

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 教授
副査 渡邊貴之 教授

2024 年 1 月

静岡県立大学経営情報学部

概要

スポーツの勝敗予測は、選手のコンディション、戦術の変化、チームの相互作用など、多数の変数に影響されるため、予測が非常に困難である。近年では、予測が困難なスポーツの試合結果の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いる。そこで本研究では、私自身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、FIFA ワールドカップ優勝国の予測を行う。

本研究ではスポーツの勝敗予測における変数の複雑性に対応するため、ランダムフォレストという機械学習手法を用いる。ランダムフォレストは他の手法に比べて、多数の決定木を利用することでデータの多様性と複雑性を捉え、過学習のリスクを軽減し、信頼性の高い予測結果を提供する。また、モデル構築とチューニングの容易であり、限られた研究期間内での成果を実現可能にする。この研究の成果は、スポーツ分析の分野でのデータ活用の高まりに応じて、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期待できる。

Abstract

Predicting winners and losers in sports is extremely difficult because it is influenced by numerous variables, such as player conditions, tactical changes, and team interactions. In recent years, statistical methods and machine learning have been used to improve the accuracy of predicting the outcome of difficult-to-predict sports matches. Therefore, in this study, I will use my own many years of soccer experience and knowledge of artificial intelligence and machine learning to predict the FIFA World Cup winning country.

To cope with the complexity of variables in sports win/loss prediction, this study uses a machine learning method called random forests. Compared to other methods, Random Forest captures the diversity and complexity of the data by utilizing a large number of decision trees, reducing the risk of overlearning and providing reliable prediction results. In addition, it is easy to build and tune models, enabling results to be achieved within a limited research timeframe. The results of this research are expected to provide useful information for those involved in the FIFA World Cup in response to the growing use of data in the field of sports analysis.

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	1
第 2 章	関連技術	3
2.1	ランダムフォレスト	3
2.1.1	特徴選択	3
2.2	データの前処理	3
2.3	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	データ詳細	4
3.2	評価指標	4
3.2.1	混合行列	4
3.2.2	正解率	4
3.2.3	適合率	4
3.2.4	再現率	4
3.2.5	F 値	4
3.3	検証内容	4
3.4	検証結果	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

第 1 章

はじめに

1.1 研究背景

スポーツの勝敗予測は、選手のコンディション、戦術の変化、チームの相互作用など、多数の変数に影響されるため、予測が非常に困難である [1]。近年では、予測が困難な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]。そこで本研究では、私自身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、FIFA ワールドカップ優勝国の予測を行う。

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高精度な予測モデルを構築し、優勝国を予測することである。本研究ではランダムフォレスト [3] という機械学習手法を用いて予測を行う。ランダムフォレストは他の手法と比較して、複数の決定木 [4] を利用することでデータの多様性と複雑な関連性を捉え、過学習 [5] のリスクを減らす [6][7]。この能力により、スポーツの試合結果に影響する様々な変数を効果的に処理し、信頼性の高い予測結果を導き出すことが期待できる。また、他の手法に比べるとモデル構築とチューニングが容易 [8] であるという特徴があり、卒業研究という限られた期間の中で成果を出すことができると考えられる。以上の理由から本研究ではランダムフォレストを採用する。

近年ではスタッツデータと呼ばれるスポーツの試合に関するデータの活用の注目が高まり、リーグやチームの選手強化や戦術立案、スポーツベッティングで利用されている [9] ことから、本研究の予測結果は 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期待できる。

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。

2 第1章 はじめに

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

2.1.1 特徴選択

2.2 データの前処理

2.3 ハイパーパラメータチューニング

第 3 章

検証実験

3.1 データ詳細

3.2 評価指標

3.2.1 混合行列

3.2.2 正解率

3.2.3 適合率

3.2.4 再現率

3.2.5 F 値

3.3 検証内容

3.4 検証結果

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] Milad Keshtkar Langaroudi, Mohammad Reza Yamaghani, “Sports Result Prediction Based on Machine Learning and Computational Intelligence Approaches A Survey” , P1, 2019
- [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] IBM”ランダムフォレストとは”,<https://www.ibm.com/jp-ja/topics/random-forest>
- [4] Cacao,”決定木分析（ディジションツリー）とは？概要や活用方法、ランダムフォレストも解説”, <https://cacao.com/ja/blog/what-is-decision-tree/>
- [5] TRYETING”機械学習における過学習とは何か？原因・回避方法をくわしく解説”, <https://www.tryeting.jp/column/6846/>
- [6] Matthias Schonlau, Rosie Yuyan Zou, “The random forest algorithm for statistical learning” , P4, 2020
- [7] Kai Liang,”Analysis and Evaluation of Sports Effect Based on Random Forest Algorithm under Big Data”,P2, 2022
- [8] Philipp Probst,”Hyperparameters and Tuning Strategies for Random Forest”,<https://arxiv.labs.arxiv.org/html/1804.03515>
- [9] 経済産業省, スポーツ庁, ”スポーツ DX レポート”,2022

付録 A