A Time Series Case-Based Predicting Model for Reservation Forecasting

Tsung-Hsien Tsai and Sheryl E. Kimes

Abstract. This study addresses how to construct sales forecasting models by using restaurant reservation data. The issues of how to retrieve booking patterns, search for influential parameters, and divide samples for training, validating, and testing are discussed. Regression and Pick Up models, which are common practice, are also built as benchmarks. We used data from a mid-sized restaurant to show that the proposed Time Series Case-Based Predicting model can significantly outperform the benchmarks in all testing cases.

Keywords: Time Series Case-Based Predicting, Advanced Booking Model, Sample Selection, Revenue Management, Data Mining.

1 Introduction

Demand forecasting is important because it provides input information for the efficient and profitable operation of business enterprises. In revenue management applications, forecasting is essential for resource allocation and overbooking. The benefit of improving forecasting accuracy in revenue management has also proven to be significant [1].

Data used for revenue management forecasting has two dimensions to it: when the reservation was booked and when the service took place. The booking information gives additional detail that can be used to update the forecast. Without this information, the forecast would be based solely on the historical information on the daily number of customers served.

Three forecasting approaches have been identified [2]. Historical booking models use historical arrival data to predict the future. For instance, all historical final sales data are used to project future sales [3]; Exponential Smoothing and Autoregressive Integrated Moving Average (ARIMA) fall into this category. Advanced booking models use information on when customers placed their reservation to develop forecasts; Pick Up and Regression models are frequently used for this purpose. Combination methods use a weighted average of the historical and

Tsung-Hsien Tsai and Sheryl E. Kimes
School of Hotel Administration, Cornell University,
14850 Ithaca, United States of America

B.-C. Chien and T.-P. Hong (Eds.): Opportunities and Challenges, SCI 214, pp. 53–58.
springerlink.com

advanced booking models [4]. Much of the sales forecasting research uses historical booking models but few papers have studied advanced booking models. The aim of this study was to construct an advanced booking model with the potential to improve predictive accuracy.

In this study, we proposed a novel advanced booking model based on the concept of pattern retrieval. A four-stage procedure was provided to project forecasts; furthermore, the issue of sample selection was discussed. We tested models on reservations data from a 100-seat restaurant.

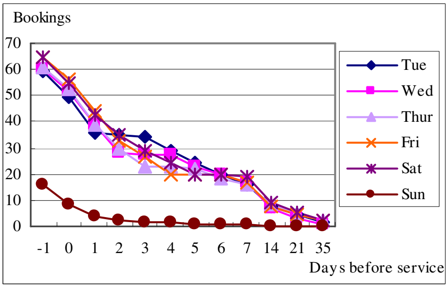
2 Booking Data

A restaurant manager may open seats for reservations via different channels several weeks (or months) before a specific service date. Customers who do not make reservations and show up in the restaurant are categorized as walk-ins. When customers make their reservations, the restaurant's computer system records the reservation date and the number of seats reserved. If we accumulate this volume of reservations over the whole booking period plus walk-ins, then we can obtain the volume of final sales.

We used 20 months of detailed reservations data from a 100-seat restaurant to develop our model. The restaurant has a busy lunch period and tracks their reservations and walk-ins with a reservation system called OpenTable.com. OpenTable.com is used by over 8000 restaurants worldwide and seats over 2 million customers each month. Restaurants can use OpenTable to track both their reservations (regardless of whether they are made online or over the phone) and walk-in business. The data includes information on when the reservation was made, the date and time of service and the number of people in each party. This provided us with the necessary information to develop booking curves.

The graph of the complete booking data for each service date shows the number at which reservations are received (Fig. 1.). DBS (-1) represents the number of walk-ins plus all reserved customers on a service date; DBS (0) is the number of all reserved customers on a service date; DBS (k>0) represents the number of accumulated reservations k days before a service date. Historical booking models use only DBS (-1) data for model construction while advanced booking models use all booking data which availability depends on DBS (k).

We used our booking data to develop booking curves for the restaurant by service date. Average booking curves were developed by day of week (Fig. 1.) and showed that on average most reservations were made within 3 days of service. Without the influence of cancellations, the booking curves are all monotonic increase, and the phenomenon implies that the reliability of information improves over time. While the averages are helpful, the daily booking curves showed quite a bit of week-to-week variation (Fig. 2.). Fig. 1. also showed weekday and weekend effects although weekday demand was fairly similar. The variation of final sales also rendered valuable information (Fig. 3.). The data was non-stationary, and seasonality may be the driver causing this phenomenon which sales were generally higher in May than those in other months. In addition, there was no significant trend and cycle was not also expected during the research period.



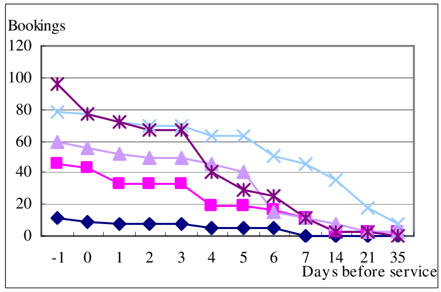


Fig. 1 Average Booking Curves

Fig. 2 Weekly variation of Tuesdays

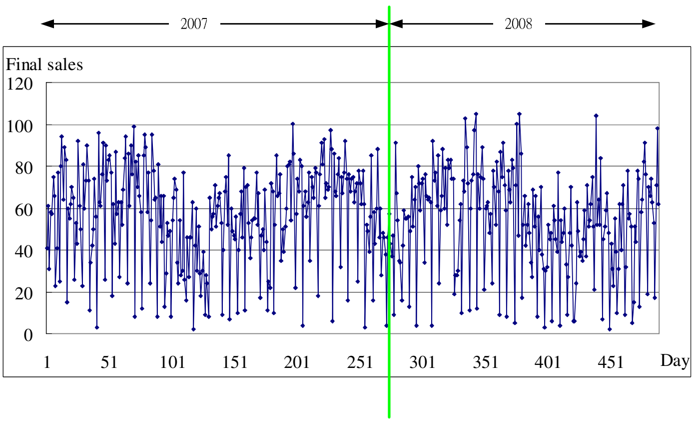


Fig. 3 The volume of final sales over the research period

3 Time Series Case-Based Predicting Model

In the following, $x_{j,k}$ is the number of accumulated reservations for service date j at booking point k. Forecasts are developed for the final sales of service date j, $x_{j,-1}$. The data collection point (DCP), m, is the point where the computer system collects data and updates forecasts. It is common to have updates once for a reservation several weeks before the service date and more frequent recalculations as the service date approaches. In this study, the DCPs used were -1, 0, 1, 2, 3, 4, 5, 6, 7, 14, 21, and 35 in terms of the data we collected.

CBP is composed of four stages, the first being similarity evaluation, which calculates similarity between booking curves in the database and the booking patterns of a targeted service date. The calculations are updated each time new reservation information becomes available at each DCP. CBP also incorporates a temporal characteristic that considers information at each DCP to be of exponential importance with a parameter α to show the reliability of information over DCPs. Equation (1) computes the distance between the service date j and a booking curve i at DCP (k) (k indicates the current DCP, and t is the first DCP).

$$D_k(j,i) = \sum_{m=k}^{t} (x_{j,m} - x_{i,m})^2 (\frac{1}{m})^{\alpha} \ . \tag{1}$$

The second phase is to select the most similar booking curves in terms of the calculation in the first stage. CBP ranks the similarity of all booking curves and selects ten most similar samples based on the patterns of the targeted service date.

The third step is to integrate the final sales of the selected booking curves. Instead of calculating a simple average for these final sales, CBP incorporates the influence of similarity and also adds an adaptive term to enhance the importance of the current reservation information (Equation (2)). β and γ are parameters to show the effects of similarity and adaptability, respectively.

$$\hat{x}_{j,-1}^{k} = \sum_{s=1}^{10} \left(\frac{\frac{1}{D_k(j,s)}}{\sum_{s=1}^{10} \frac{1}{D_k(j,s)}} \right)^{\beta} \left(\frac{x_{j,k}}{x_{s,k}} \right)^{(\frac{1}{k})^{\gamma}} x_{s,-1} \ . \tag{2}$$

The last step is to search for three parameters in Equations (1) and (2). It is difficult to apply a conventional gradient-based algorithm. As a result, we applied the Hooke-Jeeves algorithm [5], which is a direct search method, to find α, β, and γ. The weakness of direct search algorithms is the possibility to stick into local minima. In this study, the multi-start strategy, which tries different initial seeds, was applied to select the most possible global minima.

Another focus of this study was to investigate what samples in the database should be used for searching a suitable combination of parameters. Before going forward, we divided the collected booking curves into three categories: training, validating, and testing samples. In Regression and Pick up models [3], the current practice is to use all data except testing samples for estimating parameters. The parameters are then used to forecast and the predictive performance is evaluated by using testing samples. In CBP, the calibrating procedure is different because it tries to decide parameters by matching patterns between training and validating samples. One possibility is to set training samples as the base and minimize mean square errors (MSE) of the validating ones. The obtained parameters are then verified by using testing samples. The problem is how to decide training and validating samples so that a valid combination of parameters can be obtained.

We tested three mechanisms of sample division for the CBP parameter search. The first method is the sequential method (SEQ) and uses the month of data immediately before the testing sample as the validating set. The concept of this approach is to use the most recent booking trends. Randomly (RAN) selecting a certain number of samples (about one month) from all but testing data is another way to select validating samples. The idea of this approach is to learn patterns from different time periods. The last approach is to take the same month of the previous year (LAST) as the validating month. For example, May in 2008 is taken as the validating samples while predicting sales in May 2009. The logic behind this approach is to learn the patterns with the same time factor from the past.

4 Empirical Study

We tested the performance of CBP and also the above three division mechanisms. In order to have an overall consideration of time effects, we first saved one-year daily data for both training and validating purposes (2/07~1/08) and tested the performance of next month (2/08). Once the performance was computed, the testing samples were included into the training and validating database (2/07~2/08), and the model was re-optimized to test the performance of the next month (3/08). The procedure was repeated dynamically until all testing samples were exhausted (9/08). The purpose is to verify CBP's performance dynamically.

We next studied the method to select validating samples, and LAST obtained the best predictive accuracy of the three approaches. Fig. 4. shows the average improvement of MSE by using LAST in comparison with using SEQ and RAN, respectively (Equation (3)). LAST outperformed SEQ in 6 out of 8 cases and beat RAN in 5 out of 8 cases. More importantly, LAST usually was significantly better than the other two alternatives. This was because LAST seizes seasonal effects properly; SEQ and RAN systematically under- or overestimate in some situations.

Another important observation was to see whether the proposed CBP model and LAST procedure could outperform conventional Regression and Pick Up models. Fig. 5. displays the average improvement of MSE (formula is analogous to Equation (3))by using the proposed CBP in comparison with using Regression and Pick Up models. It is apparent that the proposed CBP can significantly outperform two conventional benchmarks. This result shows the value of CBP and its potential for predicting sales by using reservation data.

$$\frac{MSE^{LAST} - MSE^{Base}}{MSE^{Base}} \times 100\% \quad (3)$$

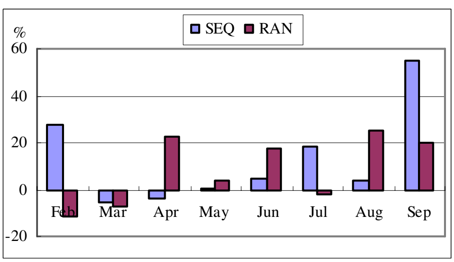


Fig. 4 Average improvement of LAST

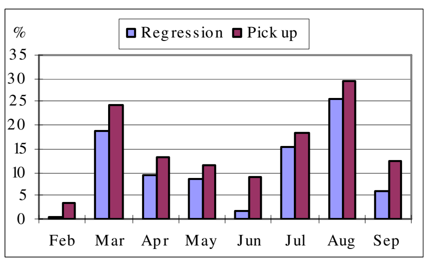


Fig. 5 Average improvement of CBP

5 Conclusions

Forecasting accurately is important for making correct decisions in daily operations. Regression and Pick Up are two common models used for arrival or sales forecasting. In the empirical study, we demonstrated that the proposed CBP model with careful selection of samples leads to a better combination of parameters and results in better predictive accuracy in comparison with the benchmarks.

In this study, the variation of final sales seems to be non-stationary, as shown in Fig. 3. As a result, the conclusions obtained in this study may be only valid for problems with similar data characteristics. It would be interesting to study what would happen if final sales have a linear trend or other periodic patterns, and how we should redesign the model so that CBP can still maintain its edge.

References

1. Lee, A.O.: Airline Reservations Forecasting: Probabilistic and Statistical Models of the Booking Process. Massachusetts Institute of Technology Press, Boston (1990)
2. Wickham, R.R.: Evaluation of Forecasting Techniques for Short-term Demand of Air Transportation. Massachusetts Institute of Technology Press, Boston (1995)
3. Weatherford, L.R., Kimes, S.E.: A Comparison of Forecasting Methods for Hotel Revenue Management. International Journal of Forecasting 19, 401–415 (2003)
4. Rajopadhye, M., Ghalia, M.B., Wang, P.P., Baker, T., Eister, C.V.: Forecasting Uncertain Hotel Room Demand. Information Sciences 132, 1–11 (2001)
5. Himmelblau, D.M.: Applied Nonlinear Programming. McGraw-Hill, U.S.A. (1972)