

NBA ボックススコアを用いた選手 EFF の予測と 予測精度に影響する選手特性の分析

福島 遥仁 (2220166)

知能工学科 データ科学講座

指導：黒木 進 准教授

1 はじめに

スポーツアナリティクスの発展に伴い、パフォーマンスを定量的に評価・予測が行われるようになった。

本研究は NBA 選手を対象に、過去のボックススコアを用いて次試合の EFF を高精度に予測モデルの作成を主目的とし、モデルの評価・考察、そして予測誤差に影響する選手の役割を分析する。

2 EFF 予測モデル

本研究では、nba_api [1] を用いて NBA のボックススコアデータを取得し、出場時間が 24 分以上かつ 100 試合以上出場した選手 218 名を分析対象とした。これらをスター選手 54 名とロールプレイヤー 164 名に分類した。説明変数には PTS や AST などの基本スタッツに加え、直近試合の平均や差分指標など計 171 項目を用いた。データは時系列性を考慮して訓練データとテストデータを 80:20 で分割し、線形回帰、XGBoost (XGB)、および分位回帰モデルにより次試合 EFF を予測した。評価指標には MAE (平均絶対誤差) と WAPE (平均絶対パーセンテージ誤差) を用い、線形回帰では SHAP 値に基づく説明変数選定を行った。

3 実験結果と考察

EFF 予測における線形回帰と XGB の性能差は小さく、全選手を対象とした MAE は、線形回帰で 6.60、XGB で 6.70 とほぼ同程度であった。選手区分別に見ても、スター選手では線形回帰が 7.45、XGB が 6.68、ロールプレイヤーでは線形回帰が 6.30、XGB が 6.38 と、いずれも大きな差は確認されなかった。これより、単純な線形回帰であっても、次試合 EFF 予測において一定の精度が得られることが分かる。一方で、選手区分による誤差特性の違いは顕著であった。スター選手では MAE が 7.45 と、全選手の 6.60 やロールプレイヤー 6.30 と比較して大きく、絶対誤差の観点から予測が困難であることが示された。これに対してロールプレイヤーでは、MAE は比較的小さいものの、WAPE は 44.7% と高く、全選手 39.1% やスター選手 30.3% を上回る結果となった。これは、スター選手では EFF 水準が高いため相対誤差が小さく評価される一方、ロールプレイヤーでは EFF 水準が低く、わずかな誤差が大きな割合誤差として現れることに起因しており、スタッツベース予測が選手の役割や試合

文脈を十分に捉えられていないことを示している。さらに、線形回帰において説明変数の選定を行った結果、予測精度を維持したままモデルの簡潔性と解釈性が向上し、特にスター選手では MAE が 7.45 から 7.39 へ低下するなど、不要な変数の除外が誤差低減に寄与した。

図 1 より、予測誤差が大きい選手は、PTS や AST、USG_PCT の高い主力選手で、試合ごとの役割変化によりパフォーマンスの変動が大きい。一方、誤差の小さい選手は REB や BLK、PF が高いインサイド選手が多く、役割が安定しているため再現性が高いと考えられる。

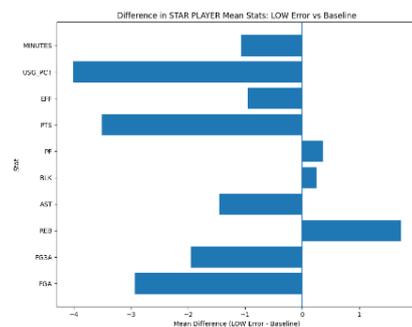
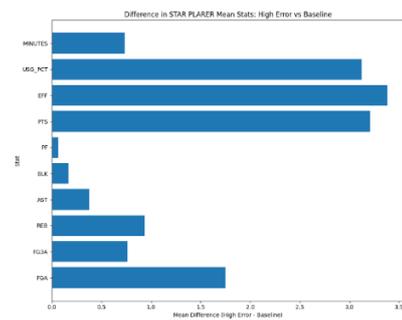


図 1：スター選手の MAE 誤差上位 20%(上)、下位 20%(下)の選手区分平均スタッツとの差

4 おわりに

本研究は NBA 選手を対象に、過去のボックススコアから次試合の EFF を予測し、予測誤差の大小と選手の役割との関連を可視化することができた。

参考文献

[1] GitHub. nba_api.

https://github.com/swar/nba_api (2025 年 12 月参照)。