

【査読論文】

ホテルの客室予約数の予測

森 竜樹 (武蔵野大学 工学部 講師)

中村 圭一 (株式会社ワールド・ヘリテージ)

東 康平 (武蔵野大学 工学部 助教)

研究協力者：大角 佳瑞 (武蔵野大学 数理工学科 3 年)、豊田 蒼馬 (武蔵野大学 数理工学科 3 年)

要約

ホテル宿泊費の最適化に注目している。最適な宿泊費とは、宿泊日の直前に満室となるように設定される客室販売価格を指す。宿泊費最適化には気象予測、周辺イベント情報、他ホテルの客室在庫数などのデータが重要だが、これらを収集・活用するのは困難である。既存のレベニューマネジメントシステム(RMS)は一部データを活用しているが、全ての情報を網羅したシステムはない。さらに、RMS の結果がブラックボックスであり、現場での納得が得られにくい問題がある。そこで本研究では、客室稼働率予測に古典的時系列解析を用い、説明可能なブッキングカーブを生成する数理モデルの構築を目指す。各チェックイン日ごとに、チェックイン日の 90 日前から当日までの期間をリードタイムとして定義し、(チェックイン日, リードタイム) を時間軸とする 2 次元の時系列データを対象とする。本論文ではこの種の 2 次元の時系列データの扱いに対して、基礎的なアイデアと数理モデルの提案を行う。

1. はじめに

急激な少子高齢化が進む日本において、国内外の交流人口を拡大し、地域の活力を維持・発展させることは日本の持続可能性に関連する重要な課題である。日本の観光庁では、この課題に対して、持続可能な観光地域づくりを行い、インバウンド回復させることは地域の消費額拡大を促し、これが地域活性化の切り札となると位置づけている [1]。消費額を拡大させる方法は多種多様であるが、中でも、旅費の占める割合が大きい宿泊費を対象にし、その適正化を行うことは消費額拡大に対し効果的であると考えられる。これは具体的には各ホテルが宿泊需要予測から、宿泊日当日に満室となるように客室販売価格を時間ご

とに変更する運用を行えばよい。価格最適化は安く部屋を販売してしまうことや、価格が高すぎて部屋が予約されないといった機会損失を未然に防ぐことができるため、消費額拡大に直結しているといえる。このような価格調整の仕組みをダイナミックプライシングと呼ぶ。

客室販売価格のダイナミックプライシングには宿泊当日の気象予測やホテル周辺のイベント情報、周辺ホテルの客室在庫数などの情報が重要であると考えられる。しかし、これらすべての情報を集めて宿泊費を決定することは、ホテルの収益管理を担当するレベニューマネジメント部門の人手不足や他ホテルの情報がオープンになっていないといった理由から困難である。現在、市場には客室販売価格のコントロールの自動化・効率化を行うレベニューマネジメントシステム（RMS）が販売されているが、宿泊当日の気象予測、ホテル周辺のイベント情報や周辺ホテルの客室在庫数といったデータをすべて収集し、運用されるシステムは存在しない（部分的にオンライントラベルエージェントと連携する RMS は存在する）。これらの RMS は過去の客室販売価格、その他情報を入力として、適切な処理を行い、部屋ごとの宿泊予定日までにどれだけ予約が入っているかを表す予想ブッキングカーブと最適な客室販売価格を出力する。しかし、実際の業務では、最終的に出力結果をホテルの支配人が確認して客室を販売するが RMS の処理がブラックボックスであるため、納得を得ることは難しい。

本研究では、客室販売価格の決定における中核要素である客室稼働率の予測に古典的な時系列解析を用いて、説明可能なブッキングカーブを生成する数理モデルの構築に取り組むことにした。特に、本研究は関西でホテルを運営している企業と連携し、客室に関する営業実績データの提供を受けている産学連携型の研究であることが特色である。

ホテル業界における客室販売価格のダイナミックプライシングに関する先行研究について述べる。植竹 - 青木 [2] では SNS の普及とオンライン客室販売の普及に伴い、レピュテーションが重要な指標となることを示唆している。また、田澤 [3] では ReVPAR（1日あたり販売可能客室数あたり客室売上）や GOPPAR（客室数あたりの利益）が重要な指標となることを示唆している。日本国内の研究ではどのような指標が重要視されるかの研究はなされているが、

国内ホテルの収益最適化やブッキングカーブを予測するような研究はなされていない。研究が盛んで無い理由として、ブッキングカーブ予測の困難さがあげられる。通常、時系列解析は時間と値が1 : 1に対応した長い組のデータを対象にする（例えば日付と株価や気温など）。このため様々な場合に大量のデータを用いることができ、正確な予測を行うことが出来ている。しかし、ブッキングカーブ予測では、ホテルの性質上、客が予約をした日からのデータしか存在しないため、使用できるデータはホテルの予約が可能となる日から宿泊当日まで（高々90日分）となる。ここから各日ごとブッキングカーブを予測するため、時系列解析の対象は時間と値は1 : 90のような組のデータとなる。各日のブッキングカーブの予測に使用できるデータは90個しかないので、正確な予測を行うことが難しくなる。このような2次元的な時系列の取り扱いについてはあまり解析がなされていないため、ブッキングカーブを予測するには基本的な時系列解析に加えて別のアイデアが必要となる。海外ホテルの収益予想に関する研究では Phumchusri-Suwatanapongched [4] により、タイにある4つ星ホテルの2016年から2019年までの営業実績データを用いて、昨年同日と同じ予想値（ベースラインモデル）、ホルト・ウィンターズ法（指数平滑化法+季節性+トレンド）、SARIMAモデル（ARモデル+差分+MAモデル+季節性）、TBATSモデル（データ変換+ARIMA誤差+三角関数による季節表現）の4つのモデル予測精度の比較がなされており、SARIMAモデルによる予測が最もよいという結果が得られている。時間と収益という1 : 1対応のデータに対して、解析することで2次元的な時系列の取り扱いをうまく避けている。

2次元的な時系列を持つ時系列解析が必要な業界として航空業界がある。航空業界では限られた飛行機の座席を出発日までに売り切ることように座席価格の変更を行っている。座席価格のダイナミックプライシングのため、ブッキングカーブ予測が必要となるが、ホテル業界と同じ理由で、2次元的な時系列の取り扱いが求められる。Mosaic ATMのWhite Papers [5] ではデータの少なさからくるブッキングカーブ予測の精度の悪さを過去データに基づく予想を組み合わせる手法が紹介されている。Ostaijen-Santos-Mitici [6] ではMarkov連鎖や非定常ポアソン過程で予約の発生プロセスを確率的に扱うことで予測を行っている。

以上より、本稿では説明可能なブッキングカーブ予測のために、深層学習や確率過程を用いず、すでにホテル収益予想の成功例である SARIMA モデルに外生変数要素を追加した SARIMAX モデルをベースにモデル構築を行い、2 次元の時系列データを取り扱うための新しい基礎的なアイデアと予測結果を報告する。

第 2 章では企業から提供される客室に関する営業実績データとその整形方法を説明し、2 次元の時系列データを取り扱うためのアイデアを述べる。第 3 章では構築した数理モデルで得られる予測結果と考察を述べる。4 章ではまとめを述べる。

2. データ整形と 2 次元の時系列データの取り扱い

企業から提供される客室に関する営業実績データは、予約の通知種別、販売先情報、予約受信日、予約番号、宿泊日、室タイプ情報、予約の変更履歴やキャンセル情報のほか宿泊プランの情報などが予約受信時間順に並んでいる。期間は 2024 年 5 月 16 日～2024 年 12 月 31 日までである。時系列解析を行う前処理として、これらの情報から日付ごとのリードタイム（予約日から宿泊当日までの時間のことを表す。今回は 90 日前までとする。以後 LT と書く）、予約された客室合計数、チェックイン日、累積利用客室合計数、曜日という形式に整形する。

整形はデータ削除、LT の計算・補完、累積利用客室合計数の計算、曜日フラグ列の追加の 4 つの手順からなる。

データ削除の手順を述べる。ここではキャンセルが含まれる予約グループの削除と、チェックイン日と予約日に欠損があるデータの削除を行う。また、チェックイン日と予約日に異常値があるデータの修正も行う。ここで異常値とは予約日が宿泊日よりも後に来ているような場合である。例えば、客が午後 12 時直前に予約無しでチェックインを行い、データ入力翌日になってしまったようなケースが該当する。この場合、LT をチェックイン日と予約日の差で計算すると、 -1 となってしまうため、チェックイン日を予約日とする。さらに、予約時の宿泊日数が 2 日以上予約を 1 泊ごとのデータに展開する。ここまでで、各日別の LT が計算可能となる。LT の計算・補完の手順を述べる。チェッ

クイン日から予約日を引くことで LT を計算する。計算後、今回対象外である LT が 90 日を超えるデータを削除する。また、LT はチェックイン日ごとに 0 から 90 まで存在するはずなので、0 から 90 までの範囲で不足している LT の数とそれに対応する予約された客室合計数を 0 で補完する。

累積利用客室合計数は LT ごとに予約された客室合計数を足すことで計算する。曜日フラグは月曜日から日曜日を意味する 7 つの列をデータに追加して、例えば、月曜日であれば 1、そうでないなら 0 といったようにフラグ付けする。

さて、上記手順で得られたデータはある宿泊日に対して、LT ごとの累積利用客室合計数のデータを 90 個持つ 2 次元的なデータとなる。このデータを古典的な時系列解析の枠組みに当てはめるため、日付と特定の LT という 1 : 1 の 1 次元時系列に変換する。したがって、予測したい日のブッキングカーブを得るためには各 LT ごとに 90 個の SARIMAX モデルを構築し、それぞれのモデルを用いて予測したい日の LT の値を計算したものを集計すればよい。なお、外生変数として曜日の情報を与える。

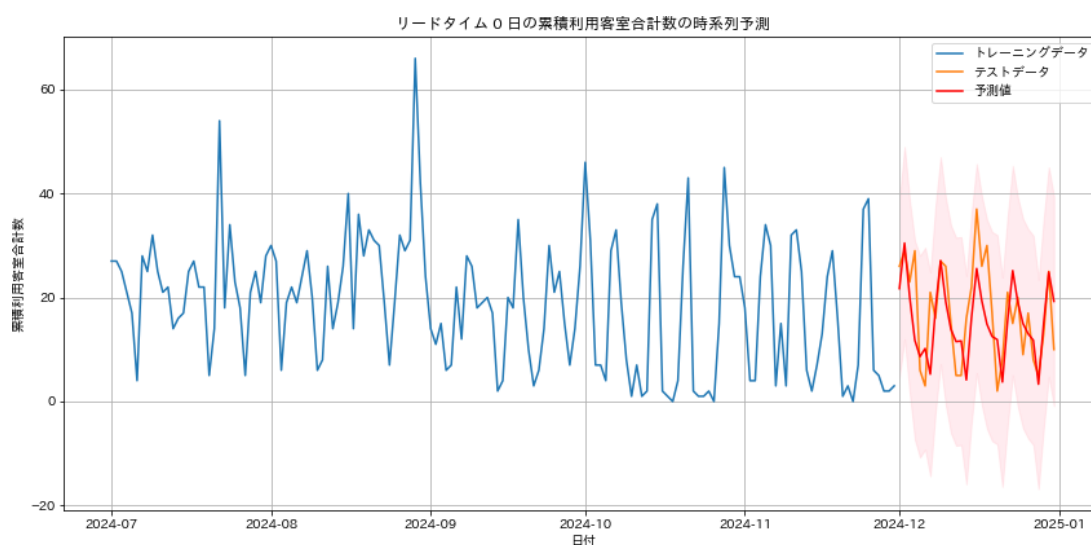


図1:LTが0日の累積利用客室合計数の予測

図 1 は LT が 0 日の累積利用客室合計数の SARIMAX モデルを作り予測を行った例である。図中の青線はトレーニングデータ、赤線は予測値、橙線は実測値を表す。この例では予約情報が十分蓄積されている 2024 年 7 月 1 日から 11 月 30 日までの LT が 0 日のデータをトレーニングデータとし SARIMAX モデルを作成した。モデルには 12 月 1 日から 30 日までの LT が 0 日のデータを予測さ

せている。このように、LT が 0 日、1 日、…、90 日までのモデルを構築し、予測データを集計すれば特定の日のブッキングカーブを予測できる。

3. 予測結果と考察

本章では 2 章で述べたアイデアで作成したモデルによる予測結果を述べる。今回作成したモデルは 2024 年 7 月 1 日から 11 月 30 日までの LT が 0 日から 90 日のデータをトレーニングデータとした SARIMAX モデルである。SARIMAX モデルに含まれる各パラメータは Optuna を用いて平均絶対値誤差 (MAE) が探索範囲では最も小さくなるように調整したものを設定している。探索空間は非季節成分の AR の次数 $p \in [0,7]$, 階差 $d \in [0,5]$, MA の次数 $q \in [0,5]$, 季節成分の AR の次数 $P \in [0,2]$, 階差 $D \in [0,1]$, MA の次数 $Q \in [0,2]$, 周期 $s \in \{7,30\}$ とした。代表例として、12 月 (LT=88) は $(p,d,q) = (7,0,3)$, $(P,D,Q,s) = (0,0,1,7)$ が選択された。作成したモデルでは 2024 年 12 月 1 日から 2024 年 12 月 31 日について各日のブッキングカーブの予測を行った。各日のブッキングカーブは 31 枚あり、紙面の都合上掲載することが難しいので特徴的なブッキングカーブが得られた日の結果を掲載する。

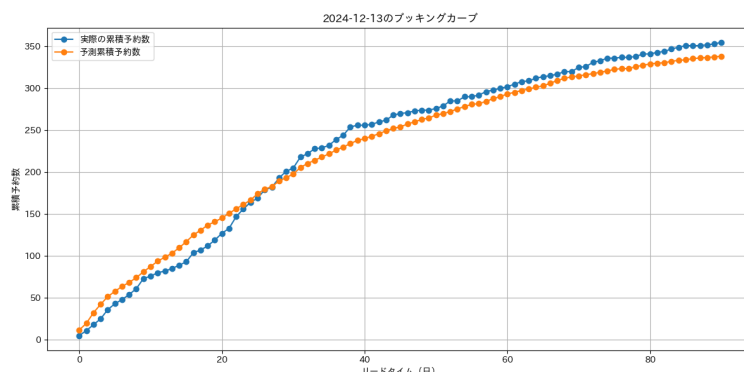


図2: 12月13日のブッキングカーブの実測値(橙線)と予測値(青線)

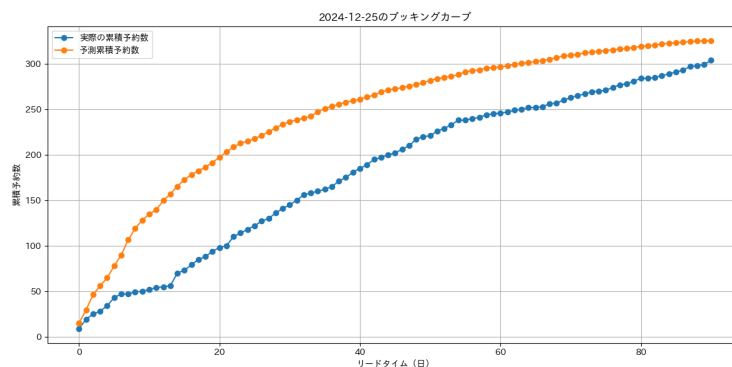


図3: 12月25日のブッキングカーブの実測値(橙線)と予測値(青線)

図2と図3はそれぞれ、MAEが最小値12.61…、最大値62.12…をとった日付のブッキングカーブを描いている。橙線が実測値で、青線が予測値を示している。今回のモデルではLTごとに31日先まで予測しているため12月の末になればなるほど予測が大きく外れると想像されていたが、これらの図から、それよりも各日を持つ様々な外的要因の影響が大きいことが示唆される。

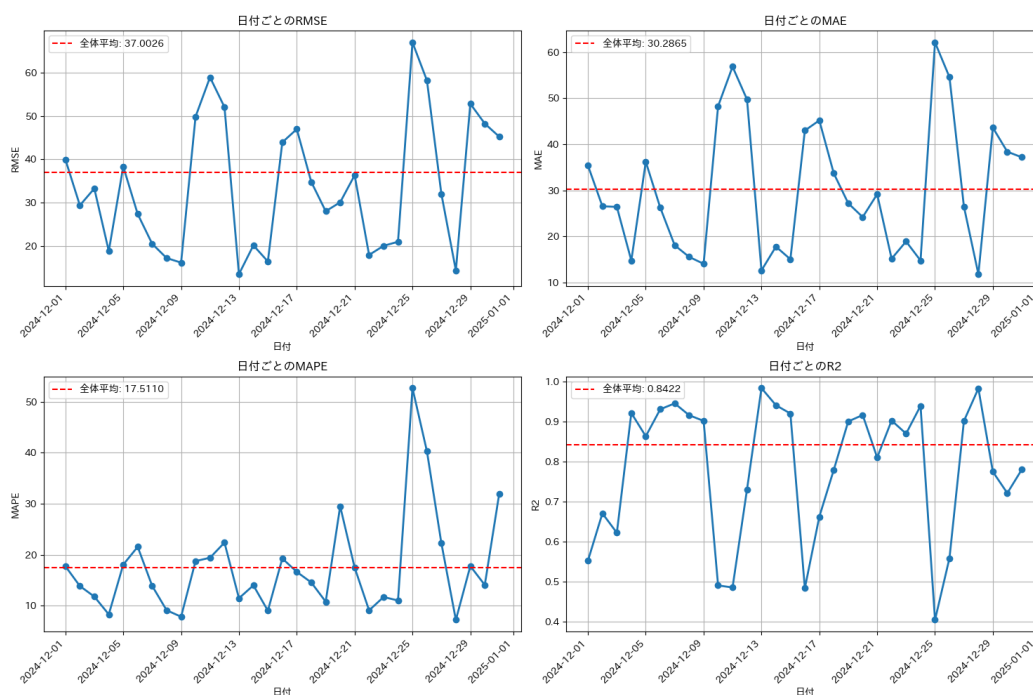


図4: 12月実測値と予測値の誤差解析結果

図4は各日ごとのブッキングカーブの実測値と予測値から得られる、平均二乗誤差 (RMSE)、MAE、平均絶対パーセント誤差 (MAPE)、決定係数 R^2 を描いたものである。横軸は日付で縦軸は各値である。それぞれの平均値はRMSEが37.0026、MAEが30.2865、MAPEが17.5110、 R^2 が0.8422であり、図中の赤色の点線で表している。なお、RMSE、MAE、MAPEの値は0に近いほうが優れており、 R^2 の値は1に近いほうが優れていることに注意する。図4からも特定の日に突出して誤差が大きくなることから各日を持つ様々な外的要因の影響が大きいことが示唆される。

季節性の影響確認のため、5月～9月で学習したモデルを用い、10月と12月の予測成績を比較する検証を実施した。図4, 5が示す通り、10月ではRMSE 30.94、MAPE 20.6%、 R^2 0.858と良好な精度を示したのに対し、12月ではRMSEが37.00に上昇し、 R^2 は0.842と同程度である一方、MAPEは

17.5%へ低下した。これは、年末にかけて宿泊需要の水準が高まるため相対誤差は縮小しやすい一方、祝日や年末商戦期などの一時的要因により日毎の絶対誤差が拡大しやすいと考えることができる。すなわち、本モデルは週次・曜日といった短周期の季節性は概ね再現できているが、年次スケールの季節要因や単発イベントに伴う変動を反映する改良の余地があることが示唆される。

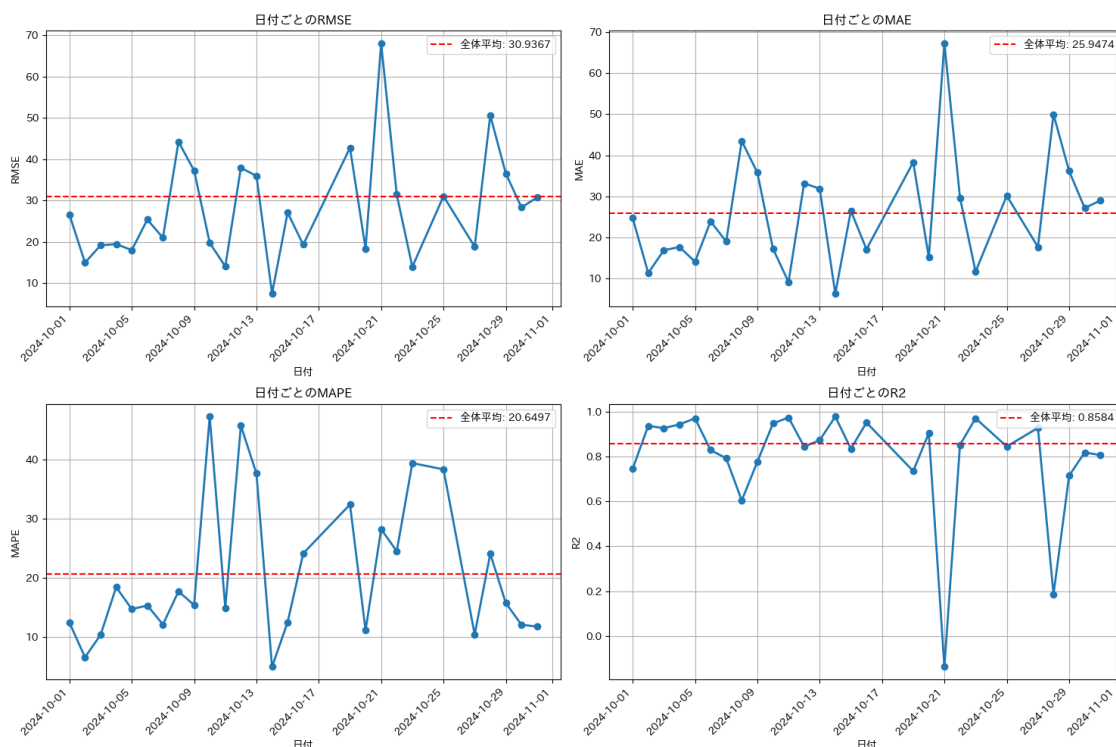


図5: 10月実測値と予測値の誤差解析結果

4. まとめ

本稿では SARIMAX モデルをベースにブッキングカーブ予測のモデル構築を行った。その中で、2 次元の時系列データを取り扱うため 1 次元の時系列データに変換を行う方法の提案を行った。今回得られた結果では、各日を持つ様々な外的要因の影響が大きいことが示唆されており、各日の性質がどのように違うのかを調べ、クラスタリングをした上で提案手法を適用する必要があることがわかった。また、年次データを蓄積し利用することと並行して、曜日以外の祝日の影響や単発イベントの影響を表現する外生変数を追加し、モデルを複数個作成、比較を行うことが求められる。さらに、2 次元の時系列データを取り扱うため 1 次元の時系列データに変換を行う方法はすべての LT を 1 列にかなげるような手法などその他のモデルとの比較を行うことが必要である。

謝辞

本論文は 2024 年度しあわせ研究費（研究テーマ：ホテルの宿泊部屋の稼働率予測のための数理モデルの構築と評価）の助成を受けたものである。

参考文献

- [1]. 観光庁 観光資源課, 観光庁における取組について, https://www.maff.go.jp/j/nousin/kouryu/nouhakusuishin/attach/pdf/suishin_kenkyu-77.pdf, 令和 6 年 3 月.
- [2]. 植竹 朋文, 青木 章通, レベニューマネジメント実施における KPI に関する研究, 専修マネジメント・ジャーナル 8 (1), 41-53, (2018).
- [3]. 田澤彌栄, 宿泊業における収益最適化とその実践に関する考察, 日本国際観光学会論文集 28 巻, 91-100 (2021).
- [4]. N. Phumchusri, P. Suwatanapongched, Forecasting hotel daily room demand with transformed data using time series methods, J. of Revenue and Pricing Management, Vol. 22, 44-56 (2023).
- [5]. Mosaic ATM, Revamping Airline Demand Forecasting, <https://mosaicatm.com/2019/01/22/airline-data-analytics-revamping-airline-demand-forecasting/#:~:text=can%20help%20airlines%20plan%20ahead,to%20make%20the%20final%20forecast>, White Papers, (2019).
- [6]. T. van Oostaijen, B. F. Santos, M. Mitici, Dynamic Airline Booking Forecasting, Conference: ATRS 2017 - Air Transport Research Society World Conference, (2017).