

卒業論文

ランダムフォレストによるサッカー W 杯の優勝国予想

2014011 石田亜斗武

指導教員 六井 淳 役職
副査 渡邊貴之 役職

9999 年 99 月

静岡県立大学経営情報学部

概要

サッカーは世界で最も人気のあるスポーツであり、その中でも FIFA ワールドカップは注目の集まるイベントである。さらに、ワールドカップのような経済効果が大きく予測不可能性が高いイベントに対して、試合結果を予測することにも注目が集まっている。しかしながら、試合の予測は多くの不確定要素を含むため困難を極める。

本研究では、これらの予測の難しさに対処し、より精度の高いモデルの構築を目的とする。ランダムフォレストという機械学習手法を用いて、優勝国の予測を行うモデルを開発する。本手法は、試合結果予測において重要な特徴量を効果的に評価することが可能であり、過去のデータから学習を行うことで、試合の結果を予測する。実験では、データの収集と前処理、特徴選択、ハイパーパラメータチューニングを行い、交差検証を通じてモデルの有効性を評価する。

本研究の結果は、賭けや投資の意思決定サポート、チームの戦略立案に対する有益な洞察を提供することが期待される。

Abstract

Soccer is the most popular sport in the world, and the FIFA World Cup is the most high-profile event of its kind. Furthermore, for an event with such a large economic impact and high unpredictability as the World Cup, there is a growing interest in predicting the outcome of matches. However, predicting matches is extremely difficult due to the many uncertainties involved.

This study aims to address these prediction difficulties and build a more accurate model. We will develop a model that predicts the winning country using a machine learning method called Random Forest. This method can effectively evaluate features that are important in predicting the outcome of a match and learns from past data. Experiments involve data collection and preprocessing, feature selection, and hyperparameter tuning to evaluate the effectiveness of the model through cross-validation.

The results of this study are expected to provide useful insights into betting and investment decision support and team strategy planning.

目次

第 1 章	はじめに	1
1.1	研究背景	1
1.2	研究目的	1
1.3	論文の構成	2
第 2 章	関連技術	3
2.1	ランダムフォレスト	3
2.2	特徴選択	3
2.3	データの前処理	3
2.4	ハイパーパラメータチューニング	3
第 3 章	検証実験	4
3.1	実験内容	4
3.2	収集データ	4
3.3	評価方法	4
3.4	実験結果	4
3.5	考察	4
第 4 章	まとめと今後の課題	5
4.1	まとめ	5
4.2	今後の課題	5
	謝辞	6
	参考文献	7
	付録 A	8

第 1 章

はじめに

1.1 研究背景

サッカーは、世界中で最も人気のあるスポーツの一つである。その頂点に位置するのが、4年に一度開催される FIFA ワールドカップであり、しばしば予想を裏切る結果が生まれる。実際に 2022 年 FIFA ワールドカップでの日本のスペインやドイツといったサッカーの強豪国に勝利したことは世界中に大きな驚きを与えた [1]。このような予測の困難さが魅力の一つとも言えるが、同時に多くのファンやステークホルダーが正確な予測を求めている。近年、スポーツの試合結果を科学的に分析するアナリティクスの分野が注目され、予測の精度を高めるために統計的手法や機械学習が利用されるようになってきている。[2] この傾向は、ワールドカップのような経済効果が大きく [3] 予測不可能性が高いイベントにおいても、重要な意味を持ち始めていると考える。そこで本研究では、自分の長年のサッカーの経験と、人工知能や機械学習の知識を使い、ワールドカップの優勝国を予測できるモデルを構築することを提案する。

1.2 研究目的

本研究の目的は、過去の FIFA ワールドカップのデータをもとに、優勝国を予測する高精度な予測モデルを構築することである。予測モデルによる予測結果を利用し、賭けや投資の意思決定のサポート、チームの戦略立案への活用など、FIFA ワールドカップに関わる人々にとって有益な情報を提供することができると期待される。本研究では、先行研究 [4][5] からスポーツの勝敗予測に適しているとされるランダムフォレストを用いて予測モデルを構築する。ランダムフォレストには特徴量の重要度の評価ができるなどの特徴 [6] があり、スポーツの勝敗予測に対して有用な特徴変数を確定することが困難である [7] という問題も改善できると考えられる。モデルの評価では、交差検証や混同行列、精度や再現率、F1 スコアなどの指標からモデルを評価する。

2 第1章 はじめに

1.3 論文の構成

第2章では、本研究の関連技術と用語について述べる。

第3章では、検証実験について述べる。

第4章では、まとめと今後について述べる。

第 2 章

関連技術

2.1 ランダムフォレスト

2.2 特徴選択

2.3 データの前処理

2.4 ハイパーパラメータチューニング

第 3 章

検証実験

3.1 実験内容

3.2 収集データ

3.3 評価方法

3.4 実験結果

3.5 考察

第 4 章

まとめと今後の課題

4.1 まとめ

4.2 今後の課題

謝辭

参考文献

- [1] 公益財団法人日本サッカー協会, FIFA ワールドカップカタール 2022, JFA.jp, https://www.jfa.jp/samuraiblack/worldcup_2022/schedule_result/
- [2] 谷岡広樹, ”スポーツアナリティクスにおけるデータと AI 活用”, P1, 2020
- [3] サッカー FIFA ワールドカップロシア大会組織委員会, <https://digitalhub.fifa.com/m/5afd3d89f0e69eb/original/ya7pgcyslzpqlmqjkykg-pdf.pdf>, 2018
- [4] Rahul Baboota, Harleen Kaur, “Predictive analysis and modelling football results using machine learning approach for English Premier League” P15 2018
- [5] Teno Gonz’alez, Dos Santos, Chunyan Wang Niklas, Carlsson Patrick Lambrix, Linköping University, Sweden. “Predicting Season Outcomes for the NBA” , P11, 2021
- [6] Matthias Schonlau, Rosie Yuyan Zou, “The random forest algorithm for statistical learning” , P4, 2020
- [7] Milad Keshtkar Langaroudi, Mohammad Reza Yamaghani, “Sports Result Prediction Based on Machine Learning and Computational Intelligence Approaches A Survey” , P1, 2019

付録 A