Pick-up method + machine learning: a proved efficient approach to forecast hotel demand

Improving hotel demand forecasting accuracy by integrating machine learning approaches with pick-up methods

1. Introduction

**hotel revenue management is important - demand forecasting is essential**

**why rm is important?**

**why demand is essential?**

* **improving accuracy can increase xx% of revenue**
* **uncertain, can drive pricing xxx**

Hotel revenue consists of price and demand, and demand forecast is essential for successful revenue management since it has high uncertainties and has direct impact on issues such as inventory management, pricing strategies, and marketing plans. Different from industries such as retail where most of the bookings happen instantly when the clients arrive, the hotel industry usually sell the room ahead of the customer arriving, and thus generate valuable advance bookings information.

**introduce advance booking (pick up method)**

Hence, advance booking method, an approach conducting forecast based on current realized bookings and historical booking patterns, is widely used in both academic and the industry. On top of the existed bookings so far (ROH, reservations on hand), the advance booking predicts what will happen from today and the target future by estimating the incremental bookings. In practice, this method takes an average of the incremental bookings in history, or average the incremental ratios in percentage, then add on or multiple to the ROH today. This method is also called “pick-up” method since it estimates the number of incremental bookings “picked up” from today’s reservation.

Pick-up methods are initially used in airline revenue management (L’heureux, 1986), and this concept has been rapidly applied in the hotels due to the highly similarities between the two industries. Generally, there are four types of pick-up models: additive & multiplicative pick-ups, and traditional & advanced pick-ups (Zakhary, Gayar, & Atiya, 2008). From the perspective of the relationship between ROH and final arrivals, Additive Pick-up models conduct prediction by adding an estimated incremental booking to current ROH, while Multiplicative pick-up models multiplied an average ratio on current ROH. On the other hand, pick-up models can also be categorized according to data completion. Traditional pick-up methods only use completed booking curves and ignored records where an arrival day is still beyond “today”. In comparison, advanced pick-up methods use both complete and incomplete booking information. Many researchers have tested the performances of pick-up models in different settings (Weatherford & Kimes, 2003), or combined pick-up models with other methodologies such as quadratic regression (Tse & Poon, 2015), exponential smoothing (Chen & Kachani, 2007), Poisson process (Lee, 2018), etc. However, most of these researches still rely on statistical assumptions to describe the relations between ROH and final arrivals. Given the uncertain demand change with various external factors’ impact, the industry urges for more practice-adapting methods to improve the accuracy.

**machine learning, on the other hand…**

* **widely used in areas…**
* **why? because…**

In recent years, machine learning method has been picking up attentions among both the industry and academia. Machine learning is a statistical method (**add more definition about machine learning**) and it has the benefit of (flexible, catching patterns, etc. )

Even though machine learning has been extensively applied in different areas (business failing prediction, stock price, exchange rate, etc.), it has not been given full attention in the hotel industry. As GlobalData (2017) mentioned, hotel has not been one of the industries considered to be at the forefront of technological innovation. Some researchers have attempted using machine learning in online review analysis (Ma et al, 2018), hotel success indicators (Phillips et al. 2015), and hotel online booking simulation (Corazza 2014), while no existed studies have used machine learning approaches in demand forecasting and revenue management.

However, machine learning approaches are specifically suitable for hotel demand forecasting because of three reasons: firstly, the amount of transaction data in the hotel industry has been surging in the recent decade, which provides a foundation for machine learning models. Secondly, machine learning is able to capture the complex and non-parametric relations between existed bookings and the final sales. Hotel sales is impacted by multiple internal and external factors, and there has not been a clear relationship to capture those patterns. By simulating the arbitrary function from the data itself, machine learning approaches have the potential to better capture the complex relations. Last but not least, machine learning models are capable of dealing with high dimensional data (in the hotel industry’s case, the long days before arrivals). Given the current blank (paucity) of using machine learning in hotel demand forecasting, this study is able to provide insights on the prediction abilities of machine learning in the hotel setting, and make the first step to introduce this powerful approach in hotel revenue management.

1. Methodology
   1. pick-up models

Pick-up models

* 1. brief intro of ml
  2. 6 sections to introduce each algorithm

1. Empirical Study
   1. data description

This research uses a one-year long booking records, with the arrival date from December 27, 2017 to December 31, 2018, of a hotel property. For each booking at this hotel, this dataset records the booking date (the day the client makes this reservation) and the arrival date (the day the client checks in). Hence, for each arrival date, the ROH can be calculated accumulatively from the earliest booking date of each day. For instance, if the hotel is predicting the number of final rooms sold on February 2021 while “today” is January 1, 2021, the realized booking (ROH) can be calculated by adding up all of the reservations for February 14, 2021 happened before January 1. In this way, a new variable, the ROHs is calculated as the main independent variable.

Another derivative variable, days before arrival (DBA), is used to describe the time window between the booking day and the arrival day. This study examines 12 different horizons: 1, 2, 3, 4, 5, 6, 7, 14, 21, 30, 60, 90, and beyond. The cutoffs are made in this way since in the hotel industry, the dynamic pricing plan is usually set following those horizons. It is noticeable that during early periods when the booking day is far away from the stay date, the reservations accumulate very slow. Therefore, a wider horizon in earlier periods allows information to accumulate, while when the stay dates are approaching closer, the booking window is broken into smaller horizons for a closer attention.

The whole dataset has 114,330 observations (370 arrival days \* 310 DBA) and is randomly hold out as 80% in the training set, 20% in the test set. For all the models, this study uses ROHs as the single independent variable on different arrival date and DBA, to make predictions. The dependent variable here is the final accumulated reservations on the arrival day, in other words, the accumulative ROHs when DBA=0.

* 1. methods/models
     1. Pick-up Models
     2. Linear Regression
     3. Machine Learning Models

(use a table to line out the pro & cons, situations, for each ML model)

* 1. results

1. Conclusion & Discussion