Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 59-70
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

Research on the Distance-based Outlier Mining Algorithm to Identify the Outpatient Visits of Teachers and Students During Holidays

Tieshi Song[1], Huipeng Jin[2], Jiutao Zhang[3], Qi Zeng[4]*

[1]Information Center, Hongqi Hospital Affiliated to Mudanjiang Medical University, Mudanjiang 157000, Heilongjiang Province, China

[2] Department of Ophthalmology, Hongqi Hospital Affiliated to Mudanjiang Medical University, Mudanjiang 157000, Heilongjiang Province, China

[3] School of Health Management, Mudanjiang Medical University, Mudanjiang 157011, Heilongjiang Province, China

[4] Continuing Education Institute, Mudanjiang Medical University, Mudanjiang 157011, Heilongjiang Province, China

*Corresponding author.

Abstract:

Background and Purpose: The health of teachers and students is particularly important to the development of the school. Forecasting of outpatient visits are benefit for hospital work efficiency. Research Methods: For forecasting of daily outpatient visits of teachers and students, distance-based outlier mining algorithm is proposed. In this approach, an algorithm for mining outliers based on distance is presented to compute holiday effect' s effective time. Solar term, as minimal time units of climate change, combining with other attributes, is used to describe the outpatient visits data of teachers and students.

***Keywords**: Outlier mining algorithm, Outpatient visits*

I. INTRODUCTION

As the first window of hospital medical services, the outpatient clinic is not only responsible for providing patients with early diagnosis and timely treatment, but also collecting information and data on different types of patients. The outpatient service level directly affects the patient's medical experience and the hospital's work efficiency [1]. Outpatient visits forecasting is an activity that predicts and judges its future development trends and conditions based on the development and change rule of outpatient visits.

Tandberg and Qualls [2] first used the time series analysis method in 1994 to predict the number of hospital emergency patients and length of hospital stay, etc., and proposed the first research on the prediction of hospital outpatient visits. Since then, scholars at home and abroad have carried out a lot of research on the prediction of outpatient visits. Specifically, the research on the prediction of outpatient visits can be divided into the following three methods: (1) Statistical prediction method, which is, using mathematical statistical methods to study the change rule of outpatient visits. For example, Li et al. [3],

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 59-70
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

Tao et al. [4], and Luo et al. [5] used Seasonal Autoregressive Integrated Moving Average Model (SARIMA model) to predict the monthly outpatient visits. Because the SARIMA model assumes a linear relationship between the data, the prediction effect for non-linear data is poor. Garg et al. [6] and Cheng et al. [7] used the fuzzy time series method to predict the monthly outpatient visits; Jiang and Hu [8] used the Markov prediction model to predict the monthly outpatient visits, but this method is only applicable to short-term outpatient visits prediction. Fang et al. [9] used the seasonal index-least square method to predict the number of monthly outpatient visits. This method obtained the seasonal index of each month firstly, and then combined the least square method to find the linear regression equation; Guo et al. [10] used the generalized autoregressive conditional heteroscedasticity model to predict the number of monthly outpatient visits; Zhang [11] and Ma et al. [12] used the gray model to predict the annual outpatient and monthly outpatient visits, respectively. This method is more effective for the original data showing an exponential change trend, and the prediction accuracy for non-exponential data will decrease; (2) Artificial intelligence method, that is, using artificial intelligence algorithms (such as neural networks, genetic algorithms) to learn the rules of outpatient visits data, such as Yang et al [13] proposed a method based on deep belief network, which conducted outpatient visits prediction in units of day, week, and month. This method firstly used the greedy layer-by-layer unsupervised algorithm to train the deep belief network, and then used the gradient optimization method to optimize the network parameters, but the method does not consider seasonal factors; Ye et al [14] use the seasonal neural network algorithm to predict the monthly outpatient visits, but it takes more time to design and train the seasonal neural network; Hadavandi et al. [15] proposed a cluster-based genetic fuzzy system to predict the monthly outpatient visits. This method firstly used clustering algorithm to segment the data, and then used genetic fuzzy system for segmentation prediction; Jiang et al. [16] used deep neural networks to make daily outpatient visits predictions; Wang et al [17] used the firefly algorithm-support vector regression algorithm to predict monthly outpatient visits for diarrhea patients; (3) Hybrid method, that is the combination of statistical prediction and artificial intelligence methods. For example, Yang [18] used the gray model-neural network to predict the monthly outpatient visits; Zhang et al. [19] also used the combination of gray model and neural network, the original data was transformed by the gray model, and the transformed data was the training samples of the neural network to predict the monthly outpatient visit; Zhang et al. [20] use genetic algorithms to determine the Autoregressive Moving Average Model (ARMA) parameters, and realized the monthly outpatient visits prediction.

Statistics on the medical treatment of teachers and students in the school are helpful to understand the physical health of teachers and students, do a good job in prevention and diagnosis and treatment, and ensure the normal teaching and learning of the school. So it is necessary to study an effective method to solve this problem.

This paper proposes a daily outpatient visits prediction method based on mathematical algorithm. Firstly, the distance-based outlier mining algorithm is used to identify the effective time range of the festival effect, and the solar term is used as the minimum time unit to represent climate change, and several other attributes are used to describe the historical data of teachers and students.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 59-70
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

II. DATA PREPROCESSING AND HOLIDAY EFFECT IDENTIFICATION

2.1 Data Description

The data used in this article is the historical data of a total of 1096 daily outpatient visits from January 1, 2015 to December 31, 2017 in a top three hospital in Mudanjiang.

2.2 Data Preparation

With reference to related literatures [13, 21], this article defines the characteristics or attribute set that describing the daily outpatient visits data as: {year, season, month, day, weekday, festival, festIndex, outpaNo}

Among them, year = {2015,2016,2017} indicates the year; season = {1,2,3,4} indicates the quarter, 1 is the first quarter, and so on; month = {1,2,…, 12} indicates month, 1 is January, and so on; day = {2015/01/01,…, 2017/12/31} indicates a specific date; weekday = {1,2,…, 7} indicates the day of the week , 1 is Monday, and so on in analogy; festival = {0,1,2,3,4,5,6,7} indicates the holiday type, 0 indicates non-holiday, 1 for New Year's Day, 2 for Spring Festival, and 3 for Qingming , 4 is the May Day Labor Day, 5 is the Dragon Boat Festival, 6 is the Mid-Autumn Festival, and 7 is the National Day; festIndex = {1,2, ..., n} represents the day of a holiday, and 1 is the first day of the holiday, and so on; so outpaNo indicates the outpatient visits for that day.

Table 1 shows the daily outpatient visits within a period of time before and after the Spring Festival in 2017. The data shows that the daily outpatient visits during the Spring Festival is decreased obviously.

TABLE I. Daily outpatient visits of period around the Spring Festival of 2017

day	*year*	*season*	*month*	*weekday*	*festival*	*festIndex*	*outpaNo*
2017/2/1	2017	1	2	1	0	-	6638
2017/2/2	2017	1	2	2	0	-	6106
2017/2/3	2017	1	2	3	0	-	5483
2017/2/4	2017	1	2	4	0	-	4945
2017/2/5	2017	1	2	5	0	-	4211
2017/2/6	2017	1	2	6	0	-	2708
2017/2/7	2017	1	2	7	2	1	**747**
2017/2/8	2017	1	2	1	2	2	**963**
2017/2/9	2017	1	2	2	2	3	**1101**
2017/2/10	2017	1	2	3	2	4	**1136**
2017/2/11	2017	1	2	4	2	5	**2670**
2017/2/12	2017	1	2	5	2	6	**2761**
2017/2/13	2017	1	2	6	2	7	**1664**

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 59-70
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

2017/2/14	2017	1	2	7	0	-	3427
2017/2/15	2017	1	2	1	0	-	6402
2017/2/16	2017	1	2	2	0	-	6792
2017/2/17	2017	1	2	3	0	-	6921
2017/2/18	2017	1	2	4	0	-	6539
2017/2/19	2017	1	2	5	0	-	6114

2.3 Prediction Evaluation index

Root Mean Square Error (RMSE) and Coefficient of determination (R2) were used as the evaluation index of the prediction result, which were defined as follows:

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

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Among them, $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$, n is the total number of daily outpatient data, yi is the ith daily outpatient visits, and $\hat{y}_i$ is the predicted value corresponding to yi.

2.4 Feature set Adjustment

The ARIMA model and the SARIMA model are currently commonly used time series prediction models to predict the daily outpatient visits [3, 4, 22]. The difference between the two models is that: SARIMA is more suitable for time series prediction with seasonal fluctuations. Figure 1 shows the prediction results of the daily outpatient visits for all three years, and Figure 2 shows the prediction results of February 2016 in detail. Among them, February 7, 2016 was the Spring Festival, and the daily outpatient visits for that day was 747, and the prediction result of SARIMA was 4532.

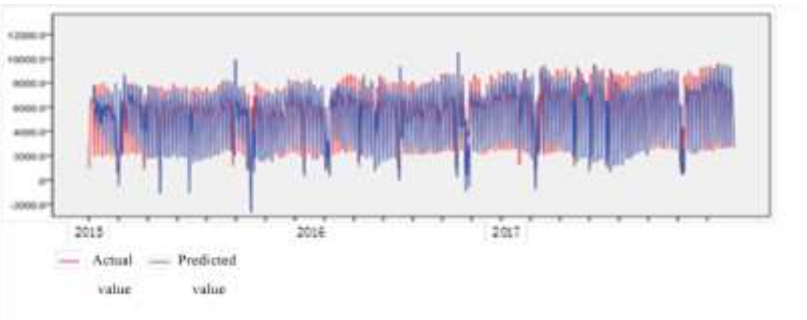

Fig 1. The forecast results from January 1, 2015 to December 31, 2017

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 59-70
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

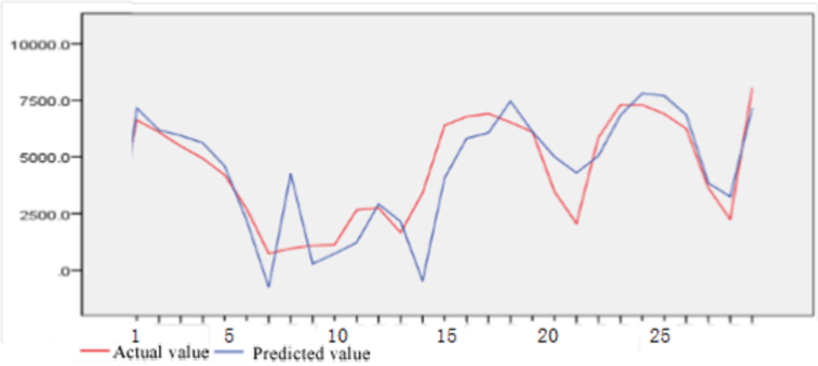

Fig 2 The forecast results of February 2016

The prediction results of SARIMA model showed that the predicted value of the daily outpatient visits has a large error from the actual value, especially on the holidays and the period before and after the holiday, the predicted value greatly deviate from the actual value. After analysis, the possible reasons for the large prediction error are as follows:

(1) The daily outpatient visits during holidays is significantly lower than that of non-holiday days, especially for important festivals, such as the Spring Festival and National Day;

(2) There is a holiday effect in the daily outpatient visits. The so-called holiday effect means that the daily outpatient data changes abnormally within a certain period of time before and after the holiday;

(3) There is a close relationship between the daily outpatient visits and the climatic conditions of the hospital's location [23], and the two temporal characteristics of the season and month are coarse in reflecting the climate change. There may be significant differences in climate on the same day of the same quarter and month in different years;

(4) The ARIMA and SARIMA models are insufficient in the prediction of daily outpatient visits. At present, these types of models are mainly used for monthly and quarterly outpatient visits prediction, such as references [3] and [4], and the prediction of daily outpatient visits is rare.

In addition, related medical research has shown that the “24 solar terms” has a significant effect on the disease [24]. Solar term is the science of seasonal change and climate change, which is summarized by the people in our long-term life practice. It reflects the laws of climate change in more detail than quarters and months. To this end, based on the analysis results of the SARIMA model and the fact that solar terms affect the disease, the feature set describing the daily outpatient data is adjusted to:

$$\{\mathrm{year}, \mathrm{month}, \mathrm{solarTerm}, \mathrm{weekIndex}, \mathrm{weekday}, \mathrm{festival}, \mathrm{festIndex}, \mathrm{festEffitive}, \mathrm{outpaNo}\}$$

Among them, the new feature is defined as follows: solarTerm = {1,2,…, 24} means 24 solar terms, 1

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 59-70
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

is Osamu, others are inferred in the order of “24 solar terms”; weekIndex = {1,2,3} means weeks in the current solar term, 1 means the first week in the current solar term, and so on. Since each solar term has a maximum of 16 days, the maximum value of weekIndex is 3; festEffitive = {-nb, ..,-1, 1,…, na} indicates the time range of the holiday effect, -1 indicates the first day before the holiday that has a holiday effect, 1 indicates the first day after the holiday that has a holiday effect, and the rest can be deduced by analogy. It should be noted that the holiday effect variables nb and na have the following characteristics: 1) The nb day before the holiday and the na day after the holiday must have a holiday effect; 2) There must be no holiday effect on the nb + 1 day before the holiday the na + 1 day after the holiday; 3) For the same holiday, nb and na may not be equal.

According to the adjusted feature set, Table 2 refreshed the data in Table 1 (nb = 2, na = 4 in the 2016 Spring Festival, calculation methods are given later). It should be noted that any day of the year can be uniquely determined by the values of the attributes year, solarTerm, weekIndex, and weekday.

TABLE II Daily outpatient visits of period around the Spring Festival of 2017 (new)

$year$	$solarT$	$montl$	$weekInde$	$weekda$	$festiva$	$festInde$	$festEffiti$	$outpaN$
2017	1	2	2	1	0	-	-	6638
2017	1	2	2	2	0	-	-	6106
2017	1	2	2	3	0	-	-	5483
2017	2	2	1	4	0	-	-	4945
2017	2	2	1	5	2	-	-2	4211
2017	2	2	1	6	2	-	-1	2708
2017	2	2	1	7	2	1	-	747
2017	2	2	2	1	2	2	-	963
2017	2	2	2	2	2	3	-	1101
2017	2	2	2	3	2	4	-	1136
2017	2	2	2	4	2	5	-	2670
2017	2	2	2	5	2	6	-	2761
2017	2	2	2	6	2	7	-	1664
2017	2	2	2	7	2	-	1	3427
2017	2	2	3	1	2	-	2	6402
2017	2	2	3	2	2	-	3	6792
2017	2	2	3	3	2	-	4	6921
2017	2	2	3	4	0	-	-	6539
2017	3	2	1	5	0	-	-	6114

2.5 Holiday Effect Identification of Outpatient

The holiday effect exists widely in the stock market, futures, and tourism. Obviously, there is a obvious holiday effect in daily outpatient visits. It refers to the effect of holidays on the fluctuation of the daily outpatient visits (excluding the holidays themselves), and there may be both pre- and post-holiday effects, that is, within a certain time range before and after the holiday, the daily outpatient visits changed

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 59-70
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

abnormally, or increased significantly or decreased significantly.

In order to identify the holiday effect of the outpatient visits, the time range of the holiday effect needs to be determined firstly, namely nb and na. This study will use the method of identifying outliers to determine the time range of the holiday effect, that is, if the daily outpatient visits of a day before or after holiday changes abnormally, the daily outpatient visit is the outlier of the daily outpatient data set in a certain time range that relative to that day. Outlier mining algorithms usually include statistics-based, distance-based, depth-based, density-based, and cluster-based methods [25]. Because the holiday effect of outpatient visit is only caused by holidays, the holiday effect of daily outpatient visit belongs to low-dimensional outlier mining. In this paper, a distance-based outlier mining algorithm was used to identify the time range of holiday effects. It should be noted that the distance-based outlier mining algorithm is particularly effective for low-dimensional data [25].

To determine the time range of the holiday effect should firstly determine whether the outpatient visit on a day before or after the holiday is an outlier. To this end, this article designed the following outlier recognition algorithm (referred to as l^{th}_Day_Check):

Step1: If the daily outpatient visit $\{outpaNo_x^l\}$ on the l^{th} day before or after the holiday x has the following characteristics, it is said that the daily outpatient visit on the l^{th} day is suspected to have a holiday effect.

$$outpaNo_x^l \geq v_u = Mean\_outpaNo_{db(x,l,k)} + \lambda(max_{db(x,l,k)} - min_{db(x,l,k)})$$

Or $\mathrm{outpaNo}_x^l \leq v_l = \mathrm{Mean\_outpaNo}_{db(x,l,k)} - \lambda(\max_{db(x,l,k)} - \min_{db(x,l,k)})$ (1)

Wherein, db (x, l, k) is a local data set composed of k + 1 daily outpatient visit, including $\mathrm{outpaNo}_x^l$ within a certain time range before and after the holiday x, that is, $db(x,l,k) = \{\mathrm{outpaNo}_x^l, \mathrm{outpaNo}_{x,1}, \mathrm{outpaNo}_{x,2}, \ldots, \mathrm{outpaNo}_{x,k}\}$, and satisfying $\mathrm{outpaNo}_x^l.\mathrm{weekday} = \mathrm{outpaNo}_{x,i}.\mathrm{weekday}$, and $\mathrm{outpaNo}_x^l$ is the normal day outpatient visit of non-suspected holiday effect or non-holiday, i = 1,2, ..., k; $\mathrm{Mean\_outpaNo}_{db(x,l,k)} = \frac{\sum_{i\in db(x,l,k)} \mathrm{outpaNo}_i}{|k+1|}$; λ is the adjustment coefficient, $\lambda \in \{0.6,1\}$. $\max_{db(x,l,k)}$ and $\min_{db(x,l,k)}$ are the maximum and minimum values of the daily outpatient visit in the data set db (x, l, k).

Step 2: Determine whether $\mathrm{outpaNo}_x^l$ is an outlier, that is: if $\mathrm{outpaNo}_x^l$ is an outlier $db(x,l,k)(p,r) -$ Outlier of db (x, l, k), at least p*100% of the data points in the data set db (x, l, k) have the distance from $\mathrm{outpaNo}_x^l$ greater than r. The parameters p, r are defined as follows:

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 59-70
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

$$p = \frac{k}{k+1} \tag{2}$$

$$r = \frac{\sum_{i,j \in db(x,l,k),\ i \neq j} |outpaNo_i - outpaNo_j|}{k(k+1)} \tag{3}$$

Take the 2016 Chinese New Year as an example (data is shown in Table 2). For the first day after the Chinese New Year (ie February 14, 2016), we should determine whether the day has a holiday effect. The day is Sunday, and assuming k = 8, λ = 0.6, the actual value of the data set db (7,1,9) is {3427,2158,1725,1687,2053,2074,2250,2469,2277}, where the first data (3427) is the daily outpatient visit of the first day, the other 8 data are the daily outpatient visits of the 4 normal dates on Sunday before and after that day. The calculation showed: $\mathrm{Mean\_outpaNo} = 2236.56$, $\max_{db(7,1,9)} = 3427$, $\min_{db(7,1,9)} = 1687$. The calculation result of formula (1) is vu = 3280.56, which is less than 3427. Therefore, it seems that there is a holiday effect on this day. From formulas (2) and (3): p = 0.89, r = 98.56. By calculation, it can be known that the distance between the daily outpatient visit of that day and the daily outpatient visit of other days in the data set db (7,1,9) is greater than r, so the daily outpatient visit of that day is the outlier of data set db (7,1,9): there is a holiday effect on the first day after the 2016 Spring Festival.

Based on the algorithm 1^{th}_Day_Check, the complete algorithm for determining the time range of the holiday effect of a holiday x is as follows:

Step1: $\mathrm{sum\_n_b} = \mathrm{sum\_n_a} = 0$

Step2: for each y∈year

Step2.1: Use 1^{th}_Day_Check algorithm to determine the holiday effect parameters $n_{b,y}$ and $n_{a,y}$ of this holiday in this year;

Step2.2: $\mathrm{sum\_n_b} \mathrel{+}= n_{b,y}$, $\mathrm{sum\_n_a} \mathrel{+}= n_{a,y}$

Step3: end for

Step4: Output the holiday effect parameters of holiday x: $n_b = \lceil \frac{\mathrm{sum\_n_b}}{|\mathrm{year}|} \rceil$, $n_a = \lceil \frac{\mathrm{sum\_n_a}}{|\mathrm{year}|} \rceil$, where $\lceil a \rceil$ represents the smallest integer greater than or equal to a.

Taking the Spring Festival as an example, calculated by 1^{th}_Day_Check algorithm, we can observe that $n_{a,2015} = 4$, $n_{a,2016} = 3$, $n_{a,2017} = 4$, $n_{b,2015} = 1$, $n_{b,2016} = 2$, $n_{b,2017} = 1$. Therefore, $n_a = \left\lceil \frac{4+3+4}{3} \right\rceil = 4$, $n_b = \left\lceil \frac{1+2+1}{3} \right\rceil = 2$.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 59-70
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Using the historical data of 1,096 daily outpatient visit in a third-level hospital as an experimental sample, the data definition was first pre-processed using the feature definition and holiday effect time range determination algorithms proposed in sections 2.4 and 2.5, and then GP was used to perform learning of predictive expressions of daily outpatient visit. The parameters of GP are as follows: the population size is 200, the maximum depth is 10, the crossover probability is 0.8, the mutation probability is 0.1, and the number of iterations is 500. The function symbol set is $\{+,-,*,/,\sin,\cos,\tan,\log_{10},\exp,\ln,\mathrm{sqrt}\}$ and the terminal symbol set is $\{X_0,X_1,X_2,X_3,X_4,X_5,X_6,X_7\} \cup \{1,2,3,5,7,11,13,100\}$. The meaning of each variable is: X0 represents the year; X1 represents the month; X2 represents the solar term; X3 represents the weekday of the solar term; X4 represents the day of the week; X5 represents the type of holiday; X6 represents the day of the holiday; X7 represents the time range of holiday effect. {1,2,3,5,7,11,13,100} is a constant set, which is mainly composed of prime numbers 1-13. Such a constant set is helpful for mining constant terms in prediction expressions [28].

In order to avoid problems such as overfitting and underfitting in the learning process, experiments are performed in the following way: if the number of samples in a certain interval to be learned is greater than 50, 75% of the data is used as the learning set and 25% of the data is used as the test set. Otherwise, the "leave-one-out" method is adopted, that is, assuming that there are k samples, only one sample is left as the test set, and the other samples are used as the learning set, learning a total of k times, and the prediction expression with the highest fitness is selected as the final result.

4.1 Experimental Results

TABLE III. The experimental results of forecasting daily outpatient visits

Function segmentation condition	Daily outpatient visit prediction expression	Evaluation index	
		R^2	RMSE
New year's day: $X_5 = 1$ and $X_7 \neq 0$	$e^{\sqrt{sin\left(e^{(X_6-1)}\right)}*(2.4255+e^{cos(X_0)}*X3)} + X_7 * (tan(X_0 - 7) - 100) + \frac{X_0}{(cos(X_5 - X_1))^2}$	0.9293	672.5944
Spring festival: $X_5 = 2$ and $X_7 \neq 0$	$X_0 * e^{cos(X_5)} + \frac{7}{X_5 + 1 + X_7 - \frac{7}{X_0}} - 0.1425 + (\frac{X_6 - 1}{5 - X_1} +200 - cos(X_4) * 100) * (X_6 - 1 + X_7)$	0.9284	575.1872
Qingming festiva:	$X_5 - e^{X_4} + e^{cos(X_5)} * X_0 + X_0 - 3.3805$	0.9385	652.2494

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 59-70
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

$X_5 = 3$ and $X_7 \neq 0$	$+(tan((X_3 + (X_1 - X_7)) * X_6 + X_3) * (100 * X_5)$		
Labor day: $X_5 = 4$ and $X_7 \neq 0$	$X_0 * (2 + (1 - (X5 * 0.5403))) + 7 * tan(sin(X_4 - 3.3805) - (X0/0.5744)) + e^{(X_2-3)+X_3-1}$	0.9088	774.0942
Dragon boat festival: $X_5 = 5$ and $X_7 \neq 0$	$\left(2 + \sqrt{tan(X_7^2) + sin(X_5)}\right) * X_0 + tan(\tan(\cos(X_4) - 8)^2))) + e^{tan(\sqrt{X_4})} * tan(X_5) + \frac{X_2}{X_4} * 174.6429$	0.9081	785.9003
Mid-autumn festival: $X_5 = 6$ and $X_7 \neq 0$	$X_0 * X_7 + X_6 - 1 + sin(X_6 - 1 + sin(tan(X_6 - 1.291)) * X_0)) * X_0 + X_0 * e^{X_7}$	0.9871	268.8304
National day: $X_5 = 7$ And $X_7 \neq 0$	$X_0 * tan\left(e^{(X_5+3)*2)}\right) + (X_2 - 1) + X_1^2 - \left(\left(\frac{5}{tan\left(\frac{2}{X_4}\right)}\right) * X_7 + 5\right)^2 * e^{5+e^{(X_1*(X_3-1)/100}} + X_6$	0.9117	707.5915
Non-holiday: $X_5 = 0$ and $X_7 = 0$	$2 * (X_2 - 1) - (X_3 - 1) - (cos(X_4) + X_4)^4 + \frac{X_0}{X_4} + (2 * (X_2 - 1) - (7 * ((X_4 * 0.7867)^2 + tan(X_0 + cos(X_0) - 100)) - X_0))) * 3$	0.9381	498.6147

The specific experimental results are shown in Table 3. All the holidays in the "function segmentation condition" column in the table indicate a certain holiday and the interval with the holiday effect of the holiday.

From the experimental results, it can be known that R2 is above 0.9, which indicates that the daily outpatient prediction expression can explain more than 90% of the factors in the outpatient visit prediction, and indicates that the daily outpatient prediction method in this paper has good prediction ability. Of which, non-holiday (ie: X5 = 0 and X7 = 0) prediction accuracy of the daily outpatient visit on the learning set is: R2 = 0.9381, RMSE = 498.6147; the prediction accuracy on the test set is: R2 = 0.98163, RMSE = 463.0586. The prediction results of the test set are shown in Figure 3. It can be observed from the figure that the prediction curve of the daily outpatient visit for non-holidays and the actual value curve basically match. The results indicate that the prediction algorithm has high generalization ability.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 59-70
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

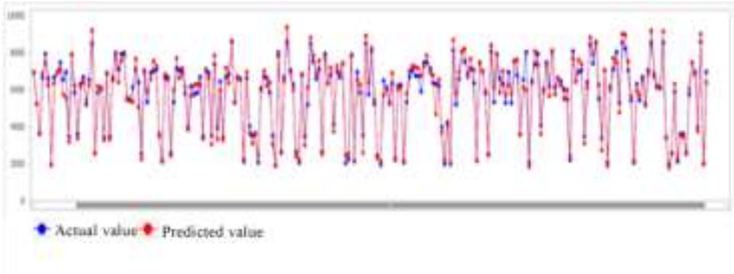

Fig 3 Comparisons between predicted values and actual values

V. CONCLUSIONS

Aiming at the problem of teachers and students outpatient visit analysis and prediction, firstly analyze the insufficient prediction results of daily outpatient visit and the SARIMA model, and distance-based holiday effect mining algorithm of daily outpatient visit was used to determine the holiday effect. In order to more accurately judge the impact of climate change on the outpatient visit, solar terms are used as the smallest unit to indicate climate change.

REFERENCES

[1] Hu JJ (2016) Improvement of outpatient service and medical experience. Chinese Hospital 4:160

[2] Tandberg D, Qualls C (1994) Time series forecasts of emergency department patient volume length of stay, and acuity. Annals of Emergency Medicine 23:299-306

[3] Li XS, Ma CL, Lei KH, Liu H (2013) Applications of SARIMA model on predicting outpatients quantity. Chinese Medical Record 3:37-40

[4] Tao Yuan, Gao YF, Liu L (2017) Applications of SARIMA model on forecasting of outpatients number.Chinese Journal of Hospital Statistics 24:391-393

[5] Luo L, Luo L, Zhang X, He XL (2017) Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. Bmc Health Services Research 17:469

[6] Garg B, Beg M M S, Ansari A Q (2012) A new computational fuzzy time series model to forecast number of outpatient visits//Fuzzy Information Processing Society

[7] Cheng C H, Wang J W, Li C H (2008) Forecasting the number of outpatient visits using a new fuzzy time series based on weighted-transitional matrix. Expert Systems with Applications 34:2568-2575

[8] Jiang YL, HU P (2009) The Application of Markov Prediction Model in Prediction of Outpatient Visits. Contemporary Medicine 15:28-29

[9] Famg F, Zhang BX, Li H, Yao MJ (2017) Application of least square method and season exponent in hospital management. Chinese Journal of Hospital Statistics 24:148-150

[10] Guo YP, Guan HJ, Rong SZ, Cui XY, Li MJ, Li XX (2013) Achievement of an application of GARCH model in forcasting hospital outpatient number by SAS software. Chinese Medical Record 14:42-45

[11] Zhang XM (2013) Prediction of hospital outpatient number based on GM(1, 1). Chinese Journal of Hospital Statistics 20:180-182

[12] Ma CL, Liu HX, Li XS, Lei HK, Zhou H (2012) Application of Grey Model (1, 1) on predicting outpatients quantity. Chinese Medical Record 12:23-25

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 59-70
Article History: Received: 30 March 2022, Revised: 8 April 2022, Accepted: 15 April 2022, Publication: 30 April 2022

[13] Yang XH, Zhong NY (2016) Forecasting of hospital oupatient based in deep brief network. Computer Science 43:26-30
[14] Ye MQ, Hu XG (2005) Curve Fitting and Forecasting of Outpatient Amount Based on Seasonal Neural Network. Journal of Engineering Graphics 12:83-86
[15] Hadavandi E, Shavandi H, Ghanbari A, Abbasian-Naghneh S (2012) Developing a hybrid artificial intelligence model for outpatient visits forecasting in hospitals. Applied Soft Computing 12:700-711
[16] Jiang S, Chin KS, Wang L, Qu G, Tsui KL (2017) Modified Genetic Algorithm-based Feature Selection Combined with Pre-trained Deep Neural Network for Demand Forecasting in Outpatient Department. Expert Systems with Applications 82:216-230
[17] Wang Y, Gu J (2014) Hybridization of Support Vector Regression and Firefly Algorithm for Diarrhoeal Outpatient Visits Forecasting// IEEE International Conference on Tools with Artificial Intelligence
[18] Yang CB (2009) An Improved Combination Forecasting Model Based on Grey Model and Artificial Neural Network and Its Application. Shandong Teachers' University
[19] Zhang YL, Yang ZS (2010) Grey RBF neural network based forecasting of outpatient capacity in modern hospital. Computer Engineering and Applications 46:225-228
[20] Zhang SS, Shang LL (2013) Season analysis and forecasting research of outpatient visits in traditional Chinese medicine hospital. Journal of Mathematical Medicine 46:225-228
[21] Zhu SZ, Wang DH, He YA, Wang Y (2015) Hospital outpatient visit analysis and forecasting using time series models. The Journal of University of Science and Technology of China 45:795-803
[22] Liang GL, Liu Y, Deng SM (2006) Applications of ARIMA model on predictive workload of out-patient department. Chinese Journal of Hospital Statistics 13:24-26
[23] Yang XF (2018) Causative analysis of sudden accidents of patients in outpatient during visiting doctors and its precaution. System Medicine 2018:7-8
[24] Hao Y, He J (2018) Analysis of solar terms characteristics of 10 common diseases occurrence in Beijing area. China Journal of Traditional Chinese Medicine and Pharmacy 2018:7-8
[25] Yao AR, Yao L, Ju SG, Chen WH, Ma HD (2008) Survey of outlier mining. Computer Science 35:13-18
[26] Koza J R (1992) Genetic programming: on the programming of computers by means of natural selection.
[27] Huang XD, Tang CJ, Li Z, Pu DH, Zeng LM (2003) A Gene Expression Programming Based Function Discovery Method//National Academic Conference on Databas.
[28] Zuo J (2004) Research on core technology of Gene Expression Programming. Sichuan University.