for in the image and classify the object. With the development of deep learning technology, convolutional neural networks are widely used because of their outstanding performance in feature extraction, which greatly improves the speed and accuracy of object detection. In recent years, reinforcement learning technology has emerged in the field of artificial intelligence, showing excellent decision-making ability to deal with problems. In order to combine the perception ability of deep learning technology with the decision-making ability of reinforcement learning technology, this paper incorporate reinforcement learning into the convolutional neural network, and propose a hierarchical deep reinforcement learning object detection model.**

Keywords-Object Detection; Deep Learning; Reinforcement Learning

I. INTRODUCTION

When observing a picture, humans can immediately know the location and category of the object in the image, and can get the information without even thinking too much. This is a breeze for us, but the computer cannot have all kinds of complicated ideas of our human brain, and it is not easy to realize it.

In computer vision, the positioning and retrieval of images will be affected by two aspects, one is the content of the image, and the other is the pros and cons of the algorithm. There are two main factors influencing the image. The first is that the background and light when taking pictures will affect the quality of the image, resulting in a decrease in the accuracy of object detection. The second is the content of the image. If there are several similar objects, or some are

DOI: 10.21307/ijanmc-2020-029

blocked by other objects, and the different angles of the object will affect the accuracy of detection. The algorithm mainly focuses on how to make the features have higher quality. Therefore, how to design an algorithm that can satisfy accurate positioning and continuously improve the object positioning speed is the key to research.

For computers, these pictures are data collections which are composed of binary digits, and the things behind the data cannot be imagined by computers. Our purpose is to let the computer simulate our human vision and simply have the ability to process the image. Human beings get a lot of information in real life every day, and most of them belongs to the information transmitted to us by vision, and only part of the information in these visual images is what human need. Therefore, by extracting the important information, positioning and identifying them accurately, human can greatly reduce the amount of data that the computer needs to process and improve the efficiency of data processing.

Reinforcement learning is an important field in machine learning. It constructs a Markov Decision Process and simulates human thinking to teach agents how to make actions that can obtain high reward values in the environment, and find the best strategy to solve the problem in such constant interaction. Based on this idea, this paper use reinforcement learning technology to simulate the human visual attention mechanism. The agent is taught to change the shape of the bounding box and focus only on a significant part of the image at a time, and then extract its features through the convolutional neural network. Finally, the object of image positioning and classification can be achieved.

II. RELATED WORK

A. Traditional object detection algorithm

Traditional object detection algorithms include primary feature extraction methods such as HOG feature extraction of objects and training SVM classifiers for recognition. Their algorithms are generally divided into three stages (see Figure 1.):

1) Select different sliding window frames according to the size of the object, and use the sliding window to select a part of the content in the figure as a candidate area.

2) Extract visual features from candidate regions.

3) Use SVM classifier for identification.

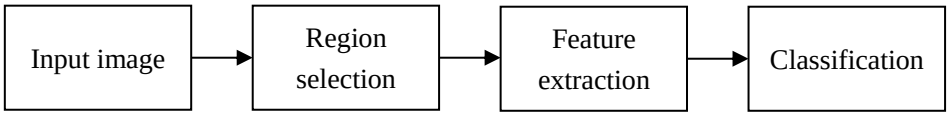


Figure 1. Traditional object detection algorithm

The traditional object search algorithm has the following disadvantages:

1) The selection strategy based on sliding windows is to slide across the entire image from beginning to end. For different object sizes, the program need windows with different size ratios to traverse. Although it can mark all the positions of the object, its brute-force enumeration search results in extremely high time complexity and a large number of windows that are not related to the object, so the speed and performance of feature extraction and classification have fallen into a bottleneck.

2) The characteristics of each object are different, which leads to the diversity of forms, and the background factors of each object will also affect the accuracy of recognition. Therefore, the features of manual design are not very robust.

B. Object detection algorithm based on deep learning

After the appearance of CNN, it has been widely used in the field of computer vision. With the continuous development of science and technology, the difficulty of obtaining a large amount of sample data has been significantly reduced, and the continuous improvement of computing capabilities has enabled CNN to have the ability to extract features from a large amount of data, which has made huge gains in computer vision.

Aiming at the shortcomings of traditional method for object detection, the object detection algorithm based on deep learning uses CPMC, Selective Search, MCG, RPN and other methods to generate candidate regions instead of window sliding strategy. These methods usually use various details of the image, such as image contrast, edge parts and color to extract higher-quality candidate regions, while reducing the number of candidate regions and time complexity.

This type of object detection method is generally divided into two types: one-stage detection algorithm and two-stage detection algorithm. The one-stage detection algorithm regards the object detection problem as a regression problem and directly obtains the category and position information of the object. The detection speed of the algorithm is fast, but the accuracy is low. The two-stage detection algorithm first generates a large number of region proposals, and then classifies these region proposals through the convolutional neural network, so the accuracy is higher, but the detection speed is slower.

C. Object detection algorithm based on deep reinforcement learning

In recent years, research on deep reinforcement learning has emerged endlessly. It has achieved excellent performance in many games than human master players, especially the success of the DeepMind team on the AlphaGO project, pushing deep reinforcement learning to a new height. In this context, many researchers try to apply deep reinforcement learning technology in the field of object detection.

In 2015, Caicedo et al. adopted a top-down search strategy, analyzed the entire scene at the beginning, and then continued to move toward the object location. That is, use a larger bounding box to frame the object, and then shrink it step by step, eventually making the object surrounded by a compact bounding box. In 2016, Mathe et al. proposed an image-based sequence search model to extract image features from a small number of pre-selected image positions in order to efficiently search for visual objects. By formulating sequential search as reinforcement learning of the search policy, their fully trainable model can explicitly balance for each class, specifically, the conflicting goals of exploration - sampling more image regions for better accuracy -, and exploitation - stopping the search efficiently when sufficiently confident about the object's location.

The above algorithm models all use reinforcement learning techniques to improve deep learning algorithms, and all have achieved good results. However, if the visual object algorithm is required to have a relatively high accuracy, it still needs to rely on a large number of candidate regions, so our research direction is to reduce the number of candidate regions while maintaining the quality of the candidate regions at a high level.

III. HIERARCHICAL OBJECT SEARCH MODEL BASED ON DRL

A. MDP formulation

This paper regard object detection as a Markov Decision Process, and find an effective object detection strategy by solving decision problems. In each process, the agent interacts with the current environment based on the current state, and decides the next search action, and gets an instant reward value. The agent continuously improves the efficiency of search in the

process of learning to obtain high cumulative reward value.

There are 6 different actions in action space A, which are composed of two different types: select action and stop action. The selection action is to frame a part of the current area as the next observation area. It consists of four borders and a center frame, which respectively reduce the current search area to different sub-regions (see Figure 2.); the stop action indicates that the object has been found, so the bounding box is no longer changed, and the search process stops.

In reinforcement learning, the state is the premise and basis for the agent to make actions. In this model, the state is composed of two aspects. One is the feature vector extracted by the convolutional neural network in the current state. The other is the historical action information performed in the process of searching for the object. This information helps to stabilize the search trajectory, so that the search process will not fall into the loop search, thereby improving the accuracy of the search.

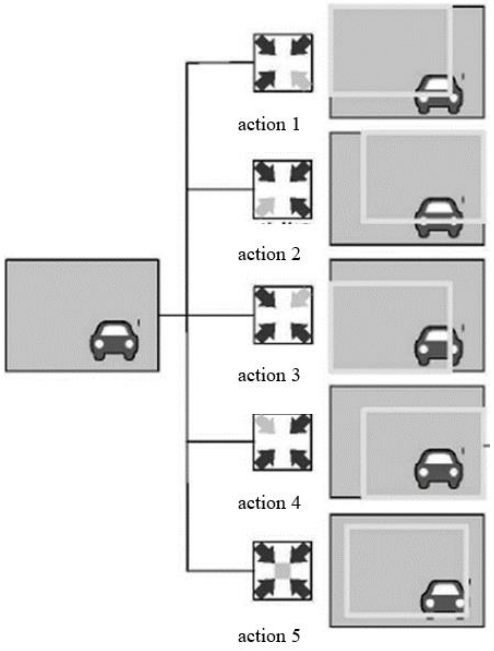


Figure 2. The selection action diagram

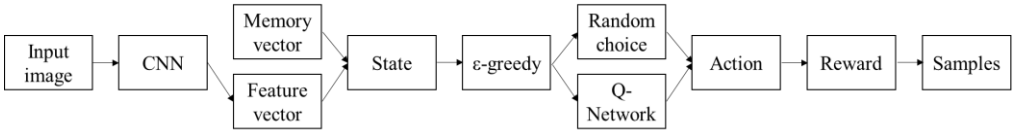


Figure 3. Sample generation process diagram based on Markov decision process

The function of the reward is to reflect the reaction obtained by the agent during the interaction with the environment. The agent judges the merits of the action according to the different rewards received, and finally learns the strategy by maximizing the cumulative reward. Since the agent have two types of actions, our calculation methods are different depending on the type of action.

The reward obtained by the agent depends on the action it takes in the current state. The model use IoU to evaluate the effect of the action, so that the accuracy of detection can be obtained. This paper take e' as the observation area after the movement, e as the observation area before the movement, and g as the area of the object area, R_m is the reward obtained after making the selection action, then our reward function can be expressed for:

$$R_m = sign(IoU(e',g) - IoU(e,g)) \quad (1)$$

If the difference of the overlap rate is positive, it means that our prediction range is closer to the object

area, if it is negative, it means that the prediction range is farther from the object area. If the decision improves the detection accuracy, the reward is positive, otherwise the reward is negative.

The model use R_n as the reward function for the stop action, and set the reward value for the stop action as δ. At the same time, the program need to add a threshold η to determine when it will end the action. When the value of IoU is greater than the threshold, indicating that object has been found, then the program can end the search and perform the stop action. At the same time as there is a reward process, there is also a punishment when IoU continues to fall below the threshold and reaches the maximum number of searches. So that the agent knows the wrong process and corrects it. The reward function for the stop action is as follows:

$$R_n = \begin{cases} +\delta, if\ IoU(e,g) \geq \eta \\ -\delta, if\ IoU(e,g) < \eta \end{cases} \quad (2)$$

B. DQN algorithm

The model use three fully connected layers to form the Q-network, its input is the information content of the image, and the activation values of the 6 neurons in the output layer represent the confidence of 6 kinds of actions, among which the highest confidence corresponding action is selected.

According to the state, action and reward function, the agent apply the Q-learning algorithm to learn the optimal strategy. Because the input image is a high-dimensional data, this paper use DQN to approximate the Q function $Q(s,a)$ in high dimensions. The Q function based on strategy π is expressed as follows:

$$Q_\pi(s,a) = E[R|s_i = s, a_i = a, \pi] \quad (3)$$

The agent selects the action with the highest Q value from the Q function, and uses the Bellman equation to continuously update the Q function:

$$Q(s,a) = r + \gamma \max Q(s',a') \quad (4)$$

Among them, s is the current state, a is the selected action in the current state, r is the immediate reward, γ is the discount factor, s' indicates the next state, and a' indicates the next action to be taken.

In order to train DQN, the program need a large number of training samples which are usually continuously sampled (see Figure 3.), but the continuity between adjacent samples will cause inefficiency and instability of Q-network learning. This paper use the experience-replay mechanism to solve this problem. When the capacity of the experience pool tends to be saturated, the program constantly replace the old samples with new samples. At the same time, in order to make most of the samples selected with nearly the same probability, the program randomly extract samples in the experience pool.

The loss function of the training process is set as follows:

$$L(w) = (r + \gamma \max Q(s',a',w) - Q(s,a,w))^2 \quad (5)$$

Among them, $Q(s,a,w)$ is the actual output of the network, $r + \gamma \max Q(s',a',w)$ is the expected output of the network, r is the current reward value, $\max Q(s',a',w)$ is the maximum expected reward value for the next decision, γ is the discount factor.

C. Hierarchical object search process

The initial candidate region of the model is the entire image. The size of the candidate region is normalized to a fixed size, and then put into a trained

CNN neural network model to extract feature values, and then rely on the greedy algorithm to use the probability ε randomly select one of all actions to search, or use the learned strategy to make action decisions with a probability of $1-\varepsilon$.

After the model made action a, it switched to a new candidate area e' which is a sub-region of the previous region, according to the reward function to give our agent the corresponding reward R_m, and at the same time normalize the new candidate area and put it into the neural network model for features extraction, combine with previous actions to get a new state s'. Repeat the above hierarchical process continuously until our action becomes a stop action, or the number of search steps reaches the upper limit. If a stop action occurs, the final reward R_n is given according to its corresponding termination reward function.

IV. EXPERIMENT

A. *Data sets and parameter settings*

Use the Pascal VOC data set to train the model, which is the most used data set for object detection. The training set uses the combination of Pascal VOC 2007 and Pascal VOC 2012, and the test set uses the Pascal VOC 2007 Test Set.

The model use three fully connected layers to form a Q-network. Its input is the information content of the image. The activation values of 6 neurons in the output layer represent the confidence of 6 actions. The parameters of the network are initialized by a standard normal distribution function. The initial value of the greedy factor ε is 1, every iteration, the ε decreases by 0.1, and stops when it decreases to 0.1. Set the size of the experience pool to 1000, the reward discount coefficient γ to 0.9, and the threshold to make the stop action is 0.5.

B. *Experimental results and analysis*

- Model training

In the process of training the model, the value of the loss function is continually declining along with the continuous iteration of the neural network, making the neural network tend to converge (see Figure 4.). When the number of training times reaches a certain level, the loss value tends to be stable, and various parameters in the network are also updated, forming a neural network model with recognition capabilities.

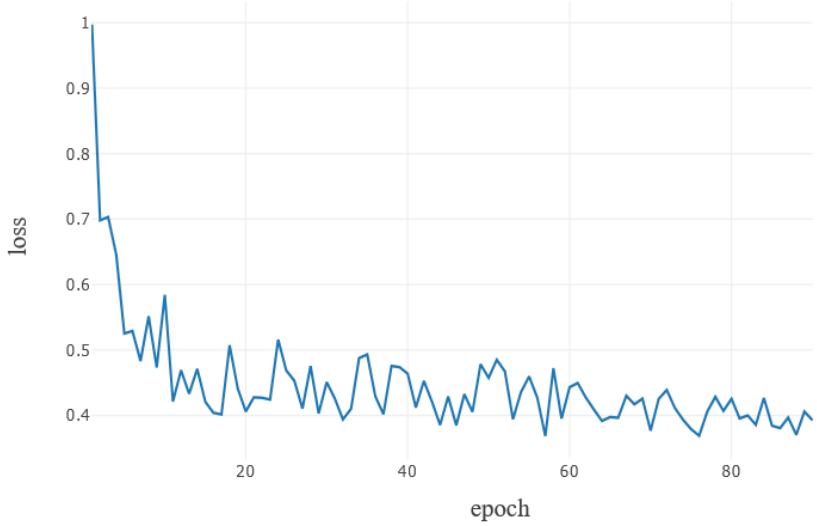


Figure 4. Schematic diagram of loss function

- Results and analysis

The model first analyzes the entire picture and finds the object through a series of frame transformation actions. Finally, the agent make the stop action indicate the end of the search. The following figure shows this hierarchical dynamic selection process in detail.

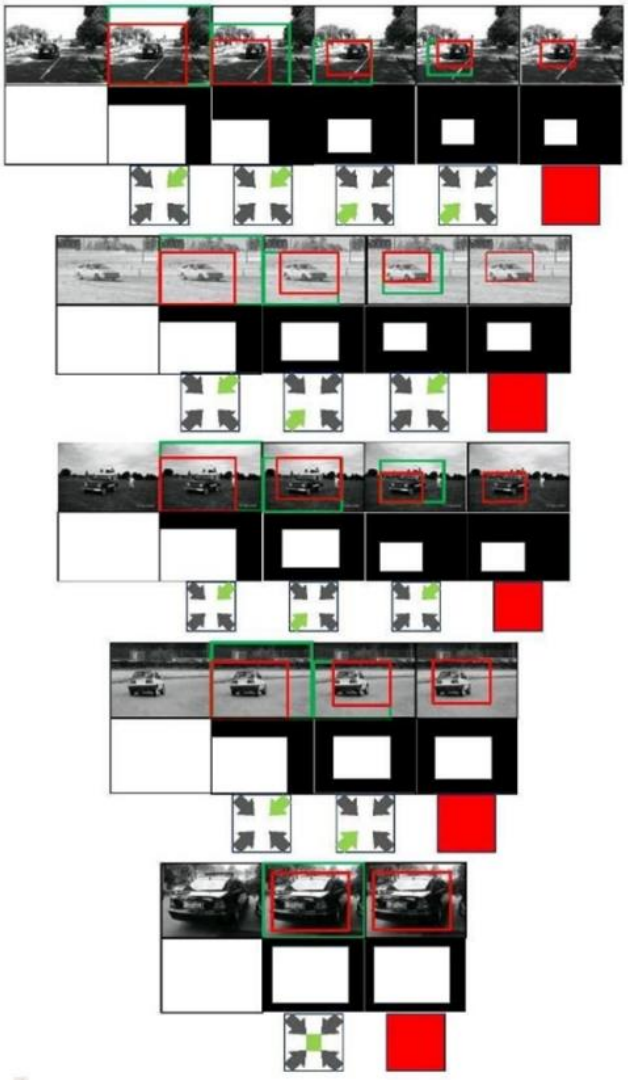


Figure 5. Hierarchical dynamic selection process

Experimental results show that the algorithm model proposed in this paper can improve the search speed and accuracy in object search. However, it can also be seen from the experiment that there may still be errors in the match between the object prediction frame and the actual bounding box of the object, because the model can only continue to select from the area selected by the previous bounding box. As a result, the predicted bounding box cannot reach other areas of the image. The model can improve the detection result by changing the appropriate proportion of the framed area.

V. CONCLUSION

This paper propose an object detection model based on deep reinforcement learning, which focuses on

different areas of the picture by performing a predefined area selection action, and iterates the process to make the bounding box tightly surround the object, Finally achieved the positioning and classification of object. Experiments show that the model can effectively detect the object in the image.

REFERENCES

[1] Sutton R S, Barto A G. Reinforcement learning: An introduction[M]. MIT press, 2018.

[2] Gagniuc P A. Markov chains: from theory to implementation and experimentation[M]. John Wiley & Sons, 2017.

[3] Hu Y, Xie X, Ma W Y, et al. Salient region detection using weighted feature maps based on the human visual attention model[C]//Pacific-Rim Conference on Multimedia. Springer, Berlin, Heidelberg, 2004: 993-1000.

[4] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.

[5] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[6] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). IEEE, 2005, 1: 886-893.

[7] Boser B E, Guyon I M, Vapnik V N. A training algorithm for optimal margin classifiers[C]//Proceedings of the fifth annual workshop on Computational learning theory. 1992: 144-152.

[8] Papandreou G, Kokkinos I, Savalle P A. Modeling local and global deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 390-399.

[9] Carreira J, Sminchisescu C. CPMC: Automatic object segmentation using constrained parametric min-cuts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 34(7): 1312-1328.

[10] Uijlings J R R, Van De Sande K E A, Gevers T, et al. Selective search for object recognition[J]. International journal of computer vision, 2013, 104(2): 154-171.

[11] Pont-Tuset J, Arbelaez P, Barron J T, et al. Multiscale combinatorial grouping for image segmentation and object proposal generation[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(1): 128-140.

[12] Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. nature, 2016, 529(7587): 484-489.

[13] Caicedo J C, Lazebnik S. Active object localization with deep reinforcement learning[C]//Proceedings of the IEEE international conference on computer vision. 2015: 2488-2496.

[14] Mathe S, Pirinen A, Sminchisescu C. Reinforcement learning for visual object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2894-2902.

[15] Watkins C J C H, Dayan P. Q-learning[J]. Machine learning, 1992, 8(3-4): 279-292.

[16] Mnih V, Kavukcuoglu K, Silver D, et al. Playing atari with deep reinforcement learning[J]. arXiv preprint arXiv:1312.5602, 2013.

Interview with the Inventor of the Future Network IPV9

Guoshao Chen
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 1825247141@qq.com

Wang Yubian
Department of Railway Transportation Control
Belarusian State University of Transport
34, Kirova street, Gomel, 246653, Republic of Belarus
E-mail: alika_wang@mail.ru

***Abstract*—With the rapid development of the Internet, from the PC terminal to the mobile terminal, from big data systems to intelligent hardware, technology has shown the unique charm of the Internet. With the development of the Internet, the network information security and network sovereignty issues involved have become increasingly prominent. Therefore, only in an environment of security, equality, and mutual assistance can the Internet play its due economic and social value. The emergence of a new generation of Internet IPV9 marks a key step for China to move towards an autonomous and controllable future network. IPV9 is to further safeguard national network sovereignty on the basis of fully guaranteeing network information security. But the defamation of IPV9 still exists. Recently, in order to further clarify the facts, we hereby interviewed Xie Jianping, the inventor of the Decimal Network, to conduct an in-depth discussion on the new generation Internet IPV9.**

Keywords-IPV9; Future network; Decimal Network

I. INTRODUCTION

The core of the current Internet (also known as the Internet) technology is IPv4 and IPv6, and its technical core is completely controlled by the United States. On December 14, 2017, the US Federal Communications Commission (FCC) officially abolished the net neutrality rule, making the Internet with obvious political color and posing a serious threat to Internet applications in various countries. The address space of the IPv4 protocol is 2 to the power of 32. Due to the insufficient estimation of the development trend of the Internet in the early stage of Internet, the insufficient setting of the address space length caused the unreasonable IP allocation. By 2010, there were no addresses to allocate. In theory, IPv6 has 2^{128} addresses, but only one eighth of the addresses can be assigned to end users, so there are only 2^{125} addresses, which is equivalent to 10^{37}. The 128 bar code in the Internet of Things is already 10^{128}, which cannot be covered, so IPv6 also has certain limitations.

Since the establishment of the Decimal Network Standards Working Group of the Ministry of Industry and Information Technology in August 2001, Shanghai Decimal Network Information Technology Co., Ltd. has conducted more than 20 years of research in the future network field, developed a complete network framework system, and completed IPV9 with independent intellectual property rights. The patent obtained by IPV9 (2001, patent number CN98122785) has been recognized by many countries including China, the United States, the United Kingdom, Russia and other countries. This

DOI: 10.21307/ijanmc-2020-030

innovative and internationally strategic new achievement has been vigorously endorsed by the Ministry of Industry and Information Technology, the National Standards Committee and other ministries Supported the establishment of a second network system other than the United States.

II. FUTURE NETWORK IPV9

In Dr Zhang's article, IPV9 is the legal version of the American IETF. The IPV9 is a version that the United States has publicly declared unsuccessful, but has never abandoned.

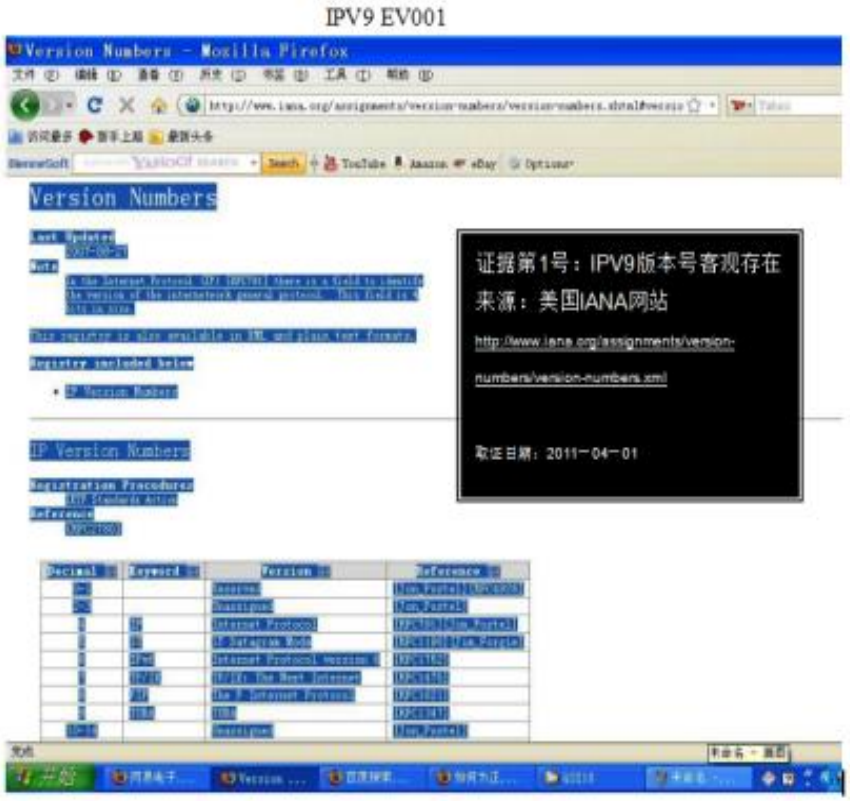


Figure 1. IPV9 EV001

As early as 1994, the United States admitted that IPV6 address length was inadequate and announced that the old system was behind The Times, and began working on IPV9, but it was never successful. At present, IPV10 is being developed in the United States. What is the concept of IPV10? Because IPV4 and IPV6 are not compatible, IPV10 solves the problem of communication between IPV4 and IPV6 mutual group machine, and the address length should reach 256 bits. The main reasons for the insufficient length of the IPV6 address are as follows: First, the length of the current bar code EAN • UCC on the Internet of Things is 10 to the 128th power, while IPV6 can only achieve 10 to the 32th power. Second, ISO has clearly mentioned that the length of the new future network should exceed 128 bits. Third, the United States is developing a new network (IPV10) with 256-bit IP addresses. This is the same as the NEW IP proposed by Huawei, which is unanimously inclined to the viewpoint that the IPV6 address length is not enough. Evidence No. 1 of IPV9 EV001 indicates that the statement “IPV9 is a joke on April Fool's Day” disseminated by Shenyang, Fang Zhouzi and a few academicians

and authoritative network experts in China is a lie. IPV9 is a protocol with an official version number and a specific technical background. The original technical documentation for IPV9, TUBA, was published by IETF two years before IETF1606 and IETF1607. IPV9 is a technology officially approved by the IETF and issued with a version number.

IPV9 is not a closed network, but is connected to the Internet in foreign countries. IPV9 can completely build a pure IPV9 network, and then connect to the old network through a gateway. The relationship between the future network and the Internet is like the relationship between the new highway and the old highway, which can operate independently or be interconnected. IPV9 information can be used in IPV9 network content circulation, digital domain name is owned by China's own root server, in foreign countries to cut off the network channel and stop the exchange of top-level domain (TLD), the use of digital domain users are not affected. In the event of foreign intervention or accident to cut off the overseas access to the Internet, China's network can still maintain a safe and stable operation. The IPV9 address of the future network is basically 256 bits long, and can be expanded to 1024 and 2048 bits. It can be expressed in simplified decimal or variable length to meet various application scenarios, while IPv6 address only 48, 64, 128 bits version and incompatible.

The future network IPV9 effective address length 256-2048 bits, can be independently and bidirectionally addressed, and can meet the needs of the Internet of Things and digital currencies. The address block architecture of IPv6 restricts its cross-city mobile and bidirectional addressing. It must deduct the 32-bit network number of each international and operator. The effective URL is only 64 bits, which cannot meet the needs of the Internet of Things mobility and the number of URLs. Therefore, IPV9 is a milestone for China to maintain national network sovereignty, guarantee the security of network information, break the monopoly of the US Internet, and promote the rapid development of a new generation of Internet with independent control and secure interconnection.

III. Future Network Patent

China IPV9 has found a way to implement it, applying for patents and copyright protection. And there are many innovations in the concept of technology, and the development achievements have been remarkable. The main results are as follows: First, the program is complete. China's future network IPV9 has formed a complete technical solution, including many core technologies, such as naming and addressing technology, three-and-four-layer composite architecture technology, character direct routing technology, terminal analysis technology, compatible interoperability technology and new network security technology. Second, it is well equipped. China has developed the key equipment for the future commercial application of network IPV9, including root server, core router, parsing server and so on. Third, the standard is leading. In the domestic standard, the country has issued a number of technical standards based on decimal network and IPV9. Internationally, the core technology concept of IPV9 has been adopted into the international standard, and many technical solutions are about to apply to the international standard for approval. It is shown in figure 2.

国家标准化管理委员会部函

标委外函[2014]46号

关于参与ISO/IEC JTC1/SC6未来网络技术两项技术报告制定的证明函

工业和信息化部计划司：

你工作组关于提请出具中国专家参与ISO/IEC JTC1/SC6/WG7未来网络标准制定工作证明的申请收悉。经核实，ISO/IEC DTR 29181-2《信息技术-未来网络-问题陈述与要求-第2部分：命名与寻址》（Information technology-Future Network-Problem statement and requirements-Part 2:Naming and addressing）和ISO/IEC DTR 29181-5《信息技术-未来网络-问题陈述与要求-第5部分：安全》（Information technology-Future Network-Problem statement and requirements-Part 5: Security）分别与2014年3月28日和2014年3月10日通过技术报告草案（DTR）投票，下一步将正式发布为技术报告（TR）。

中国专家谢建平、顾家文和张庆松作为项目编辑主要负责起草了ISO/IEC DTR 29181-2技术报告草案。中国专家刘亚东和王浩作为项目编辑主要负责起草了该ISO/IEC DTR 29181-5技术报告草案。

以上情况，特此证明。

2014年5月14日

国家标准化管理委员会部函

标委外函[2015]49号

国家标准委国际合作部关于参与ISO/IEC未来网络技术两项技术报告制定工作证明的复函

工业和信息化部科学技术司十进制网络标准工作组：

你工作组关于提请出具中国专家参与ISO/IEC JTC1/SC6/WG7未来网络标准制定工作证明的申请书收悉。经核，由我国专家主导制定的《信息技术-未来网络-问题陈述与要求-第2部分：命名与寻址》（ISO/IEC TR29181-2）和《信息技术-未来网络-问题陈述与要求-第5部分：安全》（ISO/IEC TR19181-5）等两项技术报告于2014年12月4日正式发布。

此复。

2015年5月6日

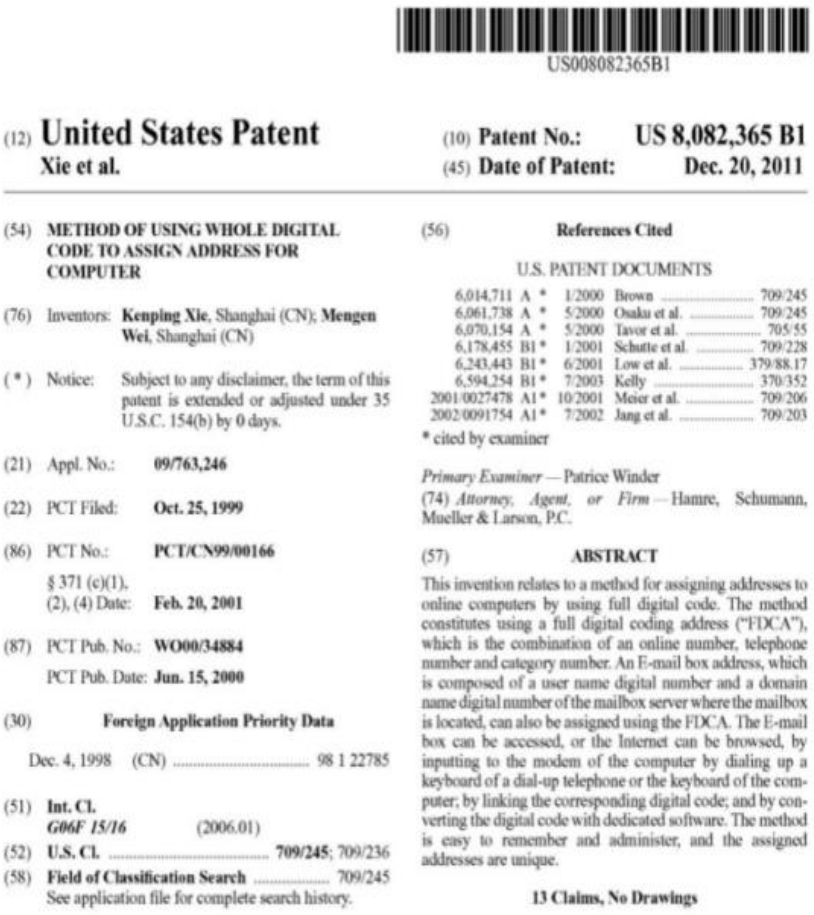
Figure 2. Official document of the future network

Domestic standards for future Networks (ISO-29181/IPV9) include: Electronic industry standard of the People's Republic of China SJ/T-11603-2016, SJ/T-11604-2016, SJ/T-11605-2016, SJ/T-11606-2016, SB/T-10530-2009, SJ/T-11271-2002. It is shown in Figure 3.

Future Network IPV9 has obtained 8 patents in China (2 inventions, 6 utility models). The method by which the computer assigns the address of the computer by the whole decimal algorithm (patent no.: ZL001351826). A method for the uniform compilation and distribution of addresses of networked computers and intelligent terminals (patent No.: ZL001168738). The method of assigning addresses to computers on the Internet by using full-digit code (patent No.: ZL981227856). Guidance code and its application system for goods and Commodity Code networking (patent No.: ZL200510027910X). A networked tax control system and its application method (patent No.: ZL2004100160308). Digital remote video monitoring system device (patent No.: ZL2004200207687).

The future network has also been awarded a US patent. It is shown in Figure 4.

Figure 3. Future Network China Standard Content



US008082365B1

(12) **United States Patent**
Xie et al.

(10) **Patent No.: US 8,082,365 B1**
(45) **Date of Patent: Dec. 20, 2011**

(54) **METHOD OF USING WHOLE DIGITAL CODE TO ASSIGN ADDRESS FOR COMPUTER**

(76) Inventors: **Kenping Xie**, Shanghai (CN); **Mengen Wei**, Shanghai (CN)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/763,246**

(22) PCT Filed: **Oct. 25, 1999**

(86) PCT No.: **PCT/CN99/00166**
§ 371 (c)(1),
(2), (4) Date: **Feb. 20, 2001**

(87) PCT Pub. No.: **WO00/34884**
PCT Pub. Date: **Jun. 15, 2000**

(30) **Foreign Application Priority Data**

Dec. 4, 1998 (CN) 98 1 22785

(51) **Int. Cl.**
G06F 15/16 (2006.01)

(52) **U.S. Cl.** **709/245**; 709/236

(58) **Field of Classification Search** 709/245
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,014,711	A *	1/2000	Brown	709/245
6,061,738	A *	5/2000	Osaku et al.	709/245
6,070,154	A *	5/2000	Tavor et al.	705/55
6,178,455	B1 *	1/2001	Schutte et al.	709/228
6,243,443	B1 *	6/2001	Low et al.	379/88.17
6,594,254	B1 *	7/2003	Kelly	370/352
2001/0027478	A1 *	10/2001	Meier et al.	709/206
2002/0091754	A1 *	7/2002	Jang et al.	709/203

* cited by examiner

Primary Examiner — Patrice Winder
(74) *Attorney, Agent, or Firm* — Hamre, Schumann, Mueller & Larson, P.C.

(57) **ABSTRACT**

This invention relates to a method for assigning addresses to online computers by using full digital code. The method constitutes using a full digital coding address ("FDCA"), which is the combination of an online number, telephone number and category number. An E-mail box address, which is composed of a user name digital number and a domain name digital number of the mailbox server where the mailbox is located, can also be assigned using the FDCA. The E-mail box can be accessed, or the Internet can be browsed, by inputting to the modem of the computer by dialing up a keyboard of a dial-up telephone or the keyboard of the computer; by linking the corresponding digital code; and by converting the digital code with dedicated software. The method is easy to remember and administer, and the assigned addresses are unique.

13 Claims, No Drawings

Figure 4. IPV9 US Patent Certificate

IV. TYPICAL APPLICATIONS OF IPV9

At present, my country has built demonstration projects of IPV9 address space, root domain name server and IPV9 backbone optical cable system in Beijing, Shanghai, Shandong, Jiangsu and Zhejiang, and is building a national military-civilian fusion IPV9 backbone optical cable and gateway bureau. The IPV9 network has now completed multi-point testing applications and has obtained good test data.

At present, China has established an "N" financial root domain name server which supports 256 bit address space. It has laid a foundation for the unified format of digital currency in China and even the whole world. In the process of issuing digital currency, the root domain name server and the top-level domain name server in the United States can avoid the management control of the overall digital currency issuing network communication system in China.

China's digital currency network communication system must have a financial root domain name server parallel to the United States and a "chn" national top-level domain name server and its supporting digital currency electronic vouchers, payment processing and other security service facilities, and adopt advanced advanced authentication before communication Technology and supporting domestic encryption technology

completely solve the problem of financial information security, ensure China's financial stability, and safeguard national sovereignty. At the same time, the establishment of a third-party platform for digital currency and physical currency conversion, electronic bills and electronic business path and identity and qualification certification based on the decimal network root domain name server, and unified national prior identification and management.

The project of Healthy Tai 'an IPV9 big data platform relies on the existing backbone optical cable and user transmission and access network of Tai 'an Branch of Shandong Radio and Television Network Co., LTD., and USES IPV9 network technology for upgrading and reconstruction. The network covers medical and health institutions at city, county, township and village levels as well as Tai'an City financial bureau, medical insurance bureau and administrative departments of Tai'an. The bandwidth meets the requirements of big data business of Tai'an health and sustainable expansion, and the compatible and safe operation of IPV9 network and IPV4 network is realized.

V. CONCLUSION

IPV9 is a new generation network architecture researched and developed by Chinese scholars. It is fully autonomous and controllable, with large address space and safe high-speed large code stream transmission. The distributed analysis of the network has low latency and is compatible with the current Internet system. Future Network is a method of empty cup design and new architecture to develop a Network system independent of the existing Internet to achieve a more secure, more economical, faster and more flexible Network. The future network will be developed in 15 years and put into preliminary commercial use around 2020.

REFERENCE

[1] RFC - Internet Standard. Internet Protocol, DARPA INTERNET PROGRAM PROTOCOL SPECIFICATION, RFC 791, 1981.09.

[2] S.Deering, R. Hinden, Network Working Group. Internet Protocol, Version 6 （IPv6）-Specification, RFC-1883, 1995.12.

[3] M. Crawford. Network Working Group. Transmission of IPv6 Packets over Ethernet Networks. RFC-2464, 1998.12.

[4] J. Onions, Network Working Group. A Historical Perspective on the usage of IP version 9. RFC1606. 1994.04.

[5] V. Cerf, Network Working Group. A VIEW FROM THE 21ST CENTURY, RFC1607. 1994.04.