

感情分析を活用したホテル価格最適化モデルの提案

信州大学工学部 電子情報システム工学科

田中・エルナン研究室 21T2803A 石渡 将伍

1. まえがき

近年、観光業界では旅行需要の多様化や市場競争の激化に伴い、宿泊施設が顧客ニーズに迅速かつ柔軟に対応することが求められている。その中で、ダイナミックプライシングは、客室の稼働率を最大化しつつ収益を最適化する重要な手法として注目されている。一方で、現在の多くのモデルは過去の価格や稼働率データに基づく定量的なアプローチが主流であり、顧客の主観的な評価や感情を反映することは困難とされている。本研究では、長野市内のホテルと共同で、顧客の口コミデータを活用した新しい価格最適化モデルを構築することを目指す。

2. 研究概要

本研究では、長野市内のホテルから提供された顧客データに加え、ホテルの口コミ文を感情分析によって評価した結果を活用し、解析を行う。その後、ロジスティック回帰を用いて特定の日における部屋の稼働率を予測するモデルを構築し、モデルの信頼性および実用性を検証する。

2.1 BERT モデル

本研究では、口コミデータを特徴量として活用するために、自然言語処理技術である BERT[1]を使用して感情分析を実施した。本モデルは、文をポジティブ、ネガティブ、中立の3クラスに分類する。これにより、顧客の主観的な評価を定量的にスコア化し、価格最適化モデルの入力特徴量として活用することが可能となる。BERTを使用した根拠として、双方向的な文脈情報を考慮した高精度な自然言語理解が挙げられる。

2.2 ロジスティック回帰

口コミ感情スコアやその他の特徴量を用いて、特定の日における部屋の稼働率を予測するために、ロジスティック回帰モデルを構築した。ロジスティック回帰は、稼働率が「稼働(1)」または「非稼働(0)」の二値分類問題であることから、適切な手法として選定された。このモデルでは、特徴量 X_1, X_2, \dots, X_n を基に、部屋が稼働する確率 $P(Y = 1|X)$ を計算する。ロジスティック回帰の回帰式を式1に示す。

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (1)$$

3. 研究方法

3.1 感情分析

本研究では、ホテルの口コミ文に特化した学習済みモデルと事前学習済み日本語モデルを組み合わせ、BERTにファインチューニングし、研究対象のデータに最適化を行った。感情分析の結果は、0~1の範囲に正規化されたスコアとして出力した。このスコアを毎月ごとに平均値として集約し、モデルの入力特徴量として活用する。

3.2 特徴量の選定

モデルの精度と信頼性を向上させるため、特徴量の精査を行った。選定した特徴量には、対象ホテル(以下、ホテルA)の価格、曜日のダミー変数、連続変数に加え、長野市内で稼働率が最も高いホテル(以下、ホテルB)の価格や口コミスコアも含めた。特徴量の選定基準として、特徴量間の相関および目的変数(部屋の稼働率)との相関を確認した。相関が高い特徴量は直接的に採用し、相関が低い場合でも、他の特徴量と相関があり、モデルの説明力向上に寄与すると判断される場合には選定対象とした。さらに、全探索による特徴量の組み合わせを評価し、精度順にソートすることで頻出する特徴量を分析した。この方法により、目的変数に対して一貫して有用な特徴量を抽出し、不要な変数を削除することで、モデルの簡略化と説明力の向上を実現した。

3.3 モデルの構築

ロジスティック回帰を用いて、部屋の稼働率を予測する二値分類モデルを構築した。このモデルでは、部屋が「稼働(1)」または「非稼働(0)」であるかを分類する。

3.4 モデルの評価

モデルの性能評価には、正解率(Accuracy)、混同行列、分類レポート(Precision, recall, F1スコア)を用いた。モデルの学習進捗に応じて、初期段階・中間段階・最終段階で評価を行い、モデルの精度向上と過学習の抑制を検証した。

4. 研究結果

4.1 特徴量

目的変数(部屋の稼働率)および各特徴量間の相関分析を行い、モデルの精度向上に寄与する特徴量を絞り込んだ。目的変数および各特徴量間の相関係数を表1に示す。

表 1: 各特徴量間および目的変数との相関係数

	Y	X1	X2	X3	X4	X5	X6
Y	1	0.12	-0.1	0.29	0.19	0.02	0.02
X1	0.12	1	0.1	0.21	0.43	0.02	0.01
X2	-0.1	0.1	1	0.14	0.27	0.01	-0.1
X3	0.29	0.21	0.14	1	0.41	0.03	0.02
X4	0.19	0.43	0.27	0.41	1	0.02	0.01
X5	0.02	0.01	0.03	0.02	0.01	1	-0.5
X6	0.02	0.01	-0.1	0.02	0.01	-0.5	1

注:

Y「目的変数」

X1「ホテル A 価格」

X2「ホテル B 価格」

X3「曜日(連続変数)」

X4「土曜日(ダミー変数)」

X5「感想スコア」

X6「評価数」

表 1 から、ホテル A 価格は土曜日との相関が特に強く、ホテル A では土曜日の価格を重視していることが示唆される。また、目的変数と曜日との相関が認められ、週末に近づくほど稼働率上昇する傾向があることが分かった。一方、感想スコアは評価数と負の相関が見られ、評価数が少ないほど感想スコアが高くなる傾向が確認された。ただし、目的変数との相関は低いため、本段階では重要な特徴量として位置付けることが難しい。次に、全探索による高精度になる特徴量の組み合わせパターンを分析した。内上位 50 パターンの 頻出特徴量を表 2 に示す。

表 2: 全探索による特徴量の頻出回数

ホテル A 価格	ホテル B 価格	感想スコア	曜日	土曜日
37	6	42	49	18

表 2 より相関の見られなかった感想スコアは一貫して目的変数の予測に寄与していることがわかる。だが、採用理由が明確ではないため、さらに分類レポートを基に評価を行った。

4.2 モデルの評価

特徴量選定の結果、最も精度が高く解釈性の高い、「ホテル A 価格」「感想スコア」「曜日(連続変数)」「土曜日」を説明変数として解析対象に選定した。感想スコアの必要性を評価するため、p 値、と回帰係数(coef)の分析を行った。その結果を表 3 に示す。

表 3: 特徴量ごとの p 値および回帰係数

	ホテル A 価格	感想スコア	曜日	土曜日
P 値	0.242	0.021	0.001	0.193
COEF	0.1993	-1.9894	0.283	0.4483

表 3 の結果から、感想スコア、曜日(連続変数)、土曜日の p 値は良好であり、モデルへの寄与が示唆された。特に曜日は、週末に近づくほど稼働率が上がる

傾向が見られる。また、土曜日は p 値が高いものの、係数が大きく、稼働率の重要な要因であることが確認された。一方、感想スコアの係数は約-2 であり、スコアが低いほど稼働率が上昇することを示している。この結果は、評価数が少ない場合に感想スコアが高くなるという特徴量分析の結果とも一致しており、感想スコアは間接的に稼働率に影響を与えている可能性が高いと判断できる。

次に、ロジスティック回帰モデルの分類レポートを表 4、表 5 に示す。

表 4: 分類レポート(非稼働)

	PRESICION	RECALL	F1	決定係数
学習前	0.48	0.55	0.49	0.03
学習後	0.75	0.66	0.70	0.23

表 5: 分類レポート(稼働)

	PRESICION	RECALL	F1	決定係数
学習前	0.51	0.56	0.52	0.03
学習後	0.62	0.69	0.66	0.23

表 4 および表 5 から、分類モデルのスコアは実務上十分な性能(0.6 以上)を示していることがわかった。ただし、表 4 の Precision がやや低いため、「稼働」の予測精度向上が課題として挙げられる。決定係数は 0.23 であり、ダイナミックプライシングモデルのようなノイズの多い環境下で有用な値であると判断できる。

5. まとめ

本研究では、感想スコアをダイナミックプライシングモデルに応用することを主目的として、ホテルの価格設定および稼働率に寄与する要因を分析した。その結果、感想スコアは直接的な相関が低いものの、評価数との関係性を通じて間接的に稼働率へ影響を与えていることがわかった。さらに、分析により、感想スコアを含む特徴量の選定が稼働率の予測精度の向上に効果をもたらすことが確認された。特に、感想スコアの係数分析から、スコアが低い場合に稼働率が高くなるという逆相関の傾向が示され、この結果は評価数の影響を反映していると考えられる。本研究により、感想スコアと稼働率の関係性を踏まえた価格戦略を構築することで、ホテルのブランド価値向上や収益最大化につなげられる可能性が示唆された。今後は本研究で得られた知見を基に、さらに高度なモデルを構築し、感想スコアをより効果的に活用した価格戦略の実現を目指す。

参考文献

[1] Nishikawa, H. "Fine-tuning Japanese BERT: Part 1." Jupyter Book, https://jupyterbook.hnishi.com/language-models/fine_tune_jp_bert_part01.html