A Self-learned Advanced Booking Model to Forecast Railway Arrivals

Tsung-Hsien Tsai
National Quemoy University
Kinmen, Taiwan

Abstract
This paper utilizes data collected during the booking period and presents a novel advanced booking model to predict railway arrivals. As the date is approaching the departure day, more and more booking information is cumulated in the railway database. We extract useful patterns based on temporal features and formulate a two-stage model. The first stage evaluates the similarity among booking curves which describe the variation of passenger behavior. Then a group of samples which are similar to the booking patterns of the forecast target are chosen to project the prediction value based on a weighted scheme. Since the proposed model has several parameters for capturing the distinctive influence of temporal features, a direct search method is incorporated to determine the values of the parameters in the second stage. In addition, two common advanced booking models termed pick-up model and regression are also constructed in this study for comparative purposes. The results show that the proposed model may achieve at least 18% improvement in terms of predictive accuracy comparing with the benchmarks. However, more evidences are needed to validate the capability of the proposed concept in the future.

Keywords: Advanced booking model; Temporal features; Sales forecasting; Unconstrained optimization; Revenue management
1. Introduction

Revenue management (RM) integrates the functions of forecasting, seat allocation, pricing and overbooking to effectively utilize resources and maximize revenue. Kimes (2005) reported that the use of RM increases 3% to 5% revenue in hotel, car rental and airline industries. Forecasting, in fact, renders demand estimation as the input for the other three systems. Improving the accuracy of forecasting directly upgrades RM’s performance; Lee (1990, p.2) indicated that 10% increase in predictive accuracy can result in 0.5% to 3% increase in revenue.

Forecasting in RM covers a wide range of problems (McGill and Van Ryzin, 1999; Van Ryzin, 2005): the demand of passenger arrivals, the behavior of cancellation, no-show and show-up rates, customers’ preference choices, the estimation of unconstrained demand, the comparison between top-down and bottom-up methods to obtain disaggregate demand. This study focuses on the problem of how to forecast the demand of passenger arrivals. One way to tackle the issue is to assume data generating processes, such as poisson and gamma distributions (Kimms and Müller-Bungard, 2007; Swan, 2002). The other is to build so-called booking models, which historical booking model (HBM), advanced booking model (ABM) and combined booking model (CBM) are three major prototypes.

The first concept to formulate passenger arrivals is HBM, which uses only historical arrival data for model construction. This implementation makes the model a typical time series forecasting problem. Potential methodologies are like moving average, exponential smoothing (Wickham, 1995), autoregressive integrated moving average and artificial neural networks (Weatherford et al., 2003; Tsai et al., 2008). Some other techniques, although they are not further tested, are also suggested such as Baysian forecasting methods and Kalman Filtering by Talluri and Van Ryzin (2004, p.450).

Another distinctive concept applies only reservation data during the booking period and forms the category of ABM. The use of reservation data to predict how many people travel at the time of departure for a specific transportation service is very common in industries. Although this is the case, there are few publications in the literature. Regression and pickup models are two common solutions (Wickham, 1995; Weatherford and Kimes, 2003); they are both easy to understand and straightforward to implement. Nevertheless, these two models only use a limited part of reservation information. Upgrading predictive performance seems to be highly possible if more booking information is considered. This study stems from this point and develops a pattern recognition model (PRM) by using all available reservation data. Schwartz and Hiemstra (1997) have proposed a similar concept for hotel reservations; their model purely focuses
on the use of booking numbers. In fact, incorporating temporal features existed in booking files may benefit model performance.

The integration of the previous two prototypes establishes CBM, which uses both historical arrival and reservation data. Here, the concept is mainly to obtain forecasts by combining outputs from several individual models. Rajopadhye et al. (2001) showed that the combination of historical and advanced booking models can improve predictive performance. In fact, the combination is indeed a powerful tool according to evidences in the literature (Hibon, M. and Evgeniou, 2005; Wong et al., 2007; Jose and Winkler, 2008); nevertheless, individual models establish the base for combining purposes. In other words, developing individual models with acceptable performance is necessary before utilizing the concept of the combination.

In the next section, the definition of booking curves is first presented. Then some general features of booking curves are discovered. A three-stage advanced booking model is proposed to incorporate these observed features. The impact of each designed factor is searched by using an unconstrained optimization algorithm. An empirical study to evaluate the model by using both disaggregated and aggregated railway data is implemented in the following section. Conclusions and feature extensions are provided in the last section.

2. Features of Booking Curves

Before constructing advanced booking models for passenger arrivals, it is essential to discover general features hidden in curves. The difference between arrival and reservation data is first explained via Figure 1 and 2. Figure 1 records only the number of the final sales of a specific service on each departure day; Figure 2 further stores how each number of final sales is achieved during the booking period. Each curve in Figure 2 is called a booking curve; the horizon axis represents days before departure, and the vertical axis indexes the number of accumulated arrivals. Figure 3 displays three typical shapes of booking curves at different demand levels (Lee et al., 2005): passengers who are risk adverse usually book early (broken lines), risk-seeking customers prefer to book late (dash lines), booking steadily over time is the sign for risk-neutral travelers (straight lines).
Figure 1 Historical arrival data

Figure 2 Reservation data
After telling the relationships between arrival and reservation data, three general features of booking curves are observed and introduced below. First of all, although booking curves may have different shapes, they generally show growing trends due to the fact of accumulating. Cancellations, nonetheless, may make booking curves descend, but the overall patterns during the booking period are still going up. This data based property functions like the first step in grey theory and aims to reduce randomness. The second feature of booking curves is that the patterns are fluctuated dynamically. Figure 4 illustrates this characteristic: curves with the similar patterns at the beginning of the booking period may result in the totally different numbers of final arrivals. Figure 5 shows another reverse example: the closed numbers of final arrivals may be obtained even though different booking patterns are observed at the beginning of the reservation period. These two figures inspire an important feature of booking curves: the reliability of information. Thus booking points which are closed to the day of departure should provide more reliable information and receive higher weights than those which are far away from. In other words, information at each booking point should be valued distinctively, and either a linear or nonlinear influence can be considered. Another vital and general feature for short-term forecasting is autocorrelations between the sales numbers. Samples which are close to the day to be forecasted should have higher impacts than those far away from.
This effect can be either a linear or nonlinear function of time order. In addition, periodic repetitions could be another possible phenomenon, and the use of time order to represent the day-of-week influence may be an option. However, the adoption of this term cannot help for picking up informative cases in the preliminary study. The concept is presented here only for completeness.

3. Advanced Booking Models

A booking curve has \( t \) booking points; point 0 represents the day of departure, and point \( t \) is the start of the booking period. In the following, training samples are used to represent data for model construction; testing samples are for days to be predicted. The aim is to predict the number of final arrivals, \( x_{j,0} \), by using reservation information, \( x_{i,k} \), where \( j \) and \( i \) are the indices of testing and training samples, respectively, and \( k \) is the index of booking points. As a result, there are \( t \) advanced booking models, depending on
the availability of reservation information, for each testing sample. For instance, model $t-k$ can utilize partial or all reservation data dated before booking point $t-k$. One of the advantages of advanced booking model is it can dynamically update the forecast when new reservation information becomes available. The proposed pattern recognition model contains three stages: similarity evaluation, sample selection, prediction generation. In the following, the function of each stage is introduced one after another; features discovered in the previous section are integrated into model designs. The proposed model is then turned into an unconstrained minimization problem in order to estimate the influence of each factor; the searching algorithm is also briefed.

3.1 Pattern Recognition Forecasting Model

3.1.1 Similarity Evaluation

Recognizing the similarity (or difference) between training and testing samples is the first stage of the proposed PRM, and only training samples with high similarity to the booking patterns of testing samples are considered to be informative. Here, similarity is defined as a measurement which calculates how close it is for two booking curves in terms of their reservation patterns. Instinctively, Euclidean distance (ED) is a suitable option as listed in Equation (1); however, it might not be robust enough to apply solely ED to calculate similarity between booking curves. Thus, the discovered features are incorporated into the formula. Equation (2) incorporates the concept of the reliability of information; if a booking point is closed to the day of departure, which $m$ is small, then it should have a higher weight. $\alpha$ reflects the significance of the factor. Furthermore, Equation (3) considers the effect of autocorrelations, and its importance is determined by $\beta$. If the date of a training sample is closed to that of the testing sample, which $j-i$ is small, then the training sample should have a higher priority to be selected than others. It should be noted that Equation (3) can be reduced to Equation (2) when $\beta = 0$ and to Equation (1) if $\alpha = 0$ and $\beta = 0$ simultaneously.

\[ D_k(j,i) = \sum_{m=k}^{i} (x_{j,m} - x_{i,m})^2 \]  
(1)

\[ D_k(j,i) = \sum_{m=k}^{i} (x_{j,m} - x_{i,m})^2 \left( \frac{1}{m} \right)^\alpha \]  
(2)

\[ D_k(j,i) = \sum_{m=k}^{i} (x_{j,m} - x_{i,m})^2 \left( \frac{1}{m} \right)^\alpha (j-i)^\beta \]  
(3)
3.1.2 Sample Selection

After evaluating the similarity, the next stage is to select the training samples with high similarity to the testing ones. General questions are then raised as how many samples are enough for obtaining acceptable performance or whether the number of selections is sensitive to predictive accuracy. In the next stage of prediction generation, a weight structure, which is based on the magnitude of similarity, is designed to predict the final sales of the testing samples. In other words, even more samples are selected, their influences would be insignificant due to trivial weights. As a result, this study arbitrarily selects ten training samples with high similarity to each testing sample. The selection mechanism, in this study, is regarded to be insensitive if the evaluation of similarity and the prediction method can both function well.

3.1.3 Prediction Generation

Once informative samples are selected, combining the numbers of arrivals of these selected samples, \( x_{s,i} \) \((s \in i)\), to obtain predictions, \( \hat{x}_{j,t} \), is the last stage of the proposed PRM. Simple average is the most prevalent and straightforward method as shown in Equation (4); each selected sample contributes equally to obtain forecasts. Moreover, as similarity shows the difference between training and testing samples, incorporating this information into Equation (4) should help for improving performance. Equation (5) uses the inverse of similarity as a weight to calculate forecasts; each selected sample now contributes differently to forecasts, depending on how similar it is to the testing sample. The parameter \( \gamma \) is used to represent the influence of this similarity based weight. In addition to the previous two ways for doing predictions, this study also includes an adaptive term, \( \frac{x_{j,k}}{x_{s,k}} \), to incorporate the deviation of reservations between the selected sample \( s \) and the testing sample \( j \) at the current booking point \( k \). However, this adaptive term should be utilized cautiously because it might also have chances to deteriorate performance if its value does not indicate the trend correctly. This phenomenon is also related to the reliability of information, which has been mentioned in the previous section. In other words, the adaptive term should not have significant influences when information is not reliable. Equation (6) shows the complete formula, which \( \delta \) determines the influence of this adaptive term. It should be noted that Equation (6) can be reduced to Equation (5) if \( \delta \to \infty \) and to Equation (4) if \( \delta \to \infty \) and \( \gamma = \sum_s \ln P - (1 - P) \sum_s \ln D_{j,s} \) at the same time.
\[ \hat{x}_{j,1} = \frac{1}{p} \sum_{i=1}^{p} x_{s,i} \]  

(4)

\[ \hat{x}_{j,1} = \frac{1}{\sum_{j=1}^{J} D_{x}(j,s)} \gamma_s x_{s,1} \]  

(5)

\[ \hat{x}_{j,1} = \frac{1}{\sum_{j=1}^{J} D_{y}(j,s)} \gamma_s \left( \frac{x_{j,k}}{x_{s,k}} \right)^{1/\gamma} x_{s,1} \]  

(6)

3.2 Direct Search Algorithm

In the first part of this section, a three-stage pattern recognition model is formulated. However, the model still needs one more step to get the work done: the decisions of \( \alpha \), \( \beta \), \( \gamma \) and \( \delta \). These four parameters reflect the effects of the discovered features, and deciding them by any subjective settings is definitely difficult and unpersuasive. As a result, this study turns the problem into an unconstrained minimization form and applies a direct search algorithm to look for these four parameters. In the following, the problem is restated, and Hooke-Jeeves algorithm is then briefed.

3.2.1 Problem Restatement

After structuring the model, the problem now is to pursue minimum predictive errors given a combination of the parameters. Here, mean square errors (MSE) are used as the objective function, and the model can then be rewritten as Equation (7), in which \( \hat{x}_{j,1} \) represents the outcome of the \( k \)th model. There are two general ways to decide the values of these parameters: the gradient search and the direct search. Although the gradient search can derive more precise directions toward the optima than the direct search, Equation (7) is apparently difficult to get the gradient information and check the convexity of the objective function. The direct search, on the other hand, is comparatively less efficient than the gradient search and cannot guarantee the global solution. However, it can search for parameters without calculating any gradient information. The direct search algorithm is then applied for determining the parameters in this study. For the weakness of inefficiency, many alternative algorithms were proposed to improve searching speed; multi-start strategy using different initial seeds is usually adopted to compensate for the problem of local solutions. In this study, Hooke-Jeeves algorithm is utilized to search for the parameters.
\[ MSE = \sum_{j} \sum_{k} (x_{j,k} - \hat{x}_{j,k}^k)^2 \]

\[ = \sum_{j} \sum_{k} (x_{j,k} - \sum_{m} (x_{j,m} - x_{s,m})^2 \left( \frac{1}{m} \right)^\alpha (j - s)^\beta \sum_{m} (x_{j,m} - x_{s,m})^2 \left( \frac{1}{m} \right)^\alpha (j - s)^\beta )^{-1} (\frac{1}{m} \right)^\gamma x_{s,k})^2 \]  

(7)

3.2.2 Hooke-Jeeves Algorithm

Hooke-Jeeves algorithm, which contains an exploratory search to look for a successful direction toward the optimum, and a pattern search to improve searching efficiency, is applied in this study. A simplified example with two decision variables is used to introduce the searching procedure; readers who are interested in the algorithm can refer to the book by Himmelblau (1972, p.142). Figure 6 demonstrates each move while searching for a solution; Figure 7 briefs the corresponding steps based on the example. Every time when Hooke-Jeeves algorithm moves a decision variable to a new point, the proposed PRM is implemented to reevaluate MSE.

![Figure 6 A simplified example for introducing Hooke-Jeeves algorithm](image)
### Figure 7 Procedure of Hooke-Jeeves algorithm

<table>
<thead>
<tr>
<th>Step 1:</th>
<th>Randomly generate an initial point (Point 1) and also make it a basic point.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 2:</td>
<td>Implement the exploratory search on each axis (broken line is a failed move). If MSE at Point 2 is less than that at point 1, which is the case here, then Point 2 becomes the next basic point.</td>
</tr>
<tr>
<td>Step 3:</td>
<td>Implement the pattern search by using the current and last basic points to move to Point 3 (dot line). Then implement another exploratory search to Point 4. If MSE at Point 4 is less than that at Point 2, then the pattern search is successful; otherwise, failed.</td>
</tr>
<tr>
<td>Step 4:</td>
<td>If the pattern search in Step 3 is successful, then repeat Step 3 until failed. If the pattern search is failed, which is the case here (Point 5), then do not move and go back to Step 2 (Point 4). If the exploratory search in Step 2 cannot reduce MSE, then shrink the step and redo step 2 again (Point 6). Here MSE at Point 6 is lower than that at Point 4, so Point 6 becomes the next basic point.</td>
</tr>
<tr>
<td>Step 5:</td>
<td>If the difference of MSE between two exploratory searches, which only happen when the pattern search fails, is smaller than a predefined number, then the procedure terminates. Otherwise, go back to Step 3. Here, the procedure checks the difference of MSE between Point 4 and Point 6 (and later between Point 8 and Point 10).</td>
</tr>
</tbody>
</table>

### 4. Empirical Study

Reservation data from December 19, 2004, to December 31, 2005, was collected from a railway company, and formed a large database. In this study, one specific origin-destination pair which has prosperous demand and, equally important, is a long haul, is focused. Passengers are liable to reserve seats when the trip is distant because they have to switch to other departures, which are not their original desires, or even other modes when tickets are sold out on the spot. Three departures of the pair are tested; moreover, the aggregate of these departures is also verified. In the following, the patterns of the departures in the real life are graphed. After that, data usage is briefed for the experiments in this study, and final results are reported.

#### 4.1 Practical Booking Curves

More than three hundred booking curves for each departure are extracted from the database. In order to avoid exhausting by showing curves of the chosen departures, only departure 1 is demonstrated and compared with the theoretical curves introduced in the second section. Figure 8 shows three typical booking patterns at two different demand levels; passengers do behave differently according to their acceptance of risks. Figure 9 shows the fact that similar patterns at the beginning of the booking period do not imply
close numbers at the departure. Figure 10, on the contrary, demonstrates another totally opposite phenomenon; similar amounts of arrivals can be obtained even booking patterns are dissimilar. Although booking curves are shown to be contrasting in the previous examples, some other curves do have anticipated patterns. Figure 11 reveals that similar booking behavior results in close number of arrivals at two different levels.

![Figure 8 Practical booking curves at different demand levels](image1)

![Figure 9 A real example of fluctuations](image2)
4.2 Data Usage

Although there are over three hundred curves for each departure; however, not all of them should be used for the estimation of the parameters. Moreover, a certain percentage of samples should be reserved for evaluating accuracy. As a result, the following setting is utilized in this study. Thirty booking curves (about one month) which are close to the present are kept as testing samples for out-of-sample evaluation. Then another thirty booking curves which are dated right before the testing samples are reserved for estimating four parameters in the proposed model; the term of validating samples is used to represent these curves. The rest of booking curves are training samples for the model to implement pattern recognition. The division of the data not only aims to estimate
parameters and evaluate out-of-sample performance, but also maintains time series order. The procedure of data usage is first to apply the training samples as the base to minimize the MSE of the validating samples and obtain the estimation of four parameters. This phrase tries to imitate the step of prediction and relieves the problem of overfitting. Then both the training and validating samples form the base to predict the numbers of final arrivals in the testing samples. Figure 12 summaries the procedure.

Figure 12 The procedure of data usage (1, 2, 3 is the order of division)

Another important setting in the data usage is the use of the multi-start strategy. Direct search methods might find only local minima, and the global minimum is not possible to guarantee. In this study, twenty-five random initial points are tested in the phase of parameter estimation, and the outcome resulting in the smallest MSE is adopted for the stage of prediction.

4.3 Out-of-Sample Performance

In this section, the predictive performance of three individual departures and their aggregate is reported. The proposed PRM model is compared with pick-up and regression models, which are widely applied in practice and taken as benchmarks in this study. As mentioned in the second section, there are $t$ models for each departure in terms of how much reservation information is used; $t$ is fourteen in this study which means passengers can start booking two weeks ago before departure. Figure 13 shows the out-of-sample performance between three models at each booking point. Upper left is the performance of departure 1; the result of departure 2 is put in upper right; lower left shows the outcome of departure 3; the performance of the aggregate is in lower right. It is apparent to see that the proposed PRM outperforms both pick-up and regression models at most booking points in the experiments; the improvement is significant and consistent.

In order to enhance the importance of accuracy at the early stage during the booking period and have an overall index of performance for comparing purposes, this study
weights MSE according to the order of booking points and forms a summarized MSE (SMSE), as shown in Equation (8), in which $m$ is the index of booking points. Figure 14 also, again, verifies the potential of the proposal model from the perspective of SMSE. In addition, Equation (9) calculates how much improvement is achieved by PRM in terms of either pick up or regression models, and Figure 15 shows that PRM can obtain significant percentage of improvement.

$$SMSE = \sum_{m=1}^{t} \frac{m}{t} \times MSE_m$$

(8)

$$\frac{SMSE^{PRM} - SMSE^{base}}{SMSE^{base}}$$

(9)

Figure 13 Performance comparisons at each booking point
4.4 Discussions of the parameters’ values

Four parameters are searched in the proposed model; these parameters decide how the model evaluates similarity, selects samples and does predictions. As a result, it is important to understand their impacts. Figure 16 displays the final results of the parameters. $\alpha$ has the largest value among four parameters; it implies that the value of information at each booking point is distinctive. In addition, autocorrelations are not always influential to all departures; departure 2 does not have the effect because of its
The use of similarity as a weight to implement predictions shows linear impacts, which the estimations of $\gamma$ are all close to unity in four data series. The modified term based on the current booking information is also helpful in all experiments. In order to understand the effect of each parameter at different booking points, further discussions are necessary.

Figure 16 Values of four parameters ($\alpha, \beta, \gamma, \delta$)

In the following, several scenarios are assumed to simulate the influences of $\alpha, \beta, \gamma, \delta$. In Figure 17, the importance of the reliability of information increases as the time approaches to the day of departure. Although the figure shows an exponential trend for four data series, the magnitude is different. For example, the value of old information during the booking period is more informative in departure 3 than that in the other three data series. The effect of the autocorrelations is shown in Figure 18; here the inverse of $(j - i)^\theta$ is used for demonstration. An exponential trend is also found in three of four tested data series. Moreover, departure 3 shows the highest autocorrelation relationships among all. Figure 19 shows that the use of similarity to weigh the selected samples has linear contributions instead of nonlinear ones, and this factor is significant in all tested data. Last but not least, Figure 20 graphs the effects of the adaptive term at each booking point given three scenarios ($\frac{X_{j,k}}{X_{s,k}} = 0.5, 1.0, 1.5$). The adaptive term in departure 2 is actually more conservative in comparison with those in departure 1, departure 3 and aggregate data series.
Figure 17 Influences of $\alpha$ at each booking point

Figure 18 Influences of $\beta$ at each booking point

Figure 19 Influences of $\gamma$ at each booking point
5. Concluding Remarks

This paper starts from the observations of booking curves to address their features. A three-stage model is then proposed to model railway passenger arrivals and compare with conventional pick up and regression models. In addition to solely apply Euclidean distance, the inclusion of the reliability of information and autocorrelations are useful for evaluating similarity between booking curves. Furthermore, both using similarity based weights and the adaptive term increase the predictive capability of the proposed model. Indeed, the model with the combination of solely using Euclidean distance and simple average cannot even compete with regression and pick up models.

There are several possibilities for future extensions. Revalidating the proposed model given different railway services such as expresses versus locals, or even data from other practical fields is one direction. Airlines, car rentals and hotels, in fact, all have the mechanism of advanced sales; their booking curves probably have similar features as railway data. Moreover, in this study, although the values of the parameters are different in respective data series, they are all the same for $t$ models. The second possibility is to discover whether the use of respective parameters for each booking point would benefit predictive performance. However, the complexity of the model would significantly
increase because of doing so. One disadvantage of the proposed model is that it cannot explain how people make choices. However, when choice models are built to explain passengers’ behavior, aggregate demand is still required to distribute demand into individual services. The proposed model also shows potential to predict aggregate demand. As a result, comparing performance between a disaggregate model and the cooperation of aggregate and choice models is another interesting extension.

References