The Prediction of New Medical Resources in China during COVID-19 Epidemic Period Based on Artificial Neural Network Model Optimized by Genetic Algorithm

Fumin Sun¹, Jie Su¹, Yong Hu¹ and Ling Zhou²*¹

¹School of Management Science and Engineering, Central University of Finance and Economics, Beijing, China
²School of Social Development and Public Policy, Beijing Normal University Beijing, China

*Corresponding author email: lingzhou_bnu@163.com

Abstract. In this paper, a neural network model is constructed to predict the number of new infections, and on the basis of this model, the demand for medical resources of new infections is further predicted. Firstly, the traditional Back Propagation neural network is optimized by genetic algorithm, which solves the problem that Back Propagation neural network is easy to fall into the local optimal solution in the prediction. Taking the number of severe patients, the number of cured patients, the number of deaths, and the number of suspected patients on the first day of the new crown period as the input variables, and the number of newly diagnosed patients on the second day as the output variable, a new Back Propagation neural network is constructed. Secondly, according to the official documents on the new crown pneumonia research and the guide for the use of medical resources, the linear function of the infection proportion and the normal distribution function of each symptom duration are constructed. Finally, combined with the above two contents, a complete process is designed to achieve the medical treatment for the newly infected patients. Prediction of the demand for medical resources.

Keywords: Genetic algorithm; Artificial neural network; Optimization; New crown infection number; Medical resources.

1. Introduction

COVID-19 has a huge impact on the whole world. There are two key issues throughout the epidemic, what is the trend of the number of patients and whether the medical resources are sufficient. In view of these problems, many experts give their own opinions. Artificial Neural Network (ANN) models have become more and more popular in the field of prediction in recent years. Yan et al. used Artificial Neural Network to predict the recovery of traffic flow, bus flow and truck flow during the epidemic, providing a way for railway departments to monitor traffic flow when responding to the national policy of resuming work and production [1]. Ye et al. took various factors affecting adolescent COVID-19 coping ability as input variables and established BP Neural Network to predict and analyze comprehensive coping ability [2]. In the process of fighting against the epidemic, the prediction of medical resources can provide a powerful reference for the government in resource dispatching. Xu et al. took the human resource scheduling of large 3A-grade hospitals during the epidemic as the main research object to provide references for the human resource scheduling of hospitals [3]. Huang et al. considered the influence of environmental factors
(temperature and humidity) as well as the implementation of control measures, to establish the global prediction system\cite{19}. Jin pointed out that many existing forecasting models lacked a reliable model basis and were ineffective, so he built a new mathematical model to analyse the scale of the epidemic\cite{20}. Liang et al. used the deep learning method to predict the probability of patients being critically ill within 5 days, 10 days and 30 days respectively, so as to conduct early triage for patients reasonably. With AILab technology as the core, this paper determines 10 patient characteristic indicators through machine learning selection variable algorithm. This model can predict the risk of COVID-19 progressing to critical illness based on the clinical characteristics of COVID-19 patients on admission\cite{21}. Wang et al. used the exponential smoothing model to predict the COVID-19 epidemic in Hubei Province\cite{22}, but the long-term prediction effect of the exponential smoothing method was poor, so we needed more advanced prediction models.

2. Data Source

2.1. Data on COVID-19 Patients

2.1.1. Number of COVID-19 patients. The data of the patients in this study were obtained from the website AkShare's Online Documentation, whose website is https://www.akshare.xyz/zh_CN/latest/. The site is a Python-based financial data interface, where data can be crawled directly by Python. In order to better cover the entire epidemic period, relevant data from January 20, 2020 to June 23, 2020 were selected. As for the data for July 2020 and beyond, due to the fact that the data of each variable is relatively small compared with the previous data and the epidemic situation is basically under control, so it is not selected. These data mainly include variables such as "confirmed", "death", "cured", "input", "severe" and "suspect". Since the activation function adopted in this paper is an S-type function, the function will be relatively gentle when the input value is large. In this paper, the values of "dead", "heal", "severe", and "suspect" the previous day are used to predict the value of "confirm" the next day. The specific data is shown below:

|     | confirm | dead | heal | input | severe | suspect |
|-----|---------|------|------|-------|--------|---------|
| 2020-01-20 | 0.000000 | 0.000000 | 0.000000 | 0.0 | 0.000000 | 0.001787 |
| 2020-01-21 | 0.001801 | 0.000948 | 0.000000 | 0.0 | 0.008516 | 0.001299 |
| 2020-01-22 | 0.003384 | 0.002377 | 0.000038 | 0.0 | 0.007932 | 0.013511 |
| 2020-01-23 | 0.006514 | 0.004105 | 0.000115 | 0.0 | 0.014778 | 0.036973 |
| 2020-01-24 | 0.012038 | 0.007583 | 0.000166 | 0.0 | 0.019786 | 0.067930 |

In addition, as the input value is small compared with other variables, this value is deleted in the analysis. Finally, the four input variables identified in this paper are: severe number, cured number, deaths number, suspected number. The output variable is newly diagnosed number.

2.1.2. Proportion of COVID-19 patients. According to relevant information, the overall proportion of each symptom in China can be determined as follows: mild patients: 22.8%, ordinary patients: 57.4%, severe patients: 13.8%, critical patients: 4.7%\cite{5}. The relevant data of France during the epidemic are as follows: mild patients: 30%, ordinary patients: 56%, severe patients: 9.4%, critical patients: 4%\cite{6}. We can see the difference is small, it can be considered that the proportions of symptoms in the two countries are similar, they are likely to satisfy the same epidemic dynamics model.

2.1.3. Symptoms of COVID-19 patients. COVID-19 infections can be classified into the following categories according to the degree of infection:

1) Mild patients: these patients have mild symptoms, either asymptomatic or mild symptoms.
2) Ordinary patients: these patients show more obvious symptoms and are able to recover by themselves to a large extent.
3) Severe patients: these patients have more severe symptoms and need oxygen treatment.
4) Critical patients: These patients have the most severe symptoms and need to be admitted to the ICU for emergency treatment.
2.1.4. The hospitalization time. According to the joint report of China and the World Health Organization, the median time from onset to clinical recovery is about 2 weeks for mild cases and about 3-6 weeks for severe or critical cases, the time from onset to severe symptoms such as hypoxia was 1 week. Among the patients who died, the time range from onset to death was 2-8 weeks.

2.2. Data on Medical Resources

At present, China adopts the following measures to treat patients with different symptoms:

Medical care shift system (4 shifts) for all patients, working hours (6h/d). For mild patients: 125:1 for doctors and 20:1 for nurses. For Ordinary patients: 10:1 for doctors, 3:1 for nurses and 10:1 for care workers. For severe patients: 5:1 for doctors, 3:1 for nurses and 5:1 for care workers. For critical patients: 1:1 for doctors, 3:1 for nurses, 2:1 for care workers, 1:1 for ventilators.

The following materials are also required: protective clothing, isolation gown (change per day), N95 medical masks (change per day), medical protective masks (change per four hours), goggles, boot/leg covers (change per day), shoe covers (change per day), gloves (change per day), and isolation caps.

3. Study on Prediction Model of the Number of Infections

Studies on the trend of COVID-19 infection can be divided into two categories, based on statistical methods or based on dynamic methods. In the study based on statistical methods Jin by establishing a mathematical model for China epidemic development of 30 provincial-level administrative region prediction, quantitative analysis on the scale of the outbreak, which indicates that the confirmed date in 30 provinces growth probability around February 19 tend to zero [16]. The only variable based on this method is the number of confirmed cases and the influence of other factors is ignored. Yang et al. obtained the specific number of infected people in Wuhan based on the analysis of the confirmed rate of COVID-19 in 50 cities in China [18]. In the study based on the dynamics method, Mei et al. proposed an IR prediction model based on the SIR model and the Extreme Learning Machine to predict the number of infected people, confirmed people and cured people during the epidemic in real time [17]. In this paper, the Artificial Neural Network is used to predict the number of infections.

3.1. Algorithm Design

3.1.1. The Meaning of optimization. The traditional BP Neural Network uses the "gradient descent" in the solution process, but the gradient descent is easy to fall into the local optimal solution in the solution process. This may result in the following two situations for the obtained weight value:

1) Fall into local optimum
2) As the number of times increases, the gradient will become smaller and smaller, resulting in worse and worse effect for weight updating.

Genetic Algorithm can help us to find the global optimal solution. If the iterative process of Neural Network can be improved and the search scope can be rapidly reduced, the domain of optimal solution can be found in a great possibility. Therefore, this paper will use Genetic Algorithm to optimize the construction process of Artificial Neural Network, in order to effectively improve the prediction effect of artificial neural network, and at the same time effectively improve the operation speed.

3.1.2. The process of optimization. The optimization of neural network method has a lot of kinds, such as optimize the topology structure or parameters, this paper is to use genetic algorithm to optimize parameters of neural network, make the parameter falls in the domain of the optimal solution as soon as possible, avoid the local optimal solution, if the optimal solution can't meet the requirements, then using traditional BP Neural Network model based on the values of parameters to simulate again.

Mathematical model:

For the i-th neuron, x1, X2, XJ are the inputs of neurons, and the inputs are usually the independent
variables that have a key impact on the system model, $W1, W2, WJ$ is the connection weight to adjust the weight ratio of each input. There are many ways to input signals to neurons. In this paper, we use S-type activation function. Finally, through this neuron, the output results are as follows:

$$\text{net} = x_1w_1 + x_2w_2 + \ldots + x_jw_j$$

$$y_k = f(\text{net}) = \frac{1}{1+e^{-\text{net}}}$$

Compared with the traditional BP neural network, this paper uses genetic algorithm to optimize its parameters in the iterative process, which needs to calculate the loss value and fitness value. Then, the roulette method is used to iterate its parameters. The corresponding values are as follows:

$$\text{Loss} = (y_k - y)^2$$

$$\text{Fitness} = 1/\text{Loss}$$

3.2. Algorithm Implementation

3.2.1. The determination of neural network topology. The neural network in this paper is three-layer, in which the input layer and the output layer are well determined, and their values are equal to the number of input variables (4) and output variables (1) respectively. When determining the intermediate variables, this paper refers to the following formula: Number of middle layer neurons = number of input layers * 2 + 1 = 9. Therefore, the number of neurons in the middle layer in this paper is 9, and the s-type activation function is selected as the activation function. At the same time, a dictionary is defined to set the initial parameters of the neuron topology structure.

3.2.2. Function involved.

```python
mn_architecture = [
    "input_dim": 4, "output_dim": 9, "activation": "sigmoid",
    "input_dim": 9, "output_dim": 1, "activation": "sigmoid"
]
```

The created weight parameters are saved in a dictionary of params_values, the key values of the dictionary are "w1", "w2", "B1" and "b2", where "w1" is a 4*9 array, "w2" is a 9*1 array, "B1" is a 9*1 array, and "b2" is a 1*1 array.

```python
def single_layer_forward_propagation(A_prev, W_curr, b_curr, activation="relu"):
    This function is a single neuron transmission function created to simulate the process of neuron transmission from one layer to another. The purpose is to make the conduction process of our neural network composed of many such processes and can adapt to a variety of neural network structures.

def full_forward_propagation(X, params_values, mn_architecture):
    The function can integrate the previous functions to realize the numerical transfer in neural network.

3.2.3. Genetic function.

```python
def initial_population(x=6):
    This is a function to generate the initial population, but the generated initial population is the parameter value we need in this neural network, namely the values of "w1", "w2", "B1" and "B2".

def theloss(x, x1, x2, y):
    First, the loss function is defined. The activation function of neural network is used in the calculation. For a single sample, the specific calculation process is as follows:

$$\text{Loss} = \left(\frac{1}{1+e^{-(x * w1)}} \cdot w2 - y\right)^2$$

Due to the large number of samples, it is a multi-dimensional vector. According to the above function, the result is a column of vectors, and the mean value of the vector is taken as the loss value.

def fitness(population, X_train, Y_train):
    This function is used to calculate fitness, but different from the common genetic algorithm, the measurement of fitness here comes from the previous loss function. The greater the loss is, the worse
the adaptability is. The fitness is defined as: Fitness=1/loss. There are six populations in the genetic algorithm, so the fitness functions of six populations can be obtained in each round of output.

```python
def selector(population, fitness):
    This function is used to realize the process of natural selection and complete the process of crossover and mutation needed in the genetic process. The crossover probability is set as 0.5 in this paper, and the mutation probability as 0.1.
```

```python
def train(X, Y, nn_architecture, epochs, learning_rate):
    This function is used to complete the parameter iteration process required in this article. The epochs is the number of cycles, which is set to 1000 in this article.
```

3.2.4. Analysis of prediction results. In this paper, the called data is divided into training set and test set, among which training set accounts for 70% of the total data set and corresponding test set accounts for 30%. In this paper, a total of 154 samples are obtained, corresponding to 108 training samples and 46 test samples. The selection process is randomly selected from the data set.

4. New Medical Resource Demand Prediction Based on Prediction Model

The prediction of medical resources in this paper aims to predict how much more medical resources will be needed by newly infected people. This paper first determined the proportion of patients with various symptoms, then determined the length of stay of patients with various symptoms based on the existing data, and finally calculated the number of new patients who would increase the demand for medical resources according to the existing treatment measures for patients with different symptoms.

4.1. Ratio of Symptoms

A simple analysis can be made of the proportion of COVID-19 patients with different symptoms in China and France, as shown in the graph of the transformation of symptoms in France:

![Figure 1. Number of COVID-19 infections and change in proportion in France](image)

According to the data from France, 30% of the infected people have mild symptoms, 80% of these symptoms will recover by themselves and do not need treatment, but 14% need to be sent to the hospital for oxygen delivery treatment, and 6% need to be sent to the ICU for emergency treatment [6]. The relationship between COVID-19 infection rates in China is shown in the figure below [5].

![Figure 2. Schematic diagram of COVID-19 disease progression in China](image)

The size of the severity and prognosis boxes in the figure reflect the proportion of cases reported as of 20 February 2020. The size of the arrow indicates the proportion of cases that recover or die. According to the data of China and France, after infection, an infected person will have four symptom
types. These symptoms are not isolated from each other, but more likely to be manifested as mild symptoms at first, possibly recovering, and then common symptoms, recovery or further deterioration. Based on the above data, in order to better simulate the proportion of the above symptoms and the conversion relationship, the following process was carried out in this paper:

1) Determine the proportion of infected patients presenting with mild symptoms. Based on 22.8% of China, assuming that it satisfies a linear distribution, the confidence interval of 90% is [11.4%, 34.2%]

2) Some people with mild symptoms heal on their own, while others get worse. The proportion of the conversion is generated by the proportion of mildly symptomatic infections determined based on 1) and a random number generator. The ratio of cure and the ratio of further deterioration (1-cure ratio) were finally calculated.

3) Circulation 2), finally the infection ratio of severe patients and critically ill patients was obtained

In the whole operation process, the following three conditions should be followed:

1) There is a 90% chance that the generated data will fall within this range

2) Its value is randomly generated by a random number generator between (0,1) when determining further worsening symptoms

3) All disease rates follow a linear distribution

4.2. The Hospitalization Time

In this paper, it is assumed that the hospitalization time follows a normal distribution, so a normal distribution function can be fitted according to the limit range of clinical time of each symptom. The hospitalization time of each person can be generated according to this normal function. In this paper, patients with mild symptoms and patients with normal symptoms shared a normal function distribution of hospitalization time, severe and critically ill patients share a normal function distribution. It is defined that the cure time of mild patients and ordinary patients is about 1-3 weeks, and the hospitalization time of severe patients and critically ill patients is about 2-8 weeks. Assuming that they obey the normal distribution, the following is the normal distribution function of the hospitalization time of patients based on 90% confidence level and 95% confidence level.

The distribution function of hospitalization time for each symptom under 90% confidence level:

1) The distribution function of cure time for mild patients and ordinary patients is the normal distribution function of N (2,0.37):

$$\frac{1}{\sqrt{2\pi 0.61}} e^{\left(-\frac{(x-2)^2}{0.74}\right)}$$

2) Distribution function of cure time for Severe patients and critical patients:

$$\frac{1}{\sqrt{2\pi 1.83}} e^{\left(-\frac{(x-5)^2}{3.35}\right)}$$

The distribution function of hospitalization time for each symptom under 95% confidence level:

1) The distribution function of cure time for mild patients and ordinary patients is the normal distribution function of N (2,0.26):

$$\frac{1}{\sqrt{2\pi 0.51}} e^{\left(-\frac{(x-2)^2}{0.26}\right)}$$

2) Distribution function of cure time for severe patients and critical patients:

$$\frac{1}{\sqrt{2\pi 2.55}} e^{\left(-\frac{(x-5)^2}{6.51}\right)}$$

4.3. Calculated Consumption

According to the above steps, this paper has the number of new COVID-19 patients, the proportion of infection with different degrees of symptoms and the ratio of interconversion, the hospitalization time
of patients with different symptoms and the consumption of medical resources of infected patients with each symptom. The steps for calculating the consumption of each resource are shown as follows:
1) According to the proportion function of mild patients, the proportion of mild patients was randomly generated, and according to the random number generator, the proportion of ordinary patients, the proportion of severe patients and the proportion of critical patients were determined
2) The number of newly infected persons with each symptom was calculated according to the proportion of each symptom
3) The for loop was used to calculate the hospitalization time of each person according to the normal distribution function of the hospitalization time of each symptom, and according to the hospitalization time, the consumption of resources of each person belonging to each symptom was calculated
4) Summation
Finally, the results are as follows:

Table 1. Consumption of medical resources

| Demand for medical personnel         | 13–15 |
|-------------------------------------|-------|
| The doctor                          | 13–15 |
| The nurse                           | 38–45 |
| The care worker                     | 12–15 |

| Cumulative demand for medical personal protective equipment |
|-------------------------------------------------------------|
| Protective clothing                                         | 59–65 |
| Isolation gown                                              | 63–67 |
| N95 medical masks                                           | 59–65 |
| Medical protective masks                                    | 378–452|
| Goggles                                                     | 63–70 |
| Boot/leg covers                                             | 563–688|
| Shoe covers                                                 | 563–688|
| Gloves                                                      | 563–688|
| Isolation caps                                              | 563–688|

| Hospital beds and medical equipment required                |
|-------------------------------------------------------------|
| Ordinary hospital bed                                       | 132–141|
| Isolation bed                                              | 80–92 |
| ICU inpatient bed                                           | 29–33 |
| Ventilators                                                 | 29–33 |
| Oxygen inhalation equipment                                 | 29–33 |

According to the above forecast, medical resources such as gloves, shoe covers and isolation caps will require more new medical resources because they are disposable, which is much higher than other medical resources. Ventilators and oxygen inhalation equipment will also require less patients due to their small proportion of patients. This suggests that our predictions are in line with reality. Since the resource forecast quantity given in this paper is a range, it is suggested that the decision can be made according to the decision habit of the decision-maker.

5. Conclusions
In this paper, the traditional BP neural network is optimized by genetic algorithm, so that the neural network can avoid falling into the local optimal solution during prediction. According to the experimental data, the loss value of the prediction results of the improved algorithm declines faster and reaches less than 0.09315, which shows a good simulation effect. When predicting the demand of new patients for medical resources, this paper uses linear function to fit the proportion distribution of patients with various symptoms, and uses normal distribution function to fit the distribution of the hospitalization time of patients with various symptoms, and finally calculates the range value of the total demand of new patients for various medical resources. On the whole, the relative size of these values is consistent with the actual allocation of resources. However, from the point of the whole process, in this paper, based on artificial neural network model to predict resource requirements, is
based on "people infected with the supply of medical resources in the first place", under this assumption, however, in the whole course of the disease resistant medical resources assigned to uninfected also is the key step in the outbreak, should be in the same important position.

Acknowledgment

This work was supported by National Key R&D Program of China (Grant No. 2018YFC0809900).

References

[1] MY Yan, MR Shen, SY Huang, et al. Study on highway Traffic Flow Recovery Prediction during COVID-19 epidemic
[2] CM Ye, SW Zhao, XH Yang, et al. Analysis and prediction of COVID-19 outbreak response ability of young Athletes based on machine learning
[3] X Xu. A brief analysis of human resource allocation Strategy in large Grade A hospitals during epidemic situation
[4] https://www.akshare.xyz/zh_CN/latest/
[5] Chinese Center for Disease Control and Prevention. Analysis of epidemiological characteristics of COVID-19 [J]. Chinese Journal of Epidemiology, 2020, 41(2): 145-151. DOI: 10.3760/cma.j.issn.0254-6450.2020.02.003.
[6] https://www.zhihu.com/question/379446346
[7] http://www.nhc.gov.cn/jkj/s3578/202002/87fd92510d094e4b9bad597608f5cc2c/files/fa3ab9461d0540c294b9982ac22aff64d.pdf
[8] Xianlian Zhang. A review of hybrid genetic algorithms
[9] Guan WJ, Ni ZY, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China [J]. NEJM, 2020. DOI:10.1056/NEJMoa2002032. [published online ahead of print, 2020, Feb 28]
[10] World Health Organization. Operational considerations for case management of COVID-19 in health facility and community. 2020[EB/OL].(2020-03-19)[2020-04-01].https://www.who.int/publications-detail/operational-considerations-for-case-management-of-covid-19-in-health-facility-and-community.
[11] National Bureau of Statistics. China statistical yearbook-2019[M].China Statistic Press,2019.
[12] A classified emergency medical system was established to respond to the overburdened medical needs of Wuhan following the COVID-19 community outbreak. Chinese Journal of Infectious Diseases, 2020, 38. DOI: 10.3760/cma.j.cn311365-20200316-00314.
[13] Xiying Zhang. BP neural network based on genetic algorithm optimization for stock price prediction.
[14] Guiliang Dai. The combination of artificial Neural networks and Genetic algorithms: Progress and Prospect.
[15] Huiqin Chen. Theory and Application of genetic Algorithm based on artificial Neural network.
[16] Qixuan Jin. Predictive modeling and rational assessment of COVID-19 in China. R563.1;R181.8
[17] WJ Mei, Zhen Liu, JY Zhu, et al. COVID-19 Limit IR real-time prediction model.
[18] Z Yang, ZX Yuan, ZY Jia. The number of infected virus novel Coronavirus patients in Wuhan was estimated based on migration data [J]. Journal Of University Of Electronic Science And Technology Of China, 2020, DOI:10.12178/1001-0548.2020030.
[19] J. Huang et al. Global prediction system for COVID-19 pandemic[J]. Science Bulletin 65 (2020) 1884–1887.
[20] X Jin. Predictive modeling and rational assessment of COVID-19 in China [J]. Statistics & Decision, 2020,(5):11-14.
[21] Wenhua Liang, Jianhua Yao, Ailan Chen, et al. Early triage of critically ill COVID-19 patients using deep learning. 2020, 11(1):3543.
[22] X. Wang, Y.Yu, Y.HU et al. COVID-19 analysis and forecast based on Exponential Smoothing Model in Hubei Province [J]. Pub Health and Prev Med, 2020, 31(01):1-4.