Iran J Public Health, Vol. 51, No.6, Jun 2022, pp.1313-1322

Original Article

Forecasting of Daily Outpatient Visits Based on Genetic Programming

*Xiaobing Liu [1], *Fulai Gu [1,2], Zhaoyang Bai [1], Qiyang Huang [3], Ge Ma [3]*

1. *School of Economics and Management, Dalian University of Technology, Liaoning, Dalian, China*
2. *The First Affiliated Hospital of Dalian Medical University, Liaoning, Dalian, China*
3. *China Academy of Industrial Internet, Beijing, China*

***Corresponding Author:** Email: gfl128@126.com

(Received 12 Mar 2021; accepted 15 May 2021)

Abstract

Background: The forecasting of daily outpatient visits has significant practical implications in outpatient clinic operation management, not only contributing to guiding long-term resource planning and scheduling but also making tactical resolutions for short-term adjustments on special days such as holidays. We here in propose an effective genetic programming (GP)-based forecasting model to predict daily outpatient visits (OV) in a primary hospital.

Methods: In the GP-based model, the holiday-based distance outlier mining algorithm was used to determine the holiday effect. In addition, solar terms were applied as the smallest unit to more accurately determine the impact of a change in the climate on the outpatient volume. A segmental learning strategy also was used to predict the daily outpatient volume for the time series data.

Results: The GP-based prediction could more effectively extract depth information from a finite training sample size and achieve a better performance for predicting daily outpatient visits, with lower root mean square error (RMSE) and higher coefficient of determination (R^2) values, than the seasonal autoregressive integrated moving average (SARIMA) model in the time range of holidays and the holiday effect.

Conclusion: GP-based model can achieve better prediction performance by overcoming the shortcomings of the SARIMA model. The results can be applied to support decision-making and planning of outpatient clinic resources, to help managers implement periodic scheduling of available resources on the basis of periodic features, and to perform proactive scheduling of additional resources.

Keywords: Daily outpatient visits; Forecasting; Time series data; Genetic programming; Outlier analysis

Introduction

The accurate forecasting of outpatient visits (OV) plays a key role in the planning and allocation of healthcare resources. Currently, outpatient clinics use formal and informal systems to forecast patient demands and acuity levels, thereby allocating clinical resources efficiently and effectively for better service. Tandberg and Qualls (1) first developed a time series model as the benchmark to predict demands for medical services. Since then, several forecasting applications have been developed to predict OV mainly using three different models: traditional statistical models, artificial intelligence models, and hybrid models.

Statistical prediction methods such as the autoregressive integrated moving average (ARIMA) model have been widely used to forecast OV (2-8). Furthermore, ARIMA models have been extended and modified to form the seasonal ARIMA (SARIMA) model and the multiplicative SARIMA model (3-5) to capture cycles over a weekly period in daily time series data for better prediction of monthly OV. However, the relationship between the target variables and factors in many time series data is nonlinear and complex, resulting in a poor prediction for nonlinear data. Artificial intelligence (AI) models such as the neural network and the genetic algorithm have been applied to predict OV (9-15). AI strategies based on data-driven and nonparametric models provide a higher prediction accuracy with some limitations including unknown or difficult-to-describe underlying relationships among datasets. Recently, several hybrid methods that integrate statistical prediction with AI have been proposed as a new platform for the prediction of OV (16-20). The statistical, AI, and hybrid prediction methods have bottlenecks that hinder improving the accuracy of predicting the number of patient visits. First, the current prediction models mainly use the quarter or month as the unit to predict the OV. In fact, the daily OV is more instructive to the daily allocation of clinical resources. Second, the prediction of OV during holidays has barely been included in the prediction models. Finally, the holiday effect has not been considered yet in the prediction of the number of OV, where the holiday effect is defined as an abnormal change in the number of OV during a period before or after holidays.

To address the shortcomings of the current prediction models, we herein propose a genetic programing (GP)-based model to predict the number of daily OV. In this model, the solar term is used as the minimum time unit to represent the change in the climate, and the distance-based outlier mining algorithm is used to determine the effective time range of the holiday effect.

Materials and Methods

Data preprocessing and determination of the holiday effect

Data description

The historical data with a total of 1096 daily OV from Jan 1, 2015 to Dec 31, 2017 at the top three hospitals in Dalian City, China, were collected (Fig. 1). The number of overall daily OV increased every year, with a large seasonal fluctuation. The number of daily OV ranged from 619 to 9639, with an average of 5500 (σ^2 = 4,653,076).

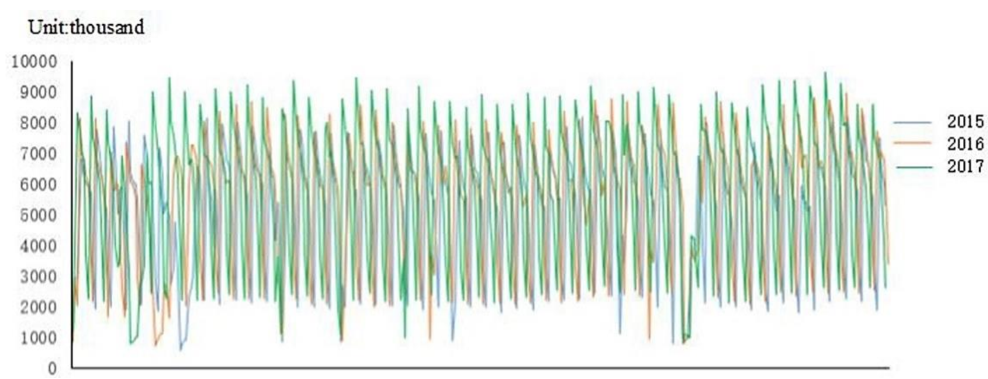

Fig. 1: The number of daily OV from January 1, 2015 to December 31, 2017 in a primary hospital located in Dalian City, China

Data preparation

The data attributes of daily outpatient volume in clinics were set as follows:

{Year, season, month, day, weekday, festival, festIndex, outpaNo}

Where year={2015, 2016, 2017}; season={1, 2, 3, 4}, "1" for the first quarter, "2" for the second quarter, and so forth; month ={1, 2, …, 12}, "1" for Jan, "2" for Feb, and so forth; day={2015/01/01, …, 2017/12/31}, presenting a specific date, "2015/01/01" for Jan 1, 2015; weekday={1, 2, …, 7}, presenting the day of the week , "1" for Monday, "2" for Tuesday, and so forth; festival = {0, 1, 2, 3, 4, 5, 6, 7}, representing holiday types, "0" for nonfestival, "1" for New Year's Day, "2" for Chinese Spring Festival, "3" for Qingming Festival, "4" for Labor Day, "5" for Dragon Boat Festival, "6" for Mid-Autumn Festival, and "7" for National Day; festIndex ={1, 2, …, n}, representing the days of the festival, "1" for the first day of the festival, "2" for the second day of the festival, and so forth; and outpaNo represents the number of OV for a specific day. Table 1 lists the daily OV within a period before and after the 2016 Chinese Spring Festival, showing a great decrease of daily outpatient volume during the Spring Festival.

Table 1: The number of daily OV with the SARIMA model and the repredicted number of daily OV with the adjusted model during the period around the 2016 Chinese Spring Festival

Daily OV								***Repredicted daily OV with the adjusted model***							
Day	Year	Season	Month	Weekday	Festival	Fest Index	Outpano	Solar Term	Month	Week Index	Weekday	Festival	Fest Index	Fest Effitive	Outpano
2016/2/1	2016	1	2	1	0	-	6638	1	2	2	1	0	-	-	6638
2016/2/2	2016	1	2	2	0	-	6106	1	2	2	2	0	-	-	6106
2016/2/3	2016	1	2	3	0	-	5483	1	2	2	3	0	-	-	5483
2016/2/4	2016	1	2	4	0	-	4945	2	2	1	4	0	-	-	4945
2016/2/5	2016	1	2	5	0	-	4211	2	2	1	5	2	-	-2	4211
2016/2/6	2016	1	2	6	0	-	2708	2	2	1	6	2	-	-1	*2708*
2016/2/7	2016	1	2	7	2	1	*747*	2	2	1	7	2	1	-	*747*
2016/2/8	2016	1	2	1	2	2	*963*	2	2	2	1	2	2	-	*963*
2016/2/9	2016	1	2	2	2	3	*1101*	2	2	2	2	2	3	-	*1101*
2016/2/10	2016	1	2	3	2	4	*1136*	2	2	2	3	2	4	-	*1136*
2016/2/11	2016	1	2	4	2	5	*2670*	2	2	2	4	2	5	-	*2670*
2016/2/12	2016	1	2	5	2	6	*2761*	2	2	2	5	2	6	-	*2761*
2016/2/13	2016	1	2	6	2	7	*1664*	2	2	2	6	2	7	-	*1664*
2016/2/14	2016	1	2	7	0	-	3427	2	2	2	7	2	-	1	*3427*
2016/2/15	2016	1	2	1	0	-	6402	2	2	3	1	2	-	2	*6402*
2016/2/16	2016	1	2	2	0	-	6792	2	2	3	2	2	-	3	*6792*
2016/2/17	2016	1	2	3	0	-	6921	2	2	3	3	2	-	4	6921
2016/2/18	2016	1	2	4	0	-	6539	2	2	3	4	0	-	-	6539
2016/2/19	2016	1	2	5	0	-	6114	3	2	1	5	0	-	-	6114

Evaluation indicators

The root mean square error (RMSE) and coefficient of determination (R^2) values were used as the forecasting evaluation indicators and are defined as follows:

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

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

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

Adjustment of data attributes

The SARIMA model was used for the preliminary prediction, and the evaluation indicators R^2 and RMSE were 0.773 and 1029.418, respectively. The numbers of predicted daily OV for all three years and in February 2016 are shown in Fig. 2 and Fig. 3, respectively. On Feb 7, 2016, the date of the Chinese Spring Festival, the actual daily outpatient volume was 747; however, the predicted volume with SARIMA was 4532. The great difference between the predicted and actual daily number of OV mainly resulted from large fluctuations in the special holiday effect. Compared with normal working days, holidays including the Chinese Spring Festival and National Day have significant effects on the numbers of daily outpatient visits. Furthermore, the numbers of daily OV were significantly associated with the climatic conditions where the outpatient clinics are located (21). Two time series attributes (season and month) cannot really reflect a change in the climate. The climate conditions on the same day among different years might be significantly different.

Medical studies have demonstrated that the "24 solar terms" have a significant effect on disease (22). The solar terms fully consider the variation of natural phenomena and reflect the changes in climate more precisely than the season and month alone.

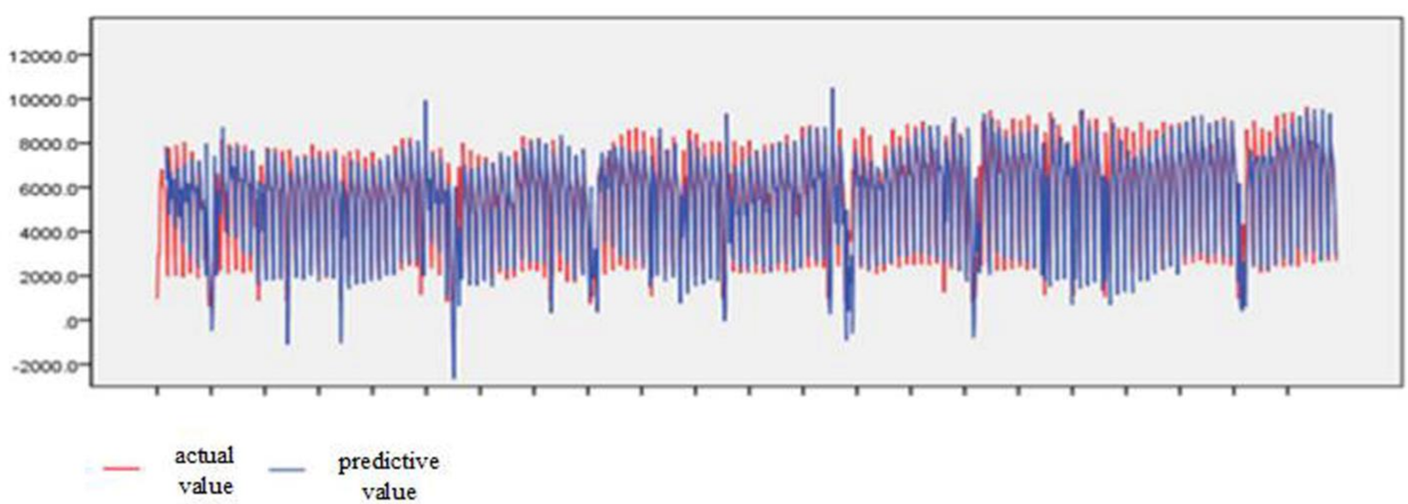

Fig. 2: The number of predicted daily OV with SARIMA from Jan 1, 2015 to Dec 31, 2017

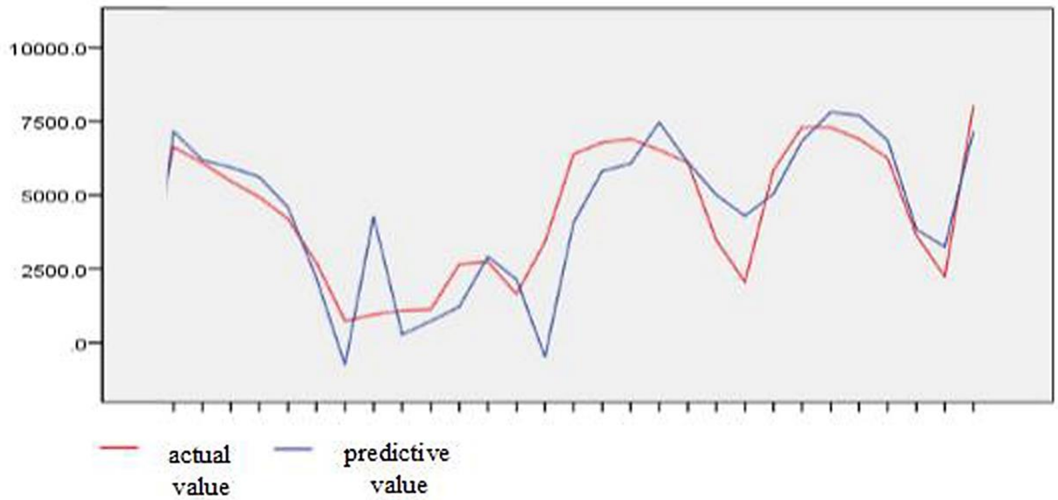

Fig. 3: The number of predicted daily OV with SARIMA in February 2016

Based on the prediction results of the SARIMA model and the effects of the solar terms on disease, the data attributes of the daily OV were adjusted as follows:

{Year, month, solarTerm, weekIndex, weekday, festival, festIndex, festEffitive, outpaNo}

Where solarTerm = {1, 2, …, 24}, representing the 24 solar terms, "1" for Xiaohan, "2" for Dahan, and the others following the order of the 24 solar terms; weekIndex = {1, 2, 3}, representing the weeks within the current solar term, "1" for the first week corresponding to solar terms, "2" for the second week, and so forth. The maximum weekIndex is 3 because each solar term has at most 16 days. $festEffitive = \{-n_b, .., -1, 1, \ldots, n_a\}$, representing the time range of the holiday effect, "-1" for the first day before the holiday with a holiday effect, "1" for the first day after the holiday with a holiday effect, and so forth. The holiday effect variables n_b and n_a have the following characteristics: 1) the n_b day before the holiday and the n_a day after the holiday must have a holiday effect; 2) the n_b + 1 day before the holiday and the n_a + 1 day after the holiday must have no holiday effect; and 3) n_b may not be equal to n_a for the same holiday.

The numbers of daily OV during the 2016 Chinese Spring Festival were estimated using the adjusted data attributes $(n_b = 2n_a = 4)$ (Table 1). Any day of the year can be uniquely determined by the values of the attributes (year, solarTerm, weekIndex, and weekday).

Determination of the holiday effect on the outpatient volume

The time range of the holiday effect, namely n_b and n_a, was determined to estimate the holiday effect on the daily outpatient volume: if the number of daily OV for a specific day before or after a holiday changed abnormally, the number of daily OV for that day was defined as an outlier relative to the daily outpatient volume set within a certain time range. A distance-based outlier mining algorithm ($\boldsymbol{l}^{\mathbf{th}}$_Day_Check) was developed to determine n_b and n_a with two steps.

Step 1 determines whether the daily outpatient volume on the lth day has a holiday effect according to the characteristics of the daily outpatient volume $outpaNo_x^l$ on the lth day before or after the holiday x as follows:

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

or

$$outpaNo_x^l \leq v_l = Mean\_outpaNo_{db(x,l,k)} - \lambda(max_{db(x,l,k)} - min_{db(x,l,k)})$$ [1]

Where $kb(x,l,k) = \{outpaNo_x^l, outpaNo_{x,1}, outpaNo_{x,2}, \dots, outpaNo_{x,k}\}$ and $outpaNo_x^l.weekday = outpaNo_{x,i}.weekday$ represents a subdataset composed of *k* + 1 daily OV including $outpaNo_x^l$ within a certain period around holiday *x*; $outpaNo_x^l$ is the normal daily outpatient volume without a holiday effect, $i = 1,2,\dots,k$. $Mean\_outpaNo_{db(x,l,k)} = \frac{\sum_{i\in db(x,l,k)} outpaNo_i}{|k+1|}$. $\lambda \in [0.6,1]$ is the adjustment coefficient. $max_{db(x,l,k)}$ and $min_{db(x,l,k)}$ are the maximum and minimum daily outpatient volumes in the dataset $db(x,l,k)$, respectively.

Step 2 determines whether $outpaNo_x^l$ is an outlier. If the distance between $outpaNo_x^l$ and at least p × 100% of the data points in the dataset $db(x,l,k)$ is greater than r, $outpaNo_x^l$ is defined as an outlier. The parameters *p* and *r* are defined as follows:

$$p = \frac{k}{k+1} \qquad [2]$$

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

[3]

The outpatient volume data during the 2016 Spring Festival shown in Table 1 were used as an example to determine if the first day (Feb 14, 2016, Sun) after the Spring Festival has a holiday effect. Given k = 8 and λ = 0.6, the actual outpatient volume in the dataset $db(7,1,9)$ is {3427, 2158, 1725, 1687, 2053, 2074, 2250, 2469, 2277}, where the first data point (3427) is the outpatient volume on the first day after the Spring Festival; and the other numbers are the daily OV in the four normal Sundays before (4 Sundays) and after (4 Sundays) Feb 14, 2016; $Mean\_outpaNo = 2236.56$; $max_{db(7,1,9)} = 3427$; and $min_{db(7,1,9)} = 1687$. According to equation (1), $v_u =$ 3280.56, which is less than 3427. Therefore, the day (Feb 14, 2016) has a holiday effect. According to formula [2] and formula [3], *P* = 0.89 and *r* = 98.56. The distances between the daily outpatient volume on February 14, 2016 and the daily outpatient volume of other days excluding Feb 14, 2016 in the dataset $db(7,1,9)$ are greater than *r*, indicating that the number of daily OV on Feb 14, 2016 is the outlier. Therefore, the first day after the 2016 Spring Festival has a holiday effect.

The complete steps for determination of the time range of the holiday effect of the holiday *x* are as follows:

Step 1:$sum\_n_b = sum\_n_a = 0$

Step 2: for each $y \in year$

Step 2.1: determine the parameters $n_{b,y}$ and $n_{a,y}$ of the holiday effect for the holiday in the year using the $\boldsymbol{l}^{\mathbf{th}}$_Day_Check algorithm.

Step 2.2:$+= n_{b,y} sum\_n_a += n_{a,y}$

Step 3: end for

Step 4: export $n_{b,y}$ and $n_{a,y}$: $n_b = \lceil\frac{sum\_n_b}{|year|}\rceil$,$n_a = \lceil\frac{sum\_n_a}{|year|}\rceil$, where $\lceil a\rceil$ represents the smallest integer greater than or equal to *a*.

The outpatient volume data during the 2016 Spring Festival shown in Table 1 were used as an example for the $\boldsymbol{l}^{\mathbf{th}}$_Day_Check algorithm: $n_{a,2015} = 4$, $n_{a,2016} = 3 n_{a,2017} = 4$, $n_{b,2015} = 1 n_{b,2016} = 2$, and $n_{b,2017} = 1$. According to $\boldsymbol{l}^{\mathbf{th}}$_Day_Check, $n_a = \left\lceil\frac{4+3+4}{3}\right\rceil = 4$, and $n_b = \left\lceil\frac{1+2+1}{3}\right\rceil = 2$.

GP-based daily OV forecasting algorithm

GP-based daily outpatient volume prediction algorithm review

As an artificial intelligence algorithm (23), GP uses a hierarchical structure with variable space to represent the solution space. We proposed a GP-based daily outpatient volume prediction model algorithm by integrating function symbol sets and terminal symbol sets to randomly generate the predicted daily outpatient volume (initial population) for learning operations such as selection, crossover, and mutation.

Coding

The gene value of the individual code is derived from the function symbol set composed of mathematical operation symbols consisting of the input variables and constants of the problem. The individual codes are organized in a tree structure to generate the initial group with the "growth method." The steps are as follows: the maximum depth of a tree (i.e., the number of node layers) is predetermined, and then an element is randomly selected from the function symbol set as the root node to randomly generate the same number of subtrees as the number of operands of the function corresponding to the element. If the depth of a node is less than the given maximum depth, an element is randomly selected from the function symbol set or terminal symbol set to fill the node. If the depth is equal to the maximum depth, an element is selected from the terminal symbol set to fill the node.

Fitness evaluation and selection

The tournament selection strategy is a fitness-based selection scheme used to select individuals that participate in crossover and variation as follows. A set of *m* individuals is randomly selected from the parental population that is known as the tournament set (*m*), from which the individual with the best fitness is selected. In addition, the elitist strategy is used to save the individuals with the best-fitness in each generation.

Crossover

Crossover is the main operation for generating new individuals that inherit the parental genes and generate new genes. Two individuals (P_1,P_2) are randomly selected from the population as the parents, and a point is then randomly selected from each parent as the crossover point. Two parents exchange the crossover point as the root node of a subtree to generate new individuals (O_1,O_2).

Mutation

Mutation increases the diversity of the population, which helps the algorithm to jump out of the local optimum. An individual from the population is randomly selected as the parent, then a node from the parent is randomly selected as the point. The subtree rooted in the mutation point is deleted, and a new subtree is randomly generated and inserted into the mutation point.

Segment learning

We proposed a segmented learning strategy to realize the segmental learning of the daily outpatient forecast function. A fitness threshold is preset for the GP algorithm to learn the prediction functions. If the fitness of the GP-generated prediction functions is lower than the threshold, an attribute is randomly selected to establish a segmental boundary in which the prediction function is generated to perform prediction learning until the fitness reaches the threshold.

The segmental boundary of the piecewise functions is often determined by the dichotomy (24). If we assume that the feature to be segmented is R with an interval of $[t_i, t_{i+1}]$, the midpoint of R ($t_{mid} = \frac{t_i+t_{i+1}}{2}$) is used as a segmental point for learning between $R_1 = [t_i, t_{mid}]$ and $R_2 = [t_{mid}, t_{i+1}]$. If the fitness of the prediction function reaches the threshold, $\mathrm{t_{mid}}$ is determined as is determined as the segmental point. If the fitness of the prediction function is lower than the threshold, the segmental boundary with dichotomy is repeated in the interval below the fitness threshold, until it reaches the threshold.

Fluctuations in the number of daily OV are mainly caused by holidays. Therefore, the festival feature is prioritized for segmental learning with the dichotomy method.

Results

Totally, 1096 historical data of daily outpatient volume from a primary hospital were used for the experimental sample to validate the forecasting model. The data were first preprocessed by definition of the features and algorithm for the time range of the holiday effect, and then GP-based prediction function learning was performed to determine the daily outpatient volume. The GP parameters were defined as follows: population size, 200; maximum depth, 10; crossover probability, 0.8; mutation probability, 0.1; and number of iterations, 500. The function symbol set was $\{+, -, *, /$

$, sin, cos, tan, log_{10}, exp, ln, sqrt\}$, and the terminal symbol set was $\{X_0, X_1, X_2, X_3, X_4, X_5, X_6, X_7\} \cup \{1,2,3,5,7,11,13,100\}$, where X_0 represents the year; X_1 for represents the month; X_2 represents the solar term; X_3 represents the week of the solar term; X_4 represents the day of the week; $X5_1$ represents the type of holiday; X_6 represents the day of a certain holiday; and X_7 represents the time range of the holiday effect. The constant set {1, 2, 3, 5, 7, 11, 13, 100} was mainly composed of the prime numbers of 1–13 and helped to mine the constant terms in the prediction functions (24).

In order to avoid the issues of over-fitting and under-fitting during the learning process, the validation was performed as follows: if the number of samples in a certain learning interval is greater than 50, 75% of the samples are used as the learning set and 25% of samples are used as the testing set. Otherwise, the "leave one method" is adopted, meaning among *k* samples, only one sample is left at a time as the testing set, and the other samples are used as the learning set for learning *k* times in total. The prediction function with the highest fitness is selected as the final result. R^2 remains above 0.9 in all cases, indicating that the prediction function of the daily outpatient volume explains more than 90% of the factors in the outpatient volume prediction. Our forecasting model for daily OV was proven to have a high prediction accuracy. R^2 for the nonholiday daily outpatient volume (X_5 = 0 and X_7 = 0) was 0.9381 (RMSE = 498.6147) in the learning set and 0.98163 (RMSE = 463.0586) in the testing set. It is clear from Fig. 4 that the predicted values of the daily OV on nonholiday days are very close to the actual values .

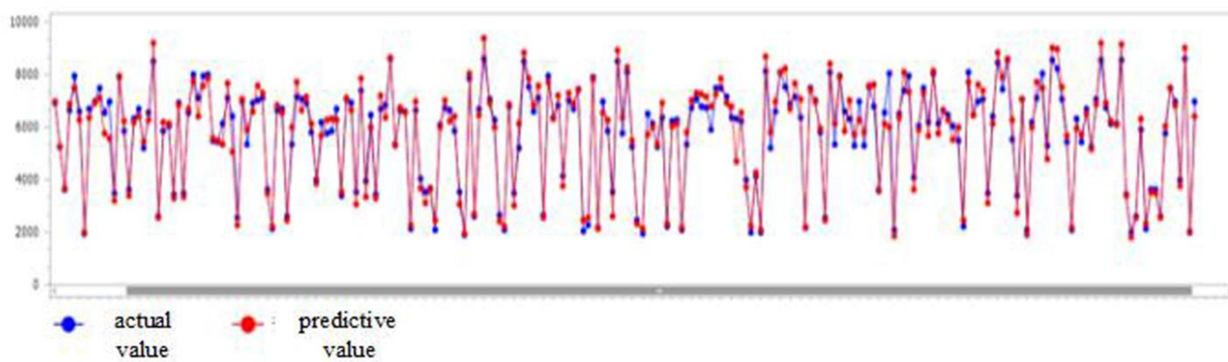

Fig. 4: Comparisons between the predicted daily OV with the GP-based model and the actual numbers of daily OV

Discussion

To the best of our knowledge, we demonstrated a GP-based prediction for the first time to more effectively extract depth information from a finite training sample size that achieves a better performance for predicting daily outpatient visits, with lower RMSE and higher R^2 values, than the SARIMA model. Time series analysis for forecasting ED arrivals has been pronounced in studies with acceptable results (1-5).

As indicated by the preliminary results of SARIMA prediction, the number of daily OV greatly fluctuates during holidays. The ARIMA algorithm proposed in our study was able to accurately predict ED visits, with a margin of error of 1.001 and an RMSE of 1.55. Further, we developed a GP-based prediction model that has the highest prediction accuracy with respect to the prediction of daily patient arrivals. In the GP-based forecasting model, the holiday-based distance outlier mining algorithm was used to determine the holiday effect. The solar terms were applied as the smallest unit to more accurately determine the impact of a change in the climate on the outpatient volume. In addition, a segmental learning strategy was proposed to realize the prediction of daily outpatient volume. Based on error measures and visual interpretation of the results, we can infer that the prediction provides a useful assessment to forecast daily ED visits.

The proposed GP-based forecasting algorithm can predict daily OV with an R^2 value greater than 0.9, which is a significant improvement in the forecasting performance compared to the SARIMA model. The proposed model is designed to improve the forecasting accuracy by overcoming the shortcomings of the SARIMA model in the time range of holidays and the holiday effect. The results can be applied to support decision-making and planning of outpatient clinic resources, to help managers implement periodic scheduling of available resources on the basis of periodic features, and to provide proactive scheduling of additional resources.

Conclusion

We developed and validated a GP-based prediction model that achieves better prediction performance by overcoming the shortcomings of the SARIMA model in the time range of holidays and the holiday effect, with a lower residual variance and a lower mean of residual errors.

Journalism Ethics considerations

Ethical issues (Including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed by the authors.

Conflict of interest

The authors declare that they have no competing interests.

References

1. Tandberg D, Qualls C (1994). Time series forecasts of emergency department patient volume, length of stay, and acuity. *Ann Emerg Med,* 23:299-306.
2. Kadri F, Harrou F, Chaabane S, Tahon C (2014). Time Series Modelling and Forecasting of Emergency Department Overcrowding. *J Med Syst,* 38:107.
3. Ting Z, Li L, Xinli Z, Yingkang S, Wenwu S (2017). Time-Series Approaches for Forecasting the Number of Hospital Daily Discharged Inpatients. *IEEE J Biomed Health Inform,* 21:515-526.
4. Calegari R, Fogliatto FS, Lucini FR (2016). Forecasting Daily Volume and Acuity of Patients in the Emergency Department. *Comput Math Methods Med,* 2016:3863268.
5. Abraham G, Byrnes GB, Bain CA (2009). Short-term forecasting of emergency inpatient flow. *IEEE Trans Inf Technol Biomed,* 13:380-8.
6. Garg B, Beg MMS, Ansari AQ (2012) A new computational fuzzy time series model to forecast number of outpatient visits. (ed)^(eds) IEEE Annual Meeting of North American of Fuzzy Information Processing Society (NAFIPS),
7. Landram FG, Abdullat A, Shah V (2011). Using seasonal and cyclical components in least squares forecasting models. *Southwestern Economic Review.*
8. Zhang X (2013). Prediction of hospital outpatient number based on GM. *Chinese Journal of Hospital Statistics,* 20:180-182.
9. Hinton GE, Osindero S, Teh YW (2006). A fast learning algorithm for deep belief nets. *Neural Comput,* 18:1527-54.
10. Bengio Y, Lamblin P, Popovici D, Larochelle H (2007). Greedy layer-wise training of deep networks.
11. Yang X, Zhong N (2016). Forecasting of hospital outpatient based in deep brief network. *Computer Science,* 43:26-30.
12. Hadavandi E, Shavandi H, Ghanbari A (2012). Developing a hybrid artificial intelligence model for outpatient visits forecasting in hospitals. *Appl Soft Comput,* 12:700-711.
13. 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 Syst Appl,* 82:216-230.
14. Mtonga K, Kumaran S, Mikeka C, et al (2019). Machine Learning-Based Patient Load Prediction and IoT Integrated Intelligent

Patient Transfer Systems. *Future Internet*, 11:236.

15. Nelson A, Herron D, Rees G (2019). Predicting scheduled hospital attendance with artificial intelligence. *npj Digit Med*, 2:26.
16. Alaeddini A, Yang K, Reeves P, Reddy CK (2015). A hybrid prediction model for noshows and cancellations of outpatient appointments. *IIE Transactions on Healthcare Systems Engineering*, 5:14-32.
17. Klute B, Homb A, Chen W (2019). Predicting Outpatient Appointment Demand Using Machine Learning and Traditional Methods. *J Med Syst,* 43(9):288.
18. Zhang S (2013). Seasonal Analysis and Forecasting Research of Outpatient Visits in Traditional Chinese Medicine Hospital. *Journal of Mathematical Medicine*.
19. Zhu S, Wang D, Yanan HE, Yan W (2015). Hospital outpatient visit analysis and forecasting using time series models. *Journal of University of Science & Technology of China*.
20. Liang GL, Liu Y (2006). Applications of ARIMA model on predictive workload of out-patient department. *Chinese Journal of Hospital Statistics*.
21. Yang X (2018). Causative analysis of sudden accidents of patients in outpatient during visiting doctors and its precaution. *System Medicine,* 7:7-8.
22. Koza, JohnR (1992). *Genetic programming : on the programming of computers by means of natural selection*. ed. MIT Press.
23. Huang X, Tang C, Li Z (2003). A Gene Expression Programming Based Function Discovery Method. *National Academic Conference on Database*.
24. Zuo J (2004). Research on core technology of Gene Expression Programming. *Sichuan University*.