Forecasting Daily Pedestrian Flows in the Tiananmen Square Based on Historical Data and Weather Conditions

Lida Huang
Institute of Public Safety Research,
Tsinghua University, Beijing, China
hld14@mails.tsinghua.edu.cn

Tao Chen*
Institute of Public Safety Research,
Tsinghua University, Beijing, China
chentao.a@tsinghua.edu.cn

Yan Wang
Institute of Public Safety Research,
Tsinghua University, Beijing, China
wangyan14@mails.tsinghua.edu.cn

Hongyong Yuan
Institute of Public Safety Research,
Tsinghua University, Beijing, China
hy-yuan@tsinghua.edu.cn

ABSTRACT

It is important to forecast the pedestrian flows for organizing crowd activities and making risk assessments. In this article, the daily pedestrian flows in the Tiananmen Square are forecasted based on historical data, the distribution of holidays and weather conditions including rain, wind, temperature, relative humidity, and AQI (Air Quality Index). Three different methods have been discussed and the Support Vector Regression based on the Adaptive Particle Swarm Optimization (APSO-SVR) has been proved the most reliable and accurate model to forecast the daily pedestrian flows. The results of this paper can help to conduct security pre-warning system and enhance emergency preparedness and management for crowd activities.

Keywords

Pedestrian flows, forecasting, historical data, weather conditions, APSO-SVR.

INTRODUCTION

With the rapid economic development, more and more crowd activities have been conducted in recent years. The crowd safety issues, meanwhile, are more and more concerned. Gatherings of large human crowds quite often end in crowd disasters such as the recent catastrophe in Shanghai Bund at 31th Dec, 2014, where 36 people died. According to incomplete statistics, thousands of people were killed by stampede in highly dense crowd around the world (Schadschneider, Klingsch, Klüpfel, Kretz, Rogsch and Seyfried, 2009). To avoid these crowd disasters, a reliable forecast of the pedestrian flow is needed and plays a major role in the emergency planning.

Accurate forecasts build a sound foundation for better emergency planning and administration. This calls for more efficient and accurate forecasting techniques. Traditional studies mainly focus on time series forecasting models, including the linear methods (Burger, C. J. S. C., Dohnal, Kathrada and Law, 2001) and the nonlinear methods (Law, 2000) , which are particularly useful when little knowledge is available or when there is no satisfactory explanatory model that relates the prediction variable to other explanatory variables (Chen, 2011) . Recent studies pay more attention to weather conditions, which indirectly influence the demand for outdoor activities. Rosselló-Nadal, J. et al. (2011) used transfer function models to investigate the effect of the short-term weather conditions on British outbound flows, and showed that the mean temperature, heat waves, air frost and sunshine days made important contribution. Singhal, A. et al. (2014) used the ordinary least square (OLS) regression model to analyze the impacts of weather on weekly and daily urban transit ridership. Chen et al. (2007) forecasted the tourism demand using SVR and the Back-Propagation Neural Networks (BPNN). Chen (2013) constructed a model combining Support Vector Regression (SVR) with Adaptive Particle Swarm Optimization (APSO) to forecast short-term tourism flow in Mount Huangshan based on the online booking information and the body comfort index (a meteorology index which is related to wind, temperature and relative humidity).

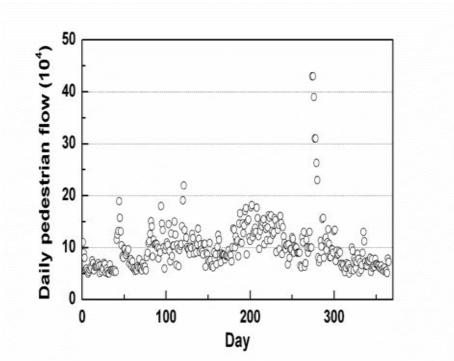

Figure 1. Time series for daily pedestrian flows.

In general, the state of art on pedestrian flows forecasting considers the effect of variability using high frequency database in order to incorporate short term flow change caused by weather anomalies. The predicting models and input variables are the two key factors of these predicting problems, which depend on specific application scenarios and need to be tested and validated.

This paper is committed to forecast the daily pedestrian flows in the Tiananmen Square based on historical data and weather conditions, which can reveal both the historical trends and uncertainties. It is organized as follows. The next section presents data used in this study and selects the relevant variables. Then three forecasting methods, the multivariable linear regression (MLR) model, BPNN and SVR model are briefly introduced and compared and the results are discussed. The last section presents the conclusions of this work.

DATA SET AND PRETREATMENT

The Tiananmen Square, located in the center of Beijing, is a landmark of China, which attracts thousands of travelers every day. This paper analyzes its daily pedestrian flows from 1st January, 2013 to 31st December, 2013. Figure 1 shows the time series for daily pedestrian flows of the Tiananmen Square. It can be seen that flows in the summer are larger than those of the winter. And there are some abnormal points, which respectively correspond to New Year's Day, Spring Festival, Tomb-sweeping Day, Labor Day and National Day. That means, statutory holidays have an influence on the daily pedestrian flow, and among these holidays, National Day is the most special for its striking contribution to the flows. Therefore, we use three 0-1 discrete variables to represent the weekend, the National Day and other holidays respectively.

Figure 2 shows the statistical histogram of daily pedestrian flows. The 50th percentile value is 80,000, which means that the daily pedestrian flows are below 80,000 in nearly half a year. And the 99th percentile value is 300,000. It means there is little possibility for the daily pedestrian flows to exceed 300,000. We can conclude that once the daily pedestrian flow is over 300,000, emergency managers should be on the alert and keep an eye on the real-time pedestrian flow.

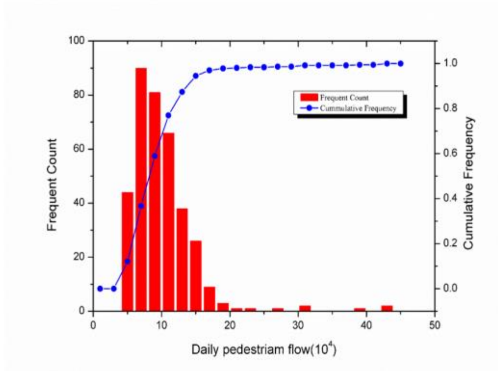

Figure 2. Statistical histogram of daily pedestrian flows.

The weather data of 2013 for Beijing were obtained from China Meteorological Data Sharing Service System (<http://cdc.cma.gov.cn>). These data was recorded at a weather station in Beijing and is assumed to be representative of weather conditions across the city. It recorded the average/ highest/lowest air temperature, rain precipitation, evaporation capacity, wind speeds, wind gusts, sunshine duration and so on. The weather data was cleaned and consolidated into semidiurnal or daily weather conditions. The Air Quality Index (AQI) data was obtained from Beijing Municipal Environmental protection Center

(<http://www.bjmemc.com.cn/g372.aspx>). These data was recorded at 28 monitoring stations in Beijing, and we use the average value to represent the AQI.

We used SPSS to select the significant predictors and 10 variables were elected, which are the average temperature, the highest temperature, the lowest temperature, rain precipitation, wind speed, relative humidity, AQI, Weekend, National Day and other holidays. In order to make a comprehensive utilization of historical data, data of the last three days before the predicted position was also concluded in the predictors. Namely, there are 13 variables considered to forecast the pedestrian flows. Table 1 lists all the predictors used in this case study.

X_1	The Average Temperature /℃	X_8	Weekend (0/1)
X_2	The Highest Temperature /℃	X_9	National Day (0/1)
X_3	The Lowest Temperature /℃	X_{10}	Other Holidays (0/1)
X_4	Rain Precipitation /um	X_{11}	Flow of Day t-1 /10^4
X_5	Wind Speed /$m \cdot s^{-1}$	X_{12}	Flow of Day t-2 /10^4
X_6	Relative Humidity /%	X_{13}	Flow of Day t-3 /10^4
X_7	AQI		

Table 1. The predictors considered in the case study.

METHODOLOGY

Three different forecasting models, MLR model, BPNN and SVR model are applied and discussed in the following. For purpose of testing the generalization performance of these models, the data of Tuesday, Thursday and Saturday was extracted as the testing set, and the remaining constituted the training set.

The most common used multivariable linear regression model is based on the ordinary least square. Given the above selected variables, the forecasting equation can be written as:

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_{13}X_{13}$$

where the meaning of X_1-X_{13} is shown in Table 1, and Y means the pedestrian flow of day t, β_0 means model intercept and β_1-β_{13} are beta parameters for estimation.

BPNN is a model-free forecasting technique widely used in the last decade. It has been demonstrated that BPNN is very effective in solving nonlinear predictive problems, which are usually hard to handle by traditional methods (Chen, 2011). The core of BPNN is the back-propagation gradient descent algorithm, which tries to improve the performance of the neural network by reducing the total error. The concrete implementation of BPNN is introduced in reference (Huang, Hwang and Hsieh. 2002). In this article, we used a 3-layer neural network to predict the pedestrian flow. The learning procedure of BPNN is listed as follows.

Step 1. Initialize the network structure and weights.

Step 2. Define input and output variables.

Step 3. Calculate the output values of hidden layer and output layer.

Step 4. Modify network weights to minimize the global error.

Step 5. Go to Step 3 until the error is small enough or iterations exceed the set point.

Recently, a regression version of SVM has emerged as an alternative and powerful technique to solve nonlinear regression problems, which is referred to as Support Vector Regression (SVR) later in the literature. The SVR transforms the input data into a high-dimensional feature space by nonlinear mapping, where the original problem is solved as a linear regression problem. Detailed descriptions of SVR can be found in reference (Vapnik, 2000) and reference (Scholkopf and Smola 2001). There are three input parameters in SVR model which will directly affect its learning capacity, the insensitive loss coefficient ε which is employed to stabilize the estimation, the penalty coefficient C which calculates the penalty when an error occurs, and the parameter γ in kernel function which reflects the correlation among support vectors. APSO was used to optimize the parameters (ε, C, γ) and detailed descriptions of it are shown in reference (Hornik, Stinchcombe and White, 1989). The calculation procedure of APSO-SVR is listed as follows.

Step 1. Normalize the input variables.

Step 2. Define (ε, C, γ) as the APSO parameters and initialize them.

Step 3. Set the number of particles, the maximum iterations and weights for fitness calculation. Define the mean square error of SVR model as the fitness function.

Step 4. Measure the fitness of each particle, update the individual extreme to obtain each particle's own best position. After comparing the best position of all particles, calculate the best locations of all particles.

Step 5. Adjust the particle's position and speed.

Step 6. Stop if the termination criterion is satisfied; otherwise, return to step 4.

Step 7. Obtain the optimized parameters (ε, C, γ), and input them into SVR model.

RESULTS AND DISCUSSION

The regression results of MLR model are presented in Table 2. In Table 3, the correlation coefficient (R), the mean absolute percentage error (MAPE) and the normalized mean square error (NMSE) of different forecasting models are compared.

Parameters	Estimated Value	Confidence Interval
β_0	2.9384	[0.6324, 5.2445]
β_1	0.2600	[-0.1464, 0.6663]
β_2	-0.0725	[-0.2939, 0.1490]
β_3	-0.1646	[-0.3937, 0.0644]
β_4	-0.0012	[-0.0051, 0.0026]
β_5	-0.0314	[-0.3916, 0.3288]
β_6	0.0154	[-0.0113, 0.0422]
β_7	-0.0061	[-0.0106, -0.0015]
β_8	0.8239	[0.1897, 1.4582]
β_9	6.7904	[3.5289, 10.0519]
β_{10}	2.2263	[0.7196, 3.7331]
β_{11}	0.6904	[0.5810, 0.7998]
β_{12}	-0.2211	[-0.3465, -0.0958]
β_{13}	0.0785	[-0.0422, 0.1992]

R^2=0.8587 F=74.3257 p<1.467e-60 s^2=3.0431

Table 2. The regression results of the MLR model.

	MLR	BPNN	APSO-SVR
R	0.9267	0.9695	0.9206
MAPE	0.1841	0.2291	0.1295
NMSE	0.4059	0.5114	0.1363

Table 3. Comparison of evaluation indexes of different forecasting models.

Figure 3 and Figure 4 show the comparison of true values and predicting values of daily pedestrian flows by different forecasting methods.

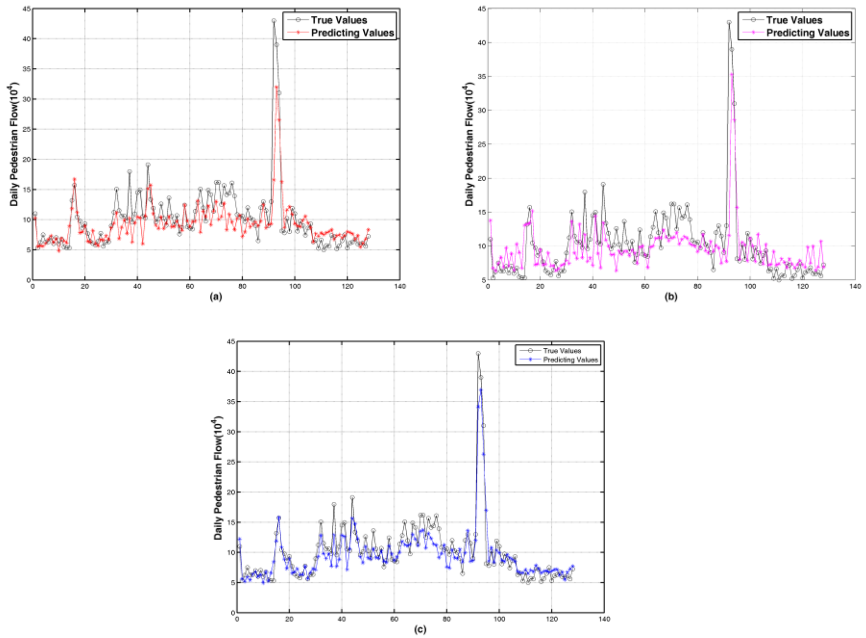

Figure 3. Pedestrian flow forecasting results on the testing set by different forecasting methods. (a) The MLR model. (b) The BPNN model. (c) The APSO-SVR model.

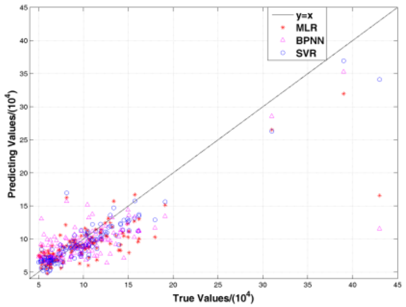

Figure 4. Comparison of true values and predicting values of daily pedestrian flows by different forecasting methods. The reference line is y = x.

From Table 2 we can find that zero is included in the confidence intervals of some parameters, which means these parameters are not significant in the model. The reliability of the forecasting equation is examined by comparing the forecasting results with the true values of the testing set, which is shown in Figure 3(a) and Figure 4. The MAPE is 0.1841 and the NMSE is 0.4059. It can be concluded that the effect of linear regression model is not entirely satisfactory in this case study. In fact, most of the weather parameters are not independent of each other, for example, rain in hot days and rain in cold days have very different influence on the pedestrian flow. However, the linear regression model doesn't consider the mutual effect among variables, resulting in its bad performance.

The forecasting results of the BPNN model are shown in Figure 3(b). The correlation coefficient between training results and actual values is 0.9695, which means the model perform well on the training set. Meanwhile, from Figure 4 we can find that the forecasting results of the testing set is poor. It can be calculated that the MAPE is 0.2291 and the NMSE is 0.5114. The generalization performance of BPNN is highly correlated with the representativeness of the training set and it is likely to cause over-fitting in learning process.

The three key parameters (ε, C, γ) of SVR are optimized by APSO and equal (0.1230, 91.5206, 0.0410). The correlation coefficient between training results of SVR-APSO model and actual values is 0.9206. The MAPE is 0.1295 and the NMSE is 0.1363, which are the best results of these three methods. It is obvious that this model is the most accurate and generic. That is to say, APSO-SVR is an effective way to predict daily pedestrian flow in this case.

CONCLUSIONS

In this paper, the daily pedestrian flows in the Tiananmen Square are forecasted based on historical data, the distribution of holidays and weather conditions including rain, wind, temperature, relative humidity, and AQI (Air Quality Index). Some results about daily pedestrian flows are presented.

The statutory holidays, especially National Day, and Weekends have a great influence on the daily pedestrian flow in the Tiananmen Square. The population capacity of the Tiananmen Square is estimated, which is about 300,000, and it means once the daily pedestrian flow is over 300,000, emergency managers should be on the alert and keep an eye on the real-time pedestrian flow. The relationship between daily pedestrian flows and corresponding weather parameters is not linear. Among these three forecasting models we applied, the Support Vector Regression based on the Adaptive Particle Swarm Optimization (APSO-SVR) has been proved the most reliable and accurate model to forecast the daily pedestrian flows.

The forecasting results can help to conduct security pre-warning system and enhance emergency preparedness and management for crowd activities. It reveals the weather data and statutory holidays affect the pedestrian flows of Tiananmen Square, which need to be incorporated into the planning and forecasting system to improve the emergency planning process.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 71373139) and the Ministry of Science and Technology of the People's Republic of China under Grant No. 2015BAK10B02.

REFERENCES

1. Burger, C. J. S. C., Dohnal, M., Kathrada, M., & Law, R. (2001) A practitioners guide to time-series methods for tourism demand forecasting—a case study of Durban, South Africa, *Tourism management*, 22, 4, 403-409.
2. Chen, K. Y. (2011) Combining linear and nonlinear model in forecasting

tourism demand, *Expert Systems with Applications*, 38, 8, 10368-10376.

3. Chen, K. Y., & Wang, C. H. (2007) Support vector regression with genetic algorithms in forecasting tourism demand, *Tourism Management*, 28, 1, 215-226.
4. Chen, R., Liang, C., Ling, Y., & MA, Y. (2013) Forecasting Short-Term Tourism Flow of Mountain Resorts Based on Adaptive PSO-SVR, *Tourism Science*, 27, 3, 50-60.
5. Hornik, K., Stinchcombe, M., & White, H. (1989) Multilayer feedforward networks are universal approximators, *Neural networks*, 2, 5, 359-366.
6. Huang, H. C., Hwang, R. C., & Hsieh, J. G. (2002) A new artificial intelligent peak power load forecaster based on non-fixed neural networks, *International journal of electrical power & energy systems*, 24, 3, 245-250.
7. Law, R. (2000) Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting, *Tourism Management*, 21, 4, 331-340.
8. Rosselló-Nadal, J., Riera-Font, A., & Cárdenas, V. (2011) The impact of weather variability on British outbound flows, *Climatic change*, 105, 1-2, 281-292.
9. Schadschneider, A., Klingsch, W., Klüpfel, H., Kretz, T., Rogsch, C., & Seyfried, A. (2009) Evacuation dynamics: Empirical results, modeling and applications, *In Encyclopedia of complexity and systems science.* Springer New York. 3142-3176.
10. Scholkopf, B., & Smola, A. J. (2001) Learning with kernels: support vector machines, regularization, optimization, and beyond, MIT Press.
11. Singhal, A., Kamga, C., & Yazici, A. (2014) Impact of weather on urban transit ridership, *Transportation research part A: policy and practice*, 69, 379-391.
12. Vapnik, V. (2000) The nature of statistical learning theory, Springer.
13.