Design and implementation of ticket price forecasting system

Cite as: AIP Conference Proceedings **1967**, 040009 (2018); https://doi.org/10.1063/1.5039083
Published Online: 23 May 2018

Yuling Li and Zhichao Li

Design and Implementation of Ticket Price Forecasting System

Yuling Li [1, a)] and Zhichao Li [2, b)]

[1] *School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2] *Erdos City Electric Power Bureau, Inner Mongolia 017000, China*

[a)] 18813162153@163.com
[b)] 921194164@qq.com

Abstract. With the advent of the aviation travel industry, a large number of data mining technologies have been developed to increase profits for airlines in the past two decades. The implementation of the digital optimization strategy leads to price discrimination, for example, similar seats on the same flight are purchased at different prices, depending on the time of purchase, the supplier, and so on. Price fluctuations make the prediction of ticket prices have application value. In this paper, a combination of ARMA algorithm and random forest algorithm is proposed to predict the price of air ticket. The experimental results show that the model is more reliable by comparing the forecasting results with the actual results of each price model. The model is helpful for passengers to buy tickets and to save money. Based on the proposed model, using Python language and SQL Server database, we design and implement the ticket price forecasting system.

INTRODUCTION

With the development of the transportation industry, the airplane has become a more popular way of traveling. The advent of the Internet enabled people to check, book and buy air tickets online. Airline companies through revenue management policies to develop airfare discount, according to the current flight seat remaining, the seasons and other factors change, continue to adjust the price, making the same flight different purchase time prices will be different. This process is designed to maximize the airline's revenue. With the advent of the Internet, product prices are becoming more visible online, enabling consumers to effectively compare and track product prices over time. Consumers generally want to buy airfares at the lowest possible price, but changes in ticket prices depend on many different dynamic factors such as airline pricing policies, flight distances, service categories, airlines, global population movements, and more. The statistical nature of price changes shows that the higher the ticket price, not the shorter the time it takes off, or the lower the time it takes to take off. Therefore, the accurate forecast of ticket prices has good application value and broad prospects, which can help passengers to make appropriate purchase decisions and save expenses.

At present, there are many papers that use data mining techniques to forecast prices, and there are many papers that use data mining techniques to predict the ups and downs of stocks. In 2003, Hamlet algorithm is used to deal with the flight price, and the three kinds of algorithms are used to form a new model [1]. A Lantseva, and K Mukhina establish a regression model to predict the price [2]. In 2015, W Groves and M Gini use feature selection techniques to give the lowest price for all flight forecasts [3]. Gu Zhaojun and Wang Shuang establish the time series model to predict the ticket prices. In this paper, time series model is constructed based on the first order moving average model [4]. However, because of different revenue management between China and other countries, foreign forecasting methods are not suitable for China's ticket price changes.

The purpose of this study is to establish a ticket price forecasting model and to design and implement a ticket price forecasting system. This work includes the following:

- Collection, processing and fusion of data from 9 cities in china;
- Establish the prediction model based on ARMA algorithm and Random forest algorithm;

6th International Conference on Computer-Aided Design, Manufacturing, Modeling and Simulation (CDMMS 2018)
AIP Conf. Proc. 1967, 040009-1–040009-6; https://doi.org/10.1063/1.5039083
Published by AIP Publishing. 978-0-7354-1672-7/$30.00

- Design and implement ticket price forecasting system.

BRIEF THEORETICAL BACKGROUND

In this article, three different models are used to predict the ticket price. The first model is the ARMA model, which treats the data sequence formed over time as a random sequence, this set of random variables reflects the continuity of the original data in time. On the one hand, it reflects the influence of influencing factors, on the other hand, it has its own variation rule. Therefore, ARMA model is used to predict the variation of air tickets. The second model is a random forest model. The prediction of air ticket price needs to consider the direct competition of airlines, the time when the aircraft took off, the date characteristics of the takeoff, and so on, we also need to consider the interaction of variables. Therefore, we use random forest algorithm. The third model is the time series-random forest model. The prediction of the third model is used to is used to compare with the forecast results of the combined algorithm proposed in this paper. We will experimentally prove that the combination algorithm has some advantages. Before describing the complete methodology of our system, we would like to begin by presenting some theoretical background on these.

Random Forest Algorithm

Random Forest is a method of artificial intelligence, which is an integrated learning method that uses bagging algorithm combined with many decision trees for classification, regression or other tasks. Random forest is a forest composed of many decision trees, and the decision trees are independent of each other. Random forest, proposed by Leo Breiman, uses a bootstrap resampling technique [5] to randomly extract k samples randomly from the original training sample set N to generate a new set of training samples. Then generate k classification trees to form random forest, and the classification result of new data depends on the classification tree voting.

ARMA algorithm

ARMA model time series analysis is a method of using parametric model to process random vibration response data for pattern parameter identification. The parametric models include AR autoregressive model, MA sliding average model and ARMA autoregressive moving average model. The ARMA model treats the data sequence formed over time as a random sequence. The dependence of this set of random variables reflects the continuity of the original data in time.

DATA COLLECTION

We collected airfare data directly from a major travel web site. For the purpose of our pilot study, we have access to the domestic 18 routes, 9 cities ticket price data. We collect data 60 days in advance at twelve-hour intervals. The data we collected have the following characteristics: city of departure, destination, ticket purchase date, departure date, ticket options with the price, time of departure. Data processing: data cleansing, deleting data that is unrelated to the ticket forecast process.

PREDICTION MODEL

This article builds a ticket price forecasting model using a combination of ARMA algorithm and random forest algorithm. In this paper, ARMA-Random forest combination model is constructed to predict ticket prices. Based on the preliminary prediction of ARMA algorithm, the random forest algorithm is used to predict the residual error and then get the forecast price. Combination algorithm flow as shown below:

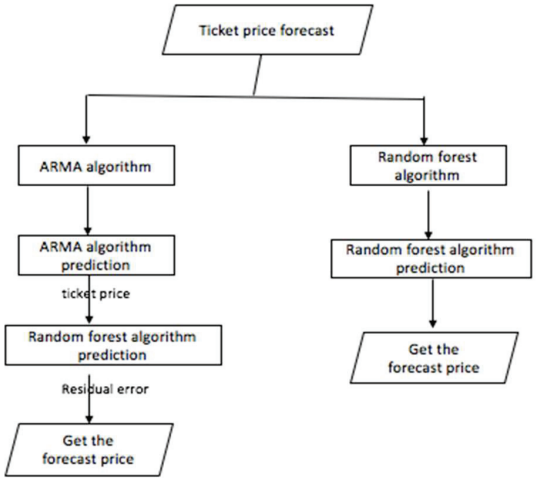

FIGURE 1. Complete system process, including ARMA and random forest algorithm

EXPERIMENTS AND RESULTS

In this section, ARMA-Random forest combining algorithm is used to forecast the ticket price respectively. We use the route PEK-HET to show the experimental results.

We establish the ARMA model based on the minimum AIC criterion. Taking Route HU7175 as an example, the following figure 2 shows the time series autocorrelation diagrams and partial autocorrelations. The AIC rule of the ARMA model is usually used to determine the most suitable model. By calculating, ARMA (2,0) in this time series has the smallest AIC and BIC, which is the best model.

Take flight HU7175 as an example. The ticket price forecast results of the ARMA-Random forest model are shown in Figure 3. As can be seen from the figure, except for a few points, ARMA-Random Forest algorithm can better fit the ticket price.

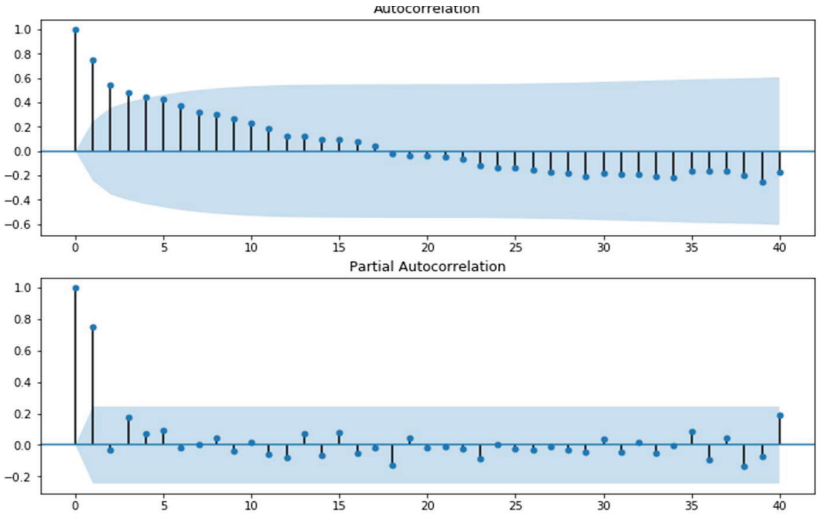

FIGURE 2. Autocorrelation diagrams and partial autocorrelations

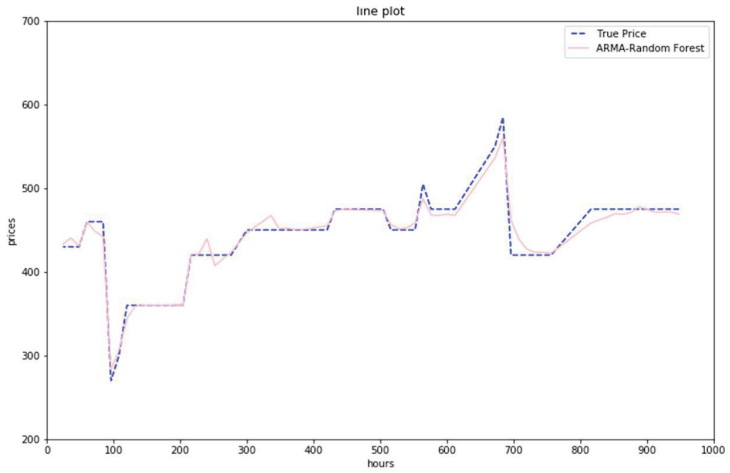

FIGURE 3. ARMA - Random forest Fitted Data versus original for HU7175 flight

Analysis of Experimental Results

In the experiment, random forest algorithm, time series - random forest algorithm and ARMA-Random forest algorithm are used to predict ticket prices. time series - random forest algorithm model is a combination of time series proposed in document 4 and random forest algorithm. Figure 7 shows the prediction of the ticket prices of the random forest prediction model, the time series - random forest prediction model and ARMA-Random forest algorithm combined model. Compare the predictions of each model. We can observe that ARMA-Random forest combination model can have the best prediction in the above model.

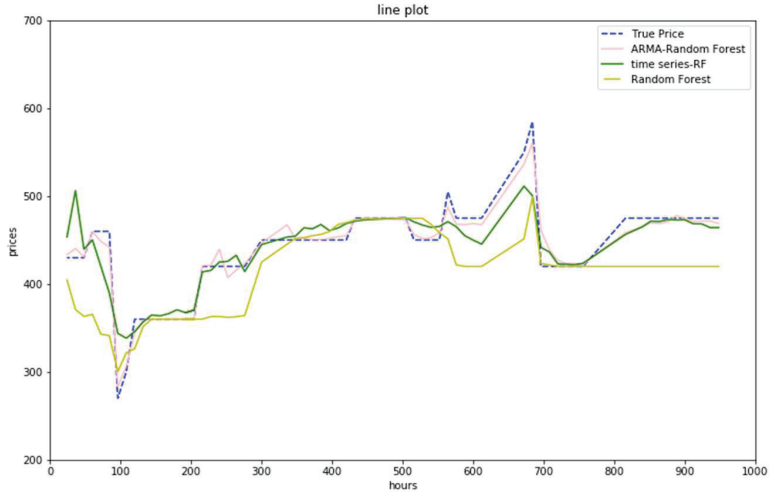

FIGURE 4. PEK-HET, HU7175 flight, fitting curve of each model

TICKET PRICE FORECAST SYSTEM

Based on the theory of combinatorial algorithm proposed above, this section designs a ticket forecasting system, which realizes the modules of forecast management, data management and user management. The architecture of the prediction system is divided into three layers, namely the view layer, the function layer and the data layer. The frame

is shown in Figure 5 below. Based on the principles of practicality, stability, maintainability and safety of system design, design and implementation of the ticket price forecasting system interface as shown in Figure 6.

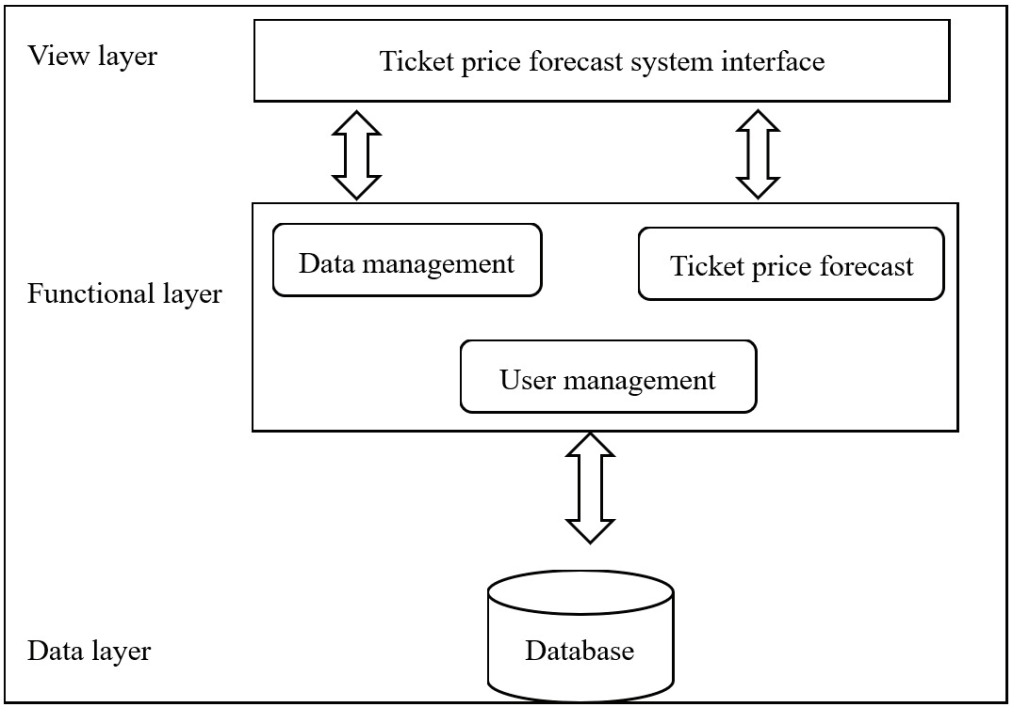

FIGURE 5. System architecture diagram

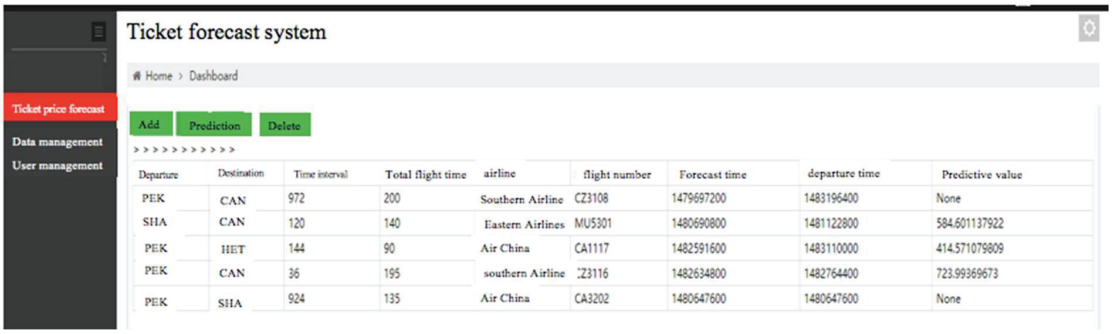

FIGURE 6. Ticket price forecast system interface

CONCLUSION AND FUTURE WORK

In conclusion, this paper presents a ticket price forecasting model based on ARMA algorithm and stochastic forest algorithm. In the experiment, random forest algorithm, time series - random forest algorithm and ARMA-Random forest algorithm are used to predict ticket prices. Through experiments, the prediction reliability of the model proposed in this paper is proved. Based on the proposed algorithm, the system of ticket price forecasting is designed and implemented to provide strong support for air tickets purchase. In the future, more ticket data needs to be acquired and algorithms need to be optimized to improve forecast accuracy

REFERENCES

1. Etzioni, O., Tuchinda, R., Knoblock, C. A., & Yates, A. To buy or not to buy: mining airfare data to minimize ticket purchase price[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, Dc, Usa, August. 2003:119-128.
2. A Lantseva, K Mukhina, A Nikishova, S Ivanov, K Knyazkov, “Data-driven Modeling of Airlines Pricing”*Procedia Computer Science*, 2015, 66:267-276.
3. Groves, William, and M. Gini. "On Optimizing Airline Ticket Purchase Timing." ACM Transactions on Intelligent Systems & Technology 7.1(2015):1-28. Gu Zhaojun, Wang Shuang, Zhao yi. "A forecasting model of air ticket price based on time series. " [J]. Journal of Civil Aviation University of China, 31.2(2013):80-84.
4. Gu Zhaojun, Wang Shuang, Zhao yi. "A forecasting model of air ticket price based on time series. " [J]. Journal of Civil Aviation University of China, 31.2(2013):80-84.
5. Breiman, L. Random forests. Machine Learning 2001,45(1), 5–32.
6. A Lantseva, K Mukhina, A Nikishova, S Ivanov, K Knyazkov. Data-driven Modeling of Airlines Pricing ☆ [J]. Procedia Computer Science, 2015, 66:267-276.
7. Webmaster. Statistics and Machine Learning[M]// Foundations of machine learning /. MIT Press, 2012:287--306.
8. Domínguez-Menchero, J. Santos, J. Rivera, and E. Torres-Manzanera. "Optimal purchase timing in the airline market." Journal of Air Transport Management 40.40(2014):137-143.