

J1リーグの観客動員数予測モデルの構築と影響要因分析

1X18C008-6 上杉和也
指導教員 大野高裕

1. 研究背景と目的

プロスポーツの経営にとって、観客からの収入は大きな収益源であり、観客数を増やすことが最も重要な目標であることは言うまでもない。日本国内のサッカーリーグであるJ1リーグでは、年間で1チームがホームとアウェイで計34試合を行う。しかし、スタジアム所在地・歴史・順位に関わらず、観客数は試合日ごとに大きなばらつきがある[1]。また、2016年のチーム平均収容率は58.4%であり、4割以上の座席が空席となっている計算になる[2]。このことから、各チームが観客数増加のための施策を行うべき状態であるにもかかわらず、観客数にばらつきがあるため、どの試合日にどのような施策・イベントを行うことが有効なのかはわからない現状がある。

予測については、従来用いられていた重回帰分析に代わり、新日鉄住金ソリューションズがAIを活用した年間入場者数を予測する技術を開発し、2016年シーズンの全チーム合計の年間入場者数を誤差率約0.5%と高い精度で予測することに成功した。しかし各チームの1試合ごとの実際の観客数は予測と大きくずれていたため、AIを導入してもなお、試合毎の観客数予測は難しいことが理解できる[2]。そのため、各チームの施策に観客数予測を生かすためにも、1試合ごとの観客数予測の必要性がある。しかし、予測を行うだけでは、どのような影響によって観客がスタジアムに足を運ぶのかを明らかにできないため、的確な施策につながりづらい。

これらの背景から本研究では、従来研究で考察されていない観客数を決定する影響要因を見つけ、重回帰分析を用いることによって要因の影響を明らかにすると同時に、機械学習を用いることで、試合ごとの観客数に対して予測モデルを構築し、この2つを組み合わせ、各チームの集客プロモーションの基礎となることを目的とする。

2. 従来研究

河合[3]らは、93年～05年のJ1の試合の入場者数について重回帰分析を行った。この結果、観客数の50.8%を説明する事ができた。しかし、データが13年分と期間が長く以前のものであるため、経済状況・視聴媒体の変化により、観客を説明する変数が現在と異なる可能性がある。

Jaume García[4]らは、スペインリーグでの4年間の試合を同様に分析し、観客数の50.5%を説明した。人気クラブであるバルセロナとレアルマドリードの試合では観客数が増加する一方で、入賞の可能性が消えた消化試合・最終節などでは減少することを明らかにした。

辻[5]らは、2013年の1年間のJ1、J2の試合の観客数に対し、平均集客率の値で2群に分けて分析を行った結果、J2チームを多く含む高集客率群では、観客数の83%を説明することができた。さらに、J1チームを多く含む低集客率群では、地元出身選手数が大きな影響を与えることも明らかになった。しかし単年度を対象にした分析かつ

J2チームもデータに含んでいるため、他の年度のJ1予測に適用することができるかが明らかではない。

3. 研究方法

3.1 研究手順

本研究では図1の研究概要図に基づき研究を行う。分析にはIBM SPSS Statistics ver.27を使用する。

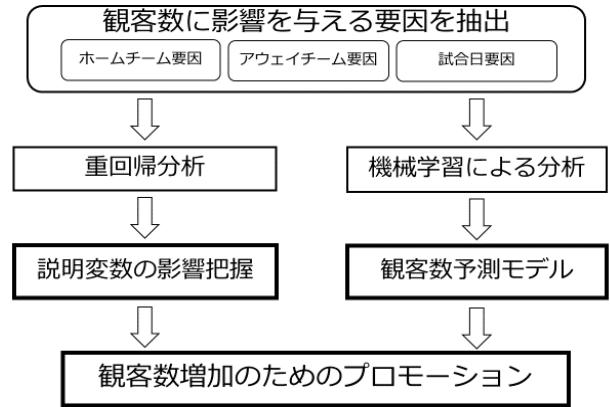


図1 研究概要図

3.1.1 観客数に影響を与える要因の抽出

観客の大多数がホームチーム・アウェイチームのどちらかのファンであるため、観客数に影響を与える要因として「ホームチーム要因」「アウェイチーム要因」を設定する。しかし、各チームの影響だけではなく、試合日当日の様々な要因が観客数に影響を与えることが考えられるため、「試合日要因」を設定する。さらに、それぞれを中要因・小要因と細分化し、その中で観客数に影響を与える要素を抽出し、それらに該当する変数を設定した。表1は、今回の分析に用いる影響要因の変数である。

表1 観客数に影響を与える要因

| 大要因 | 中要因 | 小要因 | 従来研究で考察された要素(変数) | 本研究の新たな追加要素(変数) |
|-----------|--------|--|---|--|
| ホームチーム要因 | チームの魅力 | 強さ | 現状の強さ(順位) 日本人実力選手(代表選手数) | 過去の強さ(タイトル数) 期待度(前年の移籍者数) 現状の勢い(連勝数) |
| | | 選手 | 選手の人気度(推定年俸) | |
| | 愛着 | | 生え抜き選手への愛(ユース出身選手数) | 歴史(創立年数) |
| | | | J1の喜び(昇格年) | |
| | 規模 | | 前年度チーム人気(前年度平均観客数) 最大観客数(スタジアム収容数) チーム規模(昨年度売上高) | |
| プロ野球 | | 試合盛り(プロ野球チームとの距離) 試合盛り(プロ野球の試合の有無) | | |
| アウェイチーム要因 | 経済 | | ホームタウンの金銭的余裕(所得) | 観客の金銭的余裕(お小遣い) |
| | 人口 | | 潜在観客数(人口) | |
| | チームの魅力 | 強さ | 現状の強さ(順位) 日本人実力選手(代表選手数) | 過去の強さ(タイトル数) |
| 試合日要因 | 選手 | 選手 | 選手の人気度(推定年俸) | イニエスタ・ピジャ(神戸戦) |
| | | | 前年度チーム人気(前年度平均観客数) | |
| | 移動 | | 観客の移動の大きさ(移動距離) | 特別な移動(地形特性) |
| | | | 観客の忙しさ(平日) 試合時間(キックオフ時間) 代表戦への注目(代表戦前) 代表選での注目(代表選後) | イベント(ユニフォーム配布) 夏休み(8月) |
| 注目試合 | | 入学シーズン(4月) 運動会・体育祭(10月) ライバルとの試合(ダービー) | 期待(ホーム開幕戦) シーズンまとめ(ホーム最終節) | |

3.1.2 重回帰分析による要因の検討

本研究では階層的重回帰分析により、従来研究で影響が確認されている変数を強制投入法で、本研究で加えた

変数をステップワイズ法で投入することで、新しい変数の導入によりモデルの説明力が向上したか、それらの変数が観客数にどのように影響したかを検討する。

3.1.3 機械学習による予測モデルの構築

ニューラルネットワーク分析における、パーセプトロン法を用いて、観客数予測モデルを構築する。分析結果の重要度が低い変数を1つずつ減らしていき、テストデータの相対誤差が最も低いものをモデルとする。また、「収容率」「チームの歴史」「試合日」「順位」によりデータを2群に分けて、より精度の高いモデルを構築する。

3.2 分析データ

本研究では、J1リーグの公式戦計918試合を分析対象とし、2017-2019年の3シーズンのデータを用いる。各変数に使用したデータは、Jリーグデータサイト、サッカーダイジェスト選手名鑑、Jリーグスタジアム観戦調査サマリーレポート、JFA、e-stat 政府統計窓口から抽出した。

4. 研究結果

4.1 重回帰分析

4.1.1 変数の絞り込み

ステップワイズ法により、重回帰分析を行った結果、ホームチームのタイトル数・お小遣い・前年の移籍者数が有意にならなかったため、これらの3個の変数を削除し、残りの9個の変数を新たな影響要因として分析に追加した。

4.1.2 変数の影響度

上記の9個の変数を追加した、33個の変数を説明変数として重回帰分析を行った結果が以下の図2である。モデルの説明力を示す自由度調整済み決定係数である R^2 は、0.666となり観客数の66.6%を説明することができ、複数年のデータを用いて分析を行った先行研究と比較しても、説明力の向上が認められる。

| 有意○ 正の影響 | | 有意× 正の影響 | |
|----------|-------|----------|--|
| 昇格年 | 連勝数 | 代表選手数H | |
| 前年度観客数H | タイトルA | プロ野球距離 | |
| スタジアム収容数 | 神戸戦 | プロ野球存在 | |
| 昨年度売上高 | 地形特性 | 人口 | |
| 年俵A | ユニ配布 | 代表選手数A | |
| 前年度観客数A | 8月 | 代表戦後 | |
| ダービー | 開幕戦 | 代表戦後 | |
| | 最終節 | | |

| 有意○ 負の影響 | | 有意× 負の影響 | |
|----------|------|----------|--|
| 順位H | 順位A | 年俵H | |
| 創立年数 | 移動距離 | ユース出身選手数 | |
| キックオフ時間 | 平日 | 所得 | |
| 4月 | | 10月 | |

図2 重回帰分析の結果

4.2 機械学習による予測モデル

図2の33個の変数からニューラルネットワークを用いて予測モデルの構築を行った。重要度が低い変数を削除していき逐次分析を行った結果、変数が22個の時に最も予測精度が高くなり、学習データとの相対誤差が23%、テストデータとの相対誤差が28.9%となった。データを2群に分けて分析を行った結果、「収容率」でデータを2つに分けた際の収容率が高い群に対するモデルで、相対誤差が25.0%となり予測精度が最も高くなった一方で、収容率が低い群では相対誤差が36.7%となり、精度が低くなった。

5. 考察

アウェイチームの魅力要因である年俵・神戸戦・タイトルについて正の影響が有意に出たこと、ホームチームの魅力要因である年俵・タイトル・前年の移籍者数が有意にならなかったことが確認できたため、自チームの魅力よりも対戦相手に魅力があることで、観客数が増加すると考えられる。さらに、チーム魅力の中で、強さ・愛着よりも人気選手の要因が観客数を増加させることもわかった。また、移動距離で負の影響が出たことに対して、北海道・九州への移動を伴う試合をダミー変数で代った地形特性で正の影響が出たことから、遠距離移動は観客数が減少するが、旅行になるような長距離移動は観客数を増加させることが考えられる。さらに、従来変数の影響についての正負の別は、Jリーグの従来研究と同様であったため、視聴媒体や社会環境の変化が起こっているにも関わらず、観客数に影響する要因が変化していないことが明らかになった。

機械学習による分析では、前年度観客数、スタジアム収容数の影響が大きかった点、収容率で2群に分けた結果、予測精度の向上が見られた点から、スタジアム規模が予測において重要であると考えられる。一方、収容率の低い群における予測精度が低かった理由として、これらの群では、試合日ごとの観客数にばらつきが生じており、学習データとテストデータのずれが原因で十分な説明力が得られなかったと考えられる。

6. 結論と今後の課題

本研究では、重回帰分析でJ1リーグの観客数に影響を与える要因を明らかにし、機械学習で試合毎の観客数予測モデルを構築した。その結果、チームの魅力の中でもアウェイチームの魅力が観客数を増加させることを明らかにできたと同時に、ニューラルネットワーク法を用いて、従来研究より精度の高い予測を行うことができた。

今後の課題として、本研究では1試合の観客数全てを目的変数として分析を行ったが、座席の種類ごとの分析、年齢層・同伴者などに着目しての分析を行うことで、よりプロモーションに生かすことができる結果を得ることが望める点、機械学習による予測モデル構築でニューラルネットワークのみを用いたが、精度向上のためランダムフォレスト等他の手法の検討も必要である点が指摘できる。

参考文献

- [1] Jリーグデータサイト <https://data.j-league.or.jp/SFTD01/> (最終閲覧日:2021/12/28)
- [2] ITメディア・オンライン Jリーグの年間入場者数、AIで予測・観客増に生かす <https://www.itmedia.co.jp/business/articles/1702/17/news125.html> (最終閲覧日:2021/12/21)
- [3] 河合慎祐・平田竹男：“Jリーグの観客数に影響を与える要因に関する研究” スポーツ産業学会, Vol. 18, No. 2, pp.11-19 (2008)
- [4] JaumeGarcía Plácido Rodríguez “The determinants of football match attendance revisited” Journal of Sports Economics, Vol. 3, No. 1, pp.18-36 (2002)
- [5] 辻和真・二宮浩彰：“Jリーグのスタジアム集客率からみた入場者数の決定要因”，スポーツ産業学会, Vol. 26, No. 1, pp. 73-91 (2016)