

ラインブレイクの発生確率とゴール期待値を統合した LineBreak_xG モデルの構築

八木 翔摩^{†1}, 市川 淳^{‡2}, 一ノ瀬 元喜^{‡3}

キーワード: サッカー, データ分析, イベント予測, 機械学習, XGBoost, ロジスティック回帰

1. はじめに

サッカーにおいて, 相手ディフェンスラインを突破する「ラインブレイク」は, 得点機会を創出する上で重要なプレーである. ラインブレイクが成功すると, ゴール前での数的優位やフリーでのクロスなどを生み出し, 得点確率を高める.

しかし, 従来の分析ではラインブレイクの発生確率の予測に主眼が置かれており[1], 突破後の状況がどの程度得点に結びつきやすいかは十分に考慮されていない. 同じラインブレイクであっても, 発生位置によって得点につながる可能性は大きく異なるため, 突破後の得点機会を加味した評価が必要である.

そこで本研究では, シュート位置から得点確率を推定するゴール期待値 (xG: expected goals) の考え方[2, 3]をラインブレイクに応用し, 突破後の得点機会を定量的に評価する手法を提案する. 具体的には, ラインブレイク発生確率と, 突破直後のボール位置に基づいて算出した得点確率を統合することで, ラインブレイク後に得点へとつながる可能性を評価する新たな指標を構築する.

2. 分析方法

2-1. 使用データ

本研究では, データスタジアム株式会社により提供された明治安田生命 J 1 リーグ 2023 年シーズンの第 30-34 節, 計 45 試合分のイベントデータ (パス・シュート等の行動ラベルとその時点のボールの位置座標) とトラッキングデータ (選手全員とボールの位置座標) を分析に用いた.

2-2. モデル作成

ラインブレイクを予測する分類器には, XGBoost を用いた[4]. XGBoost は, 決定木を用いて, 勾配ブースティングを効率的かつスケーラブルに最適化した学習アルゴリズムである. また, モデルの学習には全選手の位置座標や速度, ディフェンスラインに関する情報など合計 189 次元を特徴量 (x_{189}) として使用した[1].

xG を推定するモデルには, ロジスティック回帰を用いた. ロジスティック回帰は, 入力特徴量とゴール発生確率

の関係を確率的に表現できるモデルであり, サッカー分析において xG 算出に広く用いられている. 本研究では, シュート時のボール位置から算出したゴール中心までの距離と角度の 2 変数を説明変数とし, ゴールの成否を目的変数としてモデルを学習した[2, 3]. 具体的には, あるシュート位置(x, y)に対して, 距離 d と角度 θ を次のように計算する.

$$d = \sqrt{(x - x_{goal})^2 + (y - y_{goal})^2}, \theta = \arctan\left(\frac{|x - x_{goal}|}{|y - y_{goal}|}\right)$$

ここで, $(x_{goal}, y_{goal}) = (5250, 0)$ はゴール中心の座標である. また, ロジスティック関数は以下の式で表される.

$$xG = P(\text{Goal}|d, \theta) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 d + \beta_2 \theta))}$$

ここで, β_0 は切片, β_1 は距離に対応する係数, β_2 は角度に対応する係数である.

これらを統合し, ラインブレイク確率と, その発生地点における xG を掛け合わせた指標「LineBreak_xG」を提案する:

$$\text{LineBreak_xG} = P(\text{LineBreak}|x_{189}) \times P(\text{Goal}|d, \theta)$$

これにより, ラインブレイクの発生確率だけでなく, ゴールに結びつきやすいという質的な評価も可能となった

3. 結果と考察

3-1. 予測結果例

図 1 は第 30 節横浜 F・マリノス対セレッソ大阪の試合の一場面である. 図 1 (A) のセレッソの 27 番カピシヤバ選手がパスを出す場面において今回のモデルでは確率約 0.870 でラインブレイクが起これと予測した. 図 1 (B) は 16 番の毎熊選手が相手ディフェンスの背後に抜け出し, そのパスをダイレクトでクロス上げた瞬間である. これは, ディフェンスラインの裏でボールをタッチしているため実際にラインブレイクが起こった場面である. さらに, 本研究で構築した xG モデルに基づく, この場面での xG は約 0.05, ラインブレイク確率との積である LineBreak xG は約 0.04 と算出された. これにより, 本モデルが単に突破の有無を捉えるだけでなく, その後の得点可能性を定量的に評価できることが示される.

^{†1} 静岡大学大学院総合科学技術研究科工学専攻

^{†2} 静岡大学大学院情報学領域情報科学系列

^{†3} 静岡大学大学院工学領域数理システム工学系列

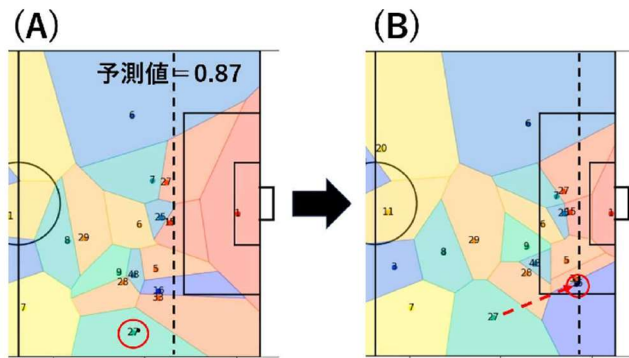


図 1 実際にラインブレイクが発生したシーンの予測例。
(A) セレッソ 27 番カピシャバ選手がパスを出す瞬間。
(B) セレッソ 16 番毎熊選手がパスをダイレクトでクロス
をあげた瞬間。

3-2. 被 LineBreak_xG と既存指標の比較

まず、従来指標との比較のため、図 2 に各チームの被クロス数と被シュート数の合計と従来指標である被ラインブレイク確率（ラインブレイクされる確率）の関係を示す。予測した各チーム 5 試合における縦軸は被クロス数と被シュート数の合計、横軸は被ラインブレイク確率の合計を示す。この散布図から、両指標の間に中程度の正の相関 ($r = 0.383$, $p = 0.117$) がみられる。

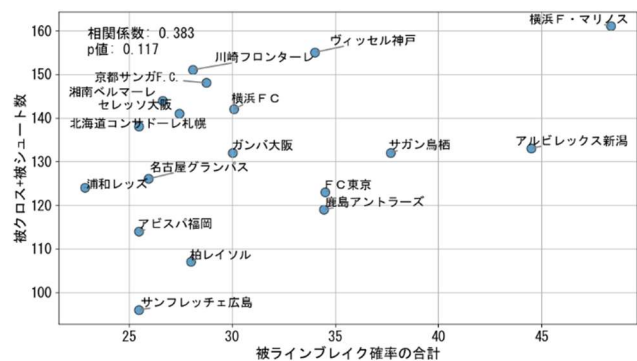


図 2 チームごとの被クロス数と被シュート数の合計と被
ラインブレイク確率の関係

さらに、図 3 は横軸を被 LineBreak_xG、縦軸を被クロス+被シュート数とした散布図である。その結果、両者の間に図 2 よりも強い正の相関 ($r = 0.469$, $p = 0.049$) がみられ、単なる突破の頻度よりも、突破後の危険度を考慮した指標である LineBreak_xG の方が被クロス・被シュート数との関連性をよりの確に捉えていることが示唆された。

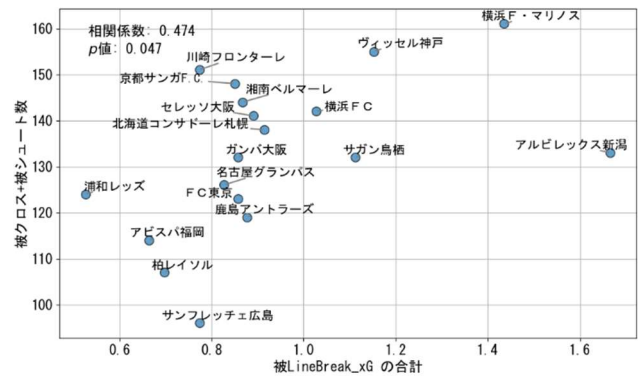


図 3 チームごとの被クロス数と被シュート数の合計と被
LineBreak_xG の関係

次に、図 3 を基にチームごとの特徴を考察する。横浜 F・マリノスはハイプレスと縦に速い攻撃を特徴とする一方で、戦列が間延びしやすく、カウンターを受けやすい構造を持つ。その結果、被クロス数・被シュート数に加え、被 LineBreak_xG も高い値を示しており、攻撃的戦術と守備リスクのトレードオフが確認できる。

4. まとめ

本研究では、ラインブレイクを単なる発生確率ではなく、「突破後にどれだけ得点機会に結びつくか」という観点から評価するため、新たな指標 LineBreak_xG を提案した。具体的には、XGBoost によりラインブレイク発生確率を予測し、その発生位置に基づく xG と統合することで、突破の危険度を定量化した。

その結果、被 LineBreak_xG は被クロス数・被シュート数と有意な相関を示し、従来の被ラインブレイク確率よりも、実際の攻撃機会に直結する守備リスクを捉えやすいことが示唆された。このことから、ラインブレイクを「起きたかどうか」だけでなく、「どれだけ危険な局面を生むか」という視点で評価する重要性が明らかとなった。

参考文献

- [1] 八木翔摩, 益留琢磨, 宮川大樹, 福嶋克茂, 市川淳, 一ノ瀬元喜. サッカーの守備突破の瞬間を捉えるためのラインブレイク予測モデルの構築と分析. 2024 年度スポーツデータサイエンスコンペティション研究報告集, pp. 199-202. 日本統計学会 スポーツデータサイエンス分科会, 2024.
- [2] I. Umami, D. H. Gutama, and H. R. Hatta. Implementing the expected goal (xG) model to predict scores in soccer matches. International Journal of Informatics and Information Systems, Vol. 4, No. 1, pp. 38-54, 2021.
- [3] A. Scholtes and O. Karaku, s. Bayes-xG: Player and position correction on expected goals (xG) using bayesian hierarchical approach. Frontiers in Sports and Active Living, Vol. 6, p. 1348983, 2024.
- [4] T. Chen and C. Guestrin. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794, 2016.