

サッカーにおける戦略策定を目的としたデータ分析

下村 勘太(21X4029) 林 雅哉(21X4118) 指導教員 劉 慶豊

1. はじめに

J リーグが創設されてから 30 年以上が経過し、サッカーは日本においてメジャーなスポーツとなりつつある。2022 年に行われたカタールワールドカップにおいて、ヨーロッパの強豪であるドイツ、スペインを破る快進撃を遂げたことは記憶に新しい。そんなサッカーでよく用いられる言葉が「スタッツ」である。スタッツとは、statistics の略で、スポーツにおけるチームや個人のプレー成績をまとめたものである。ボール支配率、シュート数、ファウル数、フリーキック数、コーナーキック数などが例として挙げられる。本稿では、このスタッツに関して分析を行い、どのような要因が強いチーム作りに寄与しているのかを解明していく。本研究では、膨大な試合のスタッツに機械学習の手法を適用して考察する。また、分析を通して得られた知見を特定チームの戦略へいかに応用ができるのか考察していく。本稿では、近年の J リーグで好成績を収めている川崎フロンターレというチームを例に用いて、戦略の提案を行っていく。

2. 用いた機械学習の手法

2.1 ニューラルネットワーク

ニューラルネットワークは、人間の脳神経回路を模倣したアルゴリズムで、機械学習やディープラーニングの基盤技術として用いられる。この手法は、多層構造（入力層、隠れ層、出力層）での処理を通して、非線形なデータの特徴を自動的に学習する。その結果、画像認識や自然言語処理、予測分析など、従来の手法では困難であった複雑な問題を効果的に解決できる点が特徴である。

2.2 ランダムフォレスト

ランダムフォレストとはたくさんの決定木を使って予測や分類を行う機械学習のアルゴリズムのことである。予測を行う際には、まず各決定木がそれぞれの予測を出す。その上、分類問題の場合では多数決で最も多くの木が選んだクラスを最終的な予測とし、回帰問題の場合では各木の予測値の平均値を最終的な予測結果とする。

2.3 勾配ブースティング

勾配ブースティングは、機械学習において強力なアンサンブル学習手法の 1 つであり、弱い予測モデル（通常は決定木）を逐次的に構築し、それぞれが前のモデルの誤差を修正することで予測性能を向上させる。過学習の抑制や高い汎化性能を持つ点が特徴で、多くの実務応用で利用されている。

3. 実験概要及び実験結果

3.1 実験概要

本稿では、先に述べたニューラルネットワーク、ランダムフォレスト、勾配ブースティングの 3 つの機械学習モデルを用いる。3 つの手法で同一のデータを分析することにより、分析結果の比較をして、考察の精度を高めていく。本稿では Footy Stats[1] のデータを使用し、J リーグ、プレミアリーグ、セリエ A、ラ・リーガ、リーグアン、ブンデスリーガの 6 つのリーグのデータを用いる。また、昨年を終了している 22-23 シーズンから遡って、過去 10 年分のデータを用いて分析を行う。本稿では、目的変数をリーグ順位とする。また、説明変数は各チームのボール支配率、シュート数などの 292 項目とし、サンプルサイズは 1154 である。

3.2 ニューラルネットワークによる分析結果

図 1 はニューラルネットワークによる分析結果で予測の上で重要である特徴量の重要度を示している。また、「カード数」、「ファウル数」と関連する項目が上位 10 項目のうち半数の 5 項目を占める結果となった。

3.3 ランダムフォレストによる分析結果

図 2 はランダムフォレストによる分析結果を示している。（詳細な分析結果は論文を参照せよ。）重要度 1 位の特徴量は先制得点試合数となり、他の項目に大きな差をつけた。また、ポゼッションと関連する項目が 2,3,4 位と続いており、ポゼッションの重要度も高いと言える。

3.4 勾配ブースティングによる分析結果

勾配ブースティングの分析結果は、ほぼランダムフォレストと同様の結果となり、他の項目と大きな差をつけて、先制得点数が重要度 1 位となった。

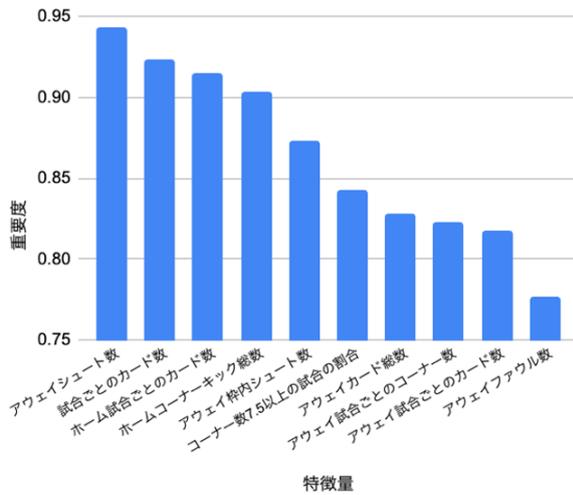


図 1: ニューラルネットワークによる分析結果

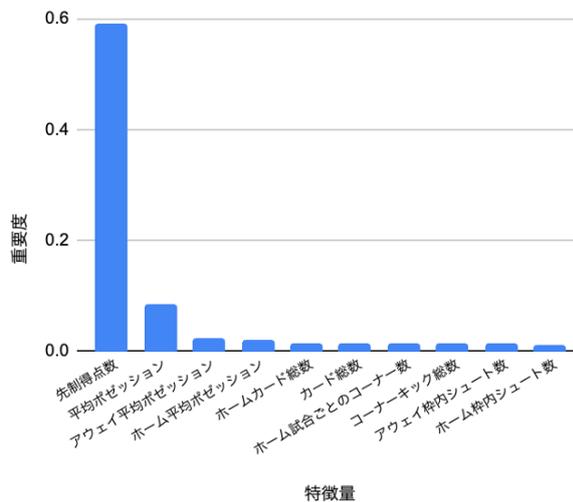


図 2: ランダムフォレストによる分析結果

4. 考察

4.1 3 手法を比較しての考察

まずランダムフォレストと勾配ブースティングにおいて、圧倒的に重要度が高い先制得点数は、ニューラルネットワークにおいても 20 位に入っており、3 手法を通して重要度が高い。また同じく上位を独占しているポゼッションに関する項目に着目すると、ニューラルネットワークにおいても 13 位に平均ポゼッションが入っており、こちらも重要度の高さが伺える。さらに、カード数に関する項目も 3 手法全てで上位に入っており、重要度が高いと言える。以上の結果より、本研究において 3 手法全てで重要度が高い、先制得点数、

ポゼッション、カード数と関連する項目がより重要度の高い項目であると考えられる。

4.2 川崎フロンターレに対する戦略提案

本研究では、重要度の高い特徴量先制得点試合数、ポゼッション、カード数に着目し、それぞれに対して戦略を考察した。そこで、川崎フロンターレに対して以下の戦略を提案する。守備ではフィジカル面で優位な選手を多く起用し守備強度を向上させ、デュエルに強い選手を活用することでノーファウルの守備を実現する。さらに、川崎フロンターレの強みであるポゼッションサッカーを活用し、ポゼッション率を向上させることで守備時間を減少させ、ファウル数や警告数を削減し、試合を優位に進めて先制得点を目指すべきである。

5. おわりに

本研究では膨大なスタッツの分析を元に機械学習の手法を利用して分析を行なった。その分析結果を根拠にサッカーチームの戦力を高める戦略を提案した。今後、より詳細なスタッツのデータを入試できるのであれば、分析を深めて、細部までより効果的な戦略提案が実現できると期待する。また、サッカーの試合に対してリアルタイムで機械学習を用いることができれば、各チーム状況に適した戦略提案をタイムリーに行うことが可能となり、戦術の幅を広げられると考えられる。

6. 参考文献

[1] FootyStats: Football Stats, Tables & Results Soccer Stats, (Data downloaded on July 27th, 2024), <https://footystats.org>.