

ゴルフ場予約サイトにおけるゴルフ場の特性を 加味したキャンセル要因の分析

西條 直哉[†] 大竹 恒平[†] 生田目 崇[†]

[†]中央大学理工学部経営システム工学科

1. はじめに

近年、インターネットの普及により EC (Electronic Commerce) 市場は急激に拡大している。EC サイトが購買チャネルとして確立する一方で、ホテルやゴルフ場といったホスピタリティ業界の予約サイトにおいては、ユーザのキャンセルによる機会損失の削減が重要な課題となっている。しかし消費者のキャンセル時の行動に着目したキャンセル要因の特定は十分に行われていない。

本研究では、ゴルフ場予約サイトを対象に、ゴルフ場の特性を加味したキャンセル要因の特定を目的とする。

2. 使用するデータ

本研究では、ゴルフ場予約サイトから提供いただいたデータを用い、キャンセル要因の分析を行う。具体的には、ゴルフ場予約データ、会員データ、ゴルフ場コーザルデータ、及びゴルフ場レビューデータを用いる。ただし、以下の条件に当てはまるデータを分析対象とする。

- ゴルフ場予約データ
 - プレー日が 2014/04/01~2015/03/31
 - プレー開始時間帯が 7 時台~10 時台
- 会員データ
 - 年齢が 20 代~70 代
- ゴルフ場コーザルデータ
 - 所在地が関東地方のゴルフ場
- ゴルフ場レビューデータ
 - ゴルフ場予約データの期間内に 1 人以上プレーしているゴルフ場

3. ゴルフ場特性によるゴルフ場の類型化

はじめに、ゴルフ場が有する特性によりユーザのキャンセル行動が異なるという考えの下、ゴルフ場コーザルデータ及びゴルフ場レビューデータを用いたゴルフ場の類型化を試みた。類型化には K-means 法を用いた。評価関数 ϕ を式

(1) と定義する。

$$\phi = \sum_{x_j \in X} \min_{i \in k} \|x_j - c_i\|^2 \quad (1)$$

ここで、 $X_j, j \in \{1, \dots, n\}$ は各項目、 n はケースの総数を示す。また、 c_i はクラス $i \{i = 1, \dots, k\}$ の中心である [1]。評価関数 ϕ を最小化する c_i を求めることによって、データ X を任意の 4 個のクラスに分割した。

K-means 法により類型化した 4 つのクラスの特徴を以下に示す。

表 1 各クラスの特徴

クラス	ゴルフ場数	ゴルフ場の特徴
1	32	・ 收容組数が少ない ・ 丘陵コースでない ・ レビュー得点のばらつきが大きい
2	25	・ 收容組数が多い ・ 価格が低い ・ 宿泊施設がある
3	33	・ 価格が高い ・ セルフプレーなし ・ レビュー得点が高い
4	36	・ 收容組数が多い ・ 価格が低い ・ レビュー数多い

表 1 の各クラスの特徴より、各クラスにおけるゴルフ場特性を以下のように定義した。

表 2 各クラスのゴルフ場特性

クラス	ゴルフ場特性
1	個性的なコース
2	宿泊できる大衆向け
3	高級で人気
4	宿泊できない大衆向け

4. ゴルフ場特性ごとのキャンセル要因の特定

次に、ゴルフ場の特性ごとにキャンセルの要因を特定するため、それぞれのクラスごとにゴルフ場予約データ及び会員データを用いた二

Analysis of Cancellation Factors Based on the Characteristics of Golf Courses in Reservation Sites
[†] Naoya Saijo, Kohei Otake and Takashi Namatame
 Faculty of Science and Engineering, Chuo University

項ロジスティック回帰分析を行った。回帰係数を β_k ，説明変数を x_k とした際のキャンセルが発生する確率 p は式(2)で示される[2]。

$$p = \frac{\exp\{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k\}}{1 + \exp\{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k\}} \quad (2)$$

上式の p を

$$p = \begin{cases} 1 \dots \text{キャンセル} \\ 0 \dots \text{プレー完了} \end{cases}$$

としてパラメータ β_k を推定した。なお各モデルではステップワイズ法を用いて変数選択を行った。各クラスタのモデルの特徴的な変数を表3に示す。表3の特徴的な変数における(+)は正の回帰係数，(-)は負の回帰係数を表す。

表3 各クラスタの特徴的な変数

クラスタ	ゴルフ場特性	特徴的な変数
共通	—	・プレー回数(+) ・予約日からプレー日の日数(+)
1	個性的なコース	・スコア管理利用(-) ・プレー開始8時台(+) ・休日(+)
2	宿泊できる大衆向け	・70代(-) ・休日(+) ・ダイレクトメール(-) ・会員になってからの日数(+)
3	高級で人気	・ハイシーズン(-) ・性別(-) ・休日(+)
4	宿泊できない大衆向け	・プレー金額(+) ・性別(-) ・ハイシーズン(-)

クラスタ共通のキャンセル要因としては、頻繁に予約を利用している人、早期に予約する人というのが挙げられる。予約サイトの利用に慣れている人ほどキャンセルへの抵抗が小さくなっているのではないかと推察できる。

個性的なコースのゴルフ場のキャンセル要因に、スコア管理を使っていない人、プレー開始時間が早い予約が挙げられる。個性的なコースのゴルフ場のため、コースよりも時間などの条件を重視する人がキャンセルしやすいのではないかと推察できる。

宿泊できる大衆向けゴルフ場のキャンセル要因に、低世代が挙げられる。仕事などの急な予

定が入りやすい若年世代がキャンセルしやすく、急な予定が入りにくい退職した60代、70代がキャンセルしにくいのではないかと推察できる。

高級で人気のゴルフ場のキャンセル要因に、女性、ハイシーズンでない予約が挙げられる。人気があるゴルフ場のため予約が埋まりやすいハイシーズンはキャンセルされにくいのではないかと推察できる。

宿泊できない大衆向けゴルフ場のキャンセル要因に、プレー金額が高い予約が挙げられる。大衆向けのゴルフ場のため金額を少しでも抑えたい人が多いのではないかと推察できる。

また、各モデルの精度を確認するために、10分割交差検証法を用いてモデルの精度推定を行った。各モデルの精度を以下に示す。

表4 各モデルの精度

クラスタ	1	2	3	4
モデルの精度(%)	61.5	63.2	61.9	63.8

5. まとめ及び今後の課題

本研究では、まずゴルフ場の特性によりキャンセル要因が異なると考え、ゴルフ場コーザルデータ及びゴルフ場レビューデータを用いてゴルフ場の類型化を行った。その結果、4つのクラスタに類型化し、各クラスタのゴルフ場特性を定義した。次に、類型化したクラスタごとにキャンセルの要因を特定するため、ゴルフ場予約データ及び会員データを用いて二項ロジスティック回帰分析を行った。これらの分析より、各ゴルフ場共通のキャンセル要因とゴルフ場が有する特性ごとのキャンセル要因の特定をすることができた。

しかし、本研究の二項ロジスティック回帰モデルの精度は、どのクラスタにおいても61%~64%ほどであり、あまり良いとは言えない。そのためモデルに使用する説明変数を工夫する、他の分析手法を用いるなどより良いモデルを構築することにより、より精度の高いキャンセルの要因を特定することが今後の課題である。

参考文献

- [1]加藤直樹，羽室行信，矢田勝俊：「データマイニングとその応用」，朝倉書店（2008）
- [2]丹後俊郎，山岡和枝，高木晴良：「ロジスティック回帰分析」，朝倉書店（1996）