Restaurant Visitor Time Series Forecasting Using Autoregressive Integrated Moving Average

Article *in* Journal of Computational and Theoretical Nanoscience · May 2018

DOI: 10.1166/jctn.2018.7345

4 authors, including:

A. Anandaraj
Sri Eshwar College of Engineering
12 PUBLICATIONS **23** CITATIONS

SEE PROFILE

Journal of
Computational and Theoretical Nanoscience
Vol. 15, 1590–1593, 2018

RESEARCH ARTICLE

Restaurant Visitor Time Series Forecasting Using Autoregressive Integrated Moving Average

G. Boomija*, A. Anandaraj, S. Nandhini, and S. Lavanya

Department of Computer Science, Narasu's Sarathy Institute of Technology, Salem 636305, Tamil Nadu, India

Many restaurants may not have proper prediction system for forecasting the visitors. The main aim of this paper is to forecast visitors to restaurants in Japan. There is a need to know how many visitors to expect each day. Forecasting helps in effective planning for purchase of ingredients and scheduling of staffs. The visitor forecasting is challenging because unpredictable factors do affect the arrival of visitors to restaurants. This paper builds forecast of visitors using regression techniques. In order to estimate the visitors, datasets of reservation and visitation of the restaurants are used. The time series data is analyzed using Box-Jenkins ARIMA models the residuals have the structure of AR (Auto Regressive). Forecasting of visitors will provide information that helps the restaurants to be much more successful in overcoming the troubles caused by factors. This also allows the restaurants to focus on creating an enjoyable dining experience for visitors.

Keywords: Forecasting, Regression Techniques, Time Series Data, Auto Regressive, ARIMA.

1. INTRODUCTION

There is a steady increase of restaurants in the 21st century and running a local restaurant is not an easy job. The number of visitors to a restaurant for a period of time is random. The advent of many a restaurants in a region marks the chance of the existing restaurants to lose scope. There is a predicament with expectation of visitors to a restaurant. It seems to be inconvenient for the successful run of the eatery in a locality thus, disturbing the economic attributes of the functioning market. Forecasting plays a major role with planning the total number of visitors using the time series data. The reservation of visitors are also a part of the prediction perspective. The challenges include in building a technically efficient model to forecast and this paper explores the appliance of time series forecasting model Autoregressive Integrated Moving Average (ARIMA). The ARIMA approach planned by Box and Jenkins is popular as it yields most accurate forecasting results. A study of time series data of visitors to restaurants in Japan over a period of 2016–2017 to create models.

2. RESTAURANTS AND FACTORS AFFECTING ITS PROCEEDS

Restaurants are public place that provide food on commercial basis. Restaurants offer food as a service in desire to satisfy the visitors. Prediction of these visitors helps the restaurants to improve its functional services and this paper focuses the purpose. Recruit Holdings has unique access to restaurant data. Hot Pepper Gourmet—a restaurant review service, AirREGl—a restaurant sales service and Restaurant Board—reservation log management software are all owned by Recruit Holdings. Using these data retrieved with assistance from websites, forecasting of visitors becomes evident. There are certain factors that might stop the restaurant's attendance. These factors include weather conditions, local competitions and new restaurants with few historical data. Weather is always unpredictable in Japan, in spite of the usual monsoons and summers. Local competition is undesirable based on location and new restaurants will have less past data which is insufficient to forecast visitors in future.

3. DATASETS INVOLVED IN PREPROCESSING

Time series datasets are available from the two sites, Hot Pepper Gourmet (hpg) and AirREGl (air). Through hpg the visitors can search restaurants and reserve online. While air is a reservation control and cash register system. The datasets of reservation and visitation for a year are mainly used. As it is a time series data, time series models are applied. The datasets used are reservation dataset of air and hpg followed by the information of restaurants where latitude and longitude is a part of locating the restaurants.

*Author to whom correspondence should be addressed.

 1546-1955/2018/15/1590/004 doi:10.1166/jctn.2018.7345

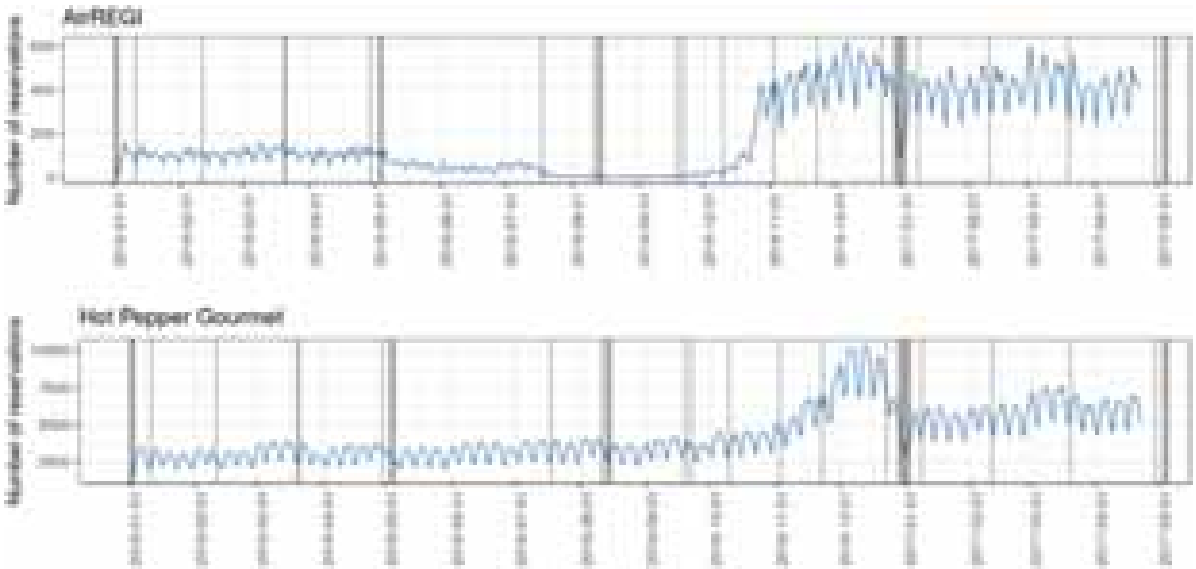

Fig. 1. Reservations made with both the sites.

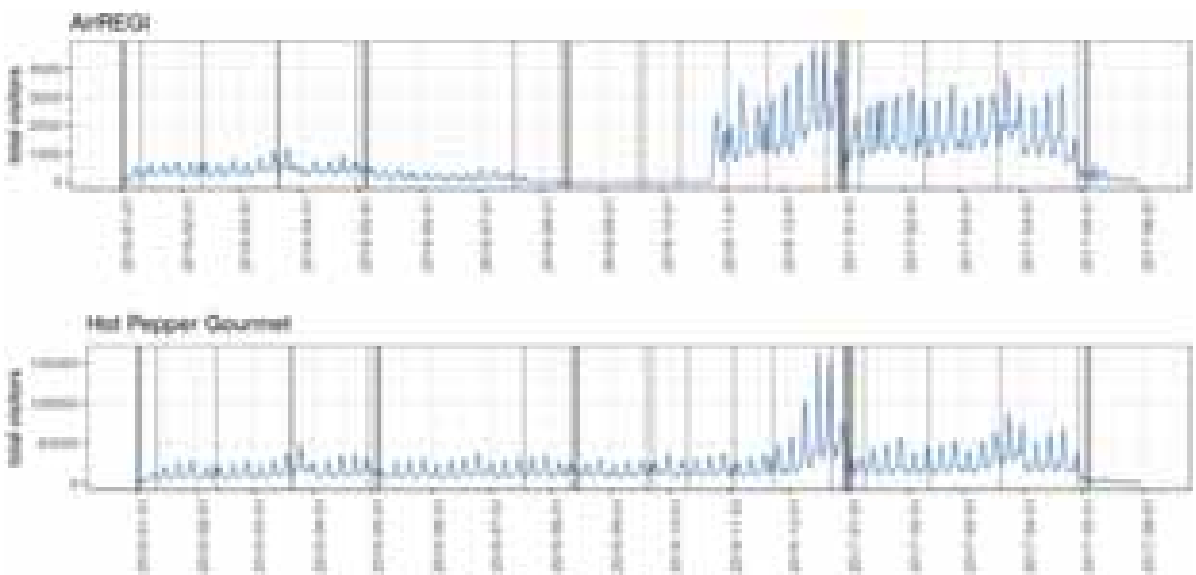

Fig. 2. Total number of visitors based on reservation.

A relation dataset is used, where the common restaurants in both the sites are filtered. Information of date is analyzed to withstand with holidays and Japan has a week's holiday called "The Golden Week." A summary of visits are analyzed as Explanatory Data along with holiday and reservations made. There is a comparison between restaurants based on the cuisine offered in both the sites.

Figure 1 portrays the number of reservation made based on the date in both the sites.

Figure 2 depicts the number of visitors who have arrived for the reservations made on that specific date for the restaurants at both the sites.

4. PURPOSE OF TIME SERIES IN PROCESSING

A time series is a structured dataset that can be annual, monthly, or in this case daily. The visit data and reserve data used here are time series data. Even date is analyzed for better results. Descriptive model is applied for making predictions. The visit data specifies the time of reservation and reserve data specifies the time when the reservation was made. This model helps in describing the trends, patterns and it uses the information in time series to forecast future values of that series. Making prediction about future is called as extrapolation. Forecasting

involves taking models that fit on historical data and use them to predict future. The purpose of time series analysis are to understand or model the stochastic mechanism that gives rise to an observed series and to predict or forecast the future values of a series based on the history of that series. Time series include four parts level, trend, seasonality and noise. Level is undesirable because it forms the baseline values and in this paper the baseline values are the data of date and time. The next important feature is seasonal variation, where seasons in Japan play a vital role. These are the parameters of time series considered in this paper.

5. FEATURE ENGINEERING FOR BETTER ESTIMATION

There is no source on how to do feature engineering and it depends on the problem. Feature engineering is a thoughtful creation of new input fields from existing. These input fields could result in better inferences. In this case of forecasting the number of visitors to a restaurant, there is an observation of weather as a factor. So, adding an extra feature of weather to the dataset can represent the result at its best. The inputs can be turned into a version which the model can understand. The data can be analyzed in two perspectives. One is to see the difference pattern between visitors on days of holidays and non-holidays and the restaurant visitor behavior. This varies from restaurant to restaurant and processing using these can add a feature to the result as well. In the datasets, we can even feature snowfall record in a particular region, pressure and humidity. Even some stations might have a record of solar radiation and cloud cover as a feature of weather. There will be some important features which can lead to accuracy in forecasting and some features that seem irrelevant. The relevant factors are featured in this paper.

From Figure 3 changes in graph can be seen. The gray lines indicate holidays and the number of visitors vary with different sites and the restaurants present in common with both the sites.

By Figure 4 the visitors for a particular week is considered and the number of visitors during Friday, Saturday and Sunday is high in comparison with the remaining days of the week.

6. ARIMA-A TIME SERIES MODEL FOR FORECASTING

Box and Jenkins developed a time series model ARIMA, which combines three types of processes. This model is used with time series data to understand the data or to predict future in series.

- AR: Auto regression, A model that uses the dependent relationship between an observation and some number of lagged observations.
- I: Integrated, The use of differencing raw observations to make the time series stationary.
- MA: Moving Average, A model that uses the dependency between an observation and residual error.

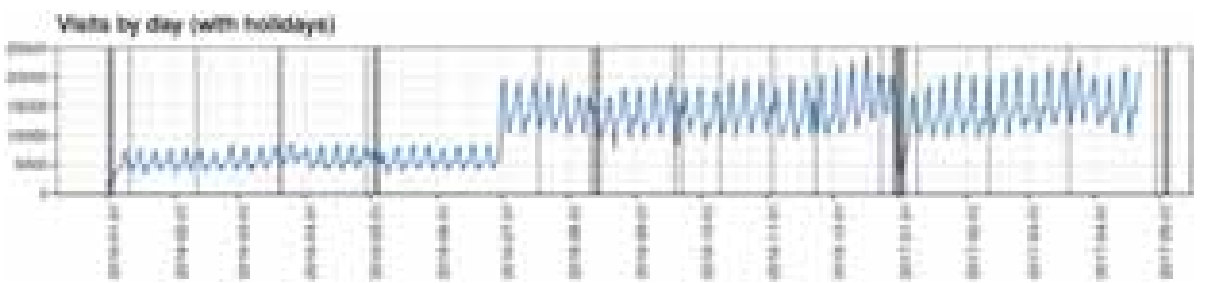

Fig. 3. Visitor forecast during holidays and non-holidays.

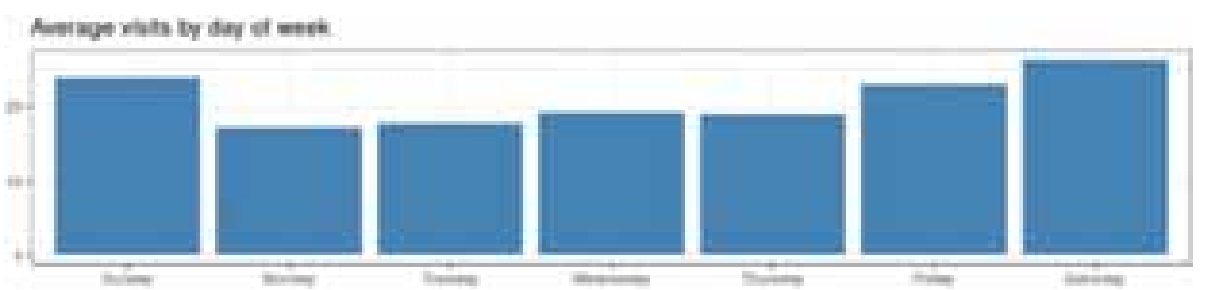

Fig. 4. Visitor forecast for a week.

ARIMA has three parameters

- p: Lag order.
- d: Degree of differencing.
- q: Order of moving average.

For building a stochastic model, three steps are involved as an iterative approach. Identification, Estimation and Diagnostic checking. ARIMA model gives better results for forecasting and this time series model is employed in forecasting the visitors to restaurants in Japan where p, q and d are integers greater than or equal to zero.

7. FORECASTING OF VISITORS USING ARIMA

The main technique being explored is ARIMA which is a forecasting technique that projects the future values of a series based entirely on its own inertia. Its main application is in the area of short term forecasting requiring at least 40 historical data points. The auto regressive part of the model determines the dependence on previous time points. The data is split into training set and testing set. The training period was different to testing period as there are different cuisine restaurants in this case. So, the train period is shortened and the test period too. Here period is weekly cycle and the visitors per cuisine and per restaurant is valuated. These data are taken for training to the ARIMA model. As ARIMA model uses (p, d, q) the series should be made stationary and choosing of order models with ACF and PACF for fitting the ARIMA to the state of train is fixed.

8. PERFORMANCE MEASURES USING RMSE AND MAPE

The prediction of visitor arrivals during testing is done and the measures of accuracy were calculated. Root mean square error and Mean absolute percentage error were used for forecasting error in time period and to measure accuracy of continuous variables. The average y value with given x value is predicted using regression. They can be positive or negative as the predicted value under or over estimates the actual value. Squaring the residuals, averaging the squares, and taking the square root gives the r.m.s error. Then use the r.m.s. error of the y values as a measure of the spread about the predicted y value. It's the square root of the average of squared differences between prediction and actual observation.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(X_{\mathrm{obs},i} - X_{\mathrm{model},i})^2}{n}}$$

MAPE measures the average magnitude errors in a set of predictions without directions. It is the average over the test sample of absolute difference between predictions and actual observations.

$$\mathrm{MAPE} = \sqrt{\frac{\sum_{i=1}^{n}(X_{\mathrm{obs},i} - X_{\mathrm{model},i})}{n}}$$

Both are used to express average and RMSE have interesting implication for large errors. RMSE is more useful than MAPE because it is steady and RMSE does not necessarily increase with the variance of the errors. RMSE increases with the variance of the frequency distribution of error magnitudes. Hence RMSE is better.

9. CONCLUSION

This paper explores a growing trend in the future of restaurants by forecasting the number of visitors each day. In this study the time series forecasting model ARIMA is functional. Based on the above an efficient prediction of visitors to come in future was analyzed for the profits of the restaurant. This prediction results in easy scheduling of ingredients and staffs for enhanced experience of visitors. Even offers can be offered during bad weather conditions for the successful running of the restaurants. The results of the study is revealed using metrics like RMSE and MAPE, where RMSE gives better accuracy in predicting the visitors. The information thus obtained from forecasting the visitors will act as an aid of support to the restaurants in focusing with the development of creating an enjoyable dining experience for the visitors. As far as the experimental results, the system thus proposed works well in forecasting visitors and is robust too.

RESEARCH ARTICLE

References

1. C. Shu and A. Tsang, Adjusting for the Chinese New Year: An Operational Approach, External Department Hong Kong Monetary Authority **(2005)**.
2. K. Torre, D. Delignieres, and L. Lemoine, *British Journal of Mathematical and Statistical* 60, 85 **(2007)**.
3. T. Trimbur, Seasonal heteroskedasticity in Census Bureau construction series, Statistical Research Division U.S. Census Bureau Washington DC. 20233–9100 **(2006)**.
4. F. L. Chu, *Tourism Management* 29, 79 **(2008)**.
5. Tourism Council of Thailand, http://www.thailandtourismcouncil.org/home.php **(2009)**.
6. L. Vaughan and E. Romero-Frías, *Journal of the American Society for Information Science and Technology* 65, 707 **(2013)**.
7. S. Vosen and T. Schmidt, Forecasting private consumption **(2011)**.
8. W. H. K. Tsui, H. O. Balli, A. Gilbey, and H. Gow, Forecasting of Hong Kong airport's passenger throughput **(2014)**.
9. J. J. Flores, M. Graff, and H. Rodriguez, Evolutive design of ARMA and ANN models for time series forecasting **(2012)**.
10. I. Khandelwal, R. Adhikari, and G. Verma, Time Series Forecasting using Hybrid ARIMA and ANN Models based on DWT Decomposition **(2015)**.

Received: 19 April 2018. Accepted: 14 May 2018.