

解釈可能な機械学習モデルを用いた 女子サッカーの得点予測

学籍番号 4122010

氏名 梅田 孝一

【目的】

統計手法と解釈可能な機械学習モデルを用いて、女子サッカーのYogibo WEリーグ(WEリーグ)と2023年FIFA Women's World Cup(女子W杯)決勝トーナメントの得点分析を行い、シュートに関わる様々な条件が得点予測に与える影響の大きさを客観的に分析することを目的とした。

【方法】

YouTubeやStatsBomb社のGitHub上のオープンリポジトリから収集されたデータに基づき、WEリーグ21試合から402本のシュート、43の得点、女子W杯決勝トーナメント16試合から330本のシュート、33のゴールを分析対象とした。分析項目は先行研究に基づきTime, Play pattern, Shot zoneなどとした。
まず大会間で各分析項目のシュートと得点割合の差異を、カイ二乗検定で検証した。
さらにPythonで解釈可能な機械学習モデルを大会毎に実装した。

【結果】

カイ二乗検定の結果、Play pattern、Shot zone、Last playでの有意差が確認された。
機械学習を用いた分析により、コーナーキックからのシュートはWEリーグにおいては得点予測にプラスの影響を与えた一方、女子W杯においてはマイナスに作用した。また、グラウンダーパスからのシュートは、女子W杯はWEリーグに比べて大きくプラスに作用していたことが分かった。またペナルティーエリアの角からのシュートはWEリーグではマイナスに作用するが、女子W杯ではプラスに作用していた。また機械学習モデルの予測精度をROC曲線下の面積AUCと、ブライアースコアで評価した結果、両大会において十分な予測精度を得られた。

【結論】

女子サッカーの得点予測を解釈可能な機械学習モデルを用いて十分な精度で行えた事に加え、二つの異なる大会においてシュートに関わる各分析項目が得点予測に与える影響を客観的に比較できた。

Using interpretable machine learning models to predicting scores for women's soccer

Student ID Number: 4122010

Name: UMEDA, Koichi

[Purpose]

The purpose of this study was to objectively analyze the magnitude of the impact of various shooting-related conditions on score prediction in women's soccer. To this end, we used statistical methods and interpretable machine learning models to analyze the scores of the Yagibo WE League and final tournament of the 2023 FIFA Women's World Cup.

[Methods]

Based on data collected from YouTube and StatsBomb's open repository on GitHub, 402 shots and 43 goals from 21 WE League games and 330 shots and 33 goals from 16 Women's World Cup Final Tournament games were analyzed. Based on previous studies, the analysis items included time, play pattern, and shot zone, among others. First, the differences in the percentage of shots and goals scored for each analysis item between tournaments were verified with a chi-square test.

Subsequently, machine learning models interpretable in Python were implemented for each tournament.

[Results]

The results of the Chi-square test confirmed significant differences in play pattern, shot zone, and last play between the two competitions. The analyses using machine learning revealed that shots from corner kicks had a positive effect on score prediction in the WE League whereas a negative effect in the Women's World Cup. Shots from passes of grounders had a larger positive effect in the Women's World Cup than in the WE League. In addition, shots from the corners of the penalty area had a negative effect in the WE League whereas a positive effect in the Women's World Cup. The prediction accuracy of the machine learning model was evaluated using the area under the curve (AUC) of receiver operating characteristics (ROC) and Brier score. The results showed that shooting scores for both competitions can be predicted with sufficient accuracy.

[Conclusion]

In addition to being able to predict scores in women's soccer with sufficient accuracy using an interpretable machine learning model, we were also able to objectively compare the impact of each of the shooting-related analyses on score prediction in two different tournaments.