

木構造型ニューラルネットワークを用いた野球における 勝率予測と意思決定に関する研究

A study on win probability and decision making in baseball using tree structured neural network

安永 直央

北條 仁志

大阪府立大学工学研究科

大阪公立大学情報学研究科

Yasunaga Nao

Hohjo Hitoshi

Osaka Prefecture University

Osaka Metropolitan University

Graduate School of Engineering

Graduate School of Informatics

1 はじめに

近年、野球における新しい指標として、従来の打率・長打率等に加えて、OPS・IsoP・WHIPといった指標が利用されている。これらの指標は、従来の指標では評価できない部分をカバーしたり、改善したり正確な評価をしたりするために生まれた。OPSは、“On-base Plus Slugging”の略で、出塁率 $((\text{安打} + \text{四死球}) \div (\text{打数} + \text{四死球} + \text{犠飛}))$ と長打率 $(\text{塁打} \div \text{打数})$ を足し合わせた数値である。打者能力をより的確に評価するために生まれた指標で、数値が高いほど優秀な打者であると評価される。IsoPは、“Isolated Power”の略で、打者の長打率をより正確に評価するための指標で、長打率 $-$ 打率で表される数値で、数値が高いほど優秀な打者であると評価される。WHIPは、“Walks plus Hits per Inning Pitched”の略で、 $(\text{被安打} + \text{与四球}) \div \text{投球回数}$ で表される数値である。1回あたり平均何人を塁に出したのかを示す指標で、数値が低いほど優秀な投手であると評価される。このように野球においてデータを統計学的見地から客観的に分析し、選手の評価や戦略などを考える分析手法は、『セイバーメトリクス』とよばれる。その分析の中で多く行われている研究分野の一つとして、試合結果予測がある。このような研究の中では、しばしば機械学習を取り入れたモデルの開発が進められており、Soto-Valero[9]は、最近傍探索・MLP・決定木・サポートベクターマシンを比較をした。機械学習において、Purucker[7]やMcCabe and Trevathan[6]などはスポーツの予測に広く使われている手法として人工ニューラルネットワークを利用した。人間の脳内部の動きをモデリングするニューラルネットワークは、任意の関数をモデル化することに関して有用である[4]ことが証明されており、分類や回帰に関して便利な手法となっている。本研究は、ニューラルネットワークを基本とし分岐構造を作成することで抽象度の違う解釈を可能とした木構造型ニューラルネットワークを利用し、野球における勝率予測を行う。

機械学習の研究として、精度よく予測するモデルが多く存在している。一方で、解釈性と呼ばれる、モデルがその結果を出した理由を説明するということに焦点を当てた研究が増えてきている。説明可能なAIはxAI(eXplainable AI)と呼ばれ、アメリカの国防高等研究計画局(DARPA)[3]などでも研究が進められている分野である。DARPAによると、『説明とはモデルの特徴を意味的な情報と連携させること』とされており、その方法として、どの特徴が判別に効いているのかを示すことやネットワークの入出力の相関関係を理解しやすいモデルなどが挙げられている。精度と解釈性はトレードオフの関係にあり、モデル自体を複雑にすればするほど、精度が向上する可能性は高いが解釈性は低下する。このことから、その両立を求めることは困難とされている。xAIの既存の研究は大きく三つの方法に大別される。一つ目が、モデル自体に解釈性を与える手法である。これは、決定木のようなルールベースのモデルを利用して解釈をしていくもので、CORELS[1]など

がある。CORELS は、Angelino et al.(2017) によって考案された、ルールリストという決定木の亜種を学習し、解釈性の高いモデルを生成する手法で特徴量に対して if-then でルールリストを作成する。カテゴリカルな特徴空間でルールリストを構築するために、組み合わせ最適問題を各種探索の枝刈りを用いて高速化を行い、最終的に解釈性の高いモデルを構築していく。二つ目が、モデルに関わらず解釈性を付与する手法である。これは、説明変数の出力に対する影響力を可視化することなどによって実現され、SHAP[5] や LIME[8] などがある。LIME は、一つの予測結果に対してのみ局所的に近似させた分類機を利用して特徴量の可視化を行う。この方法の利点は、既存の精度の高いモデルに適用させることで精度と解釈性を両立させやすいことにある。三つ目が、解釈可能なモデルで近似する手法である。これは、精度の高い予測をするモデルを近似した解釈可能なモデルを作成するもので、Born Again Tree[10] などがある。この手法は、最初に予測精度の高いモデルを構築し、その後学習したモデルを使い擬似訓練データを生成し、解釈性の高いモデルを構築する。その結果予測精度の高いモデルと解釈性の高い 2 つのモデルから予測精度と解釈性が実現されるものである。

本研究においては、モデル自体に解釈性を与える手法とモデルに関わらず解釈性を与える手法を組み合わせるによって意思決定者の望む抽象度での解釈を可能とすることを目的に研究を進めている。具体的には、分岐構造を事前に作成することでモデルに解釈性を与える下地を作成し、SHAP によって実際に解釈を行っていく。

今回は実験に野球のデータを利用している。理由は二つあり、野球のデータには本塁打数や奪三振数といったように入力一つ一つに意味を持っているということ、そして事前の分岐構造作成にあたり最低限の知識を保有していることである。そのため、今回提案するモデルは、説明変数となる入力値一つ一つに意味を持っているデータ、例えば他のスポーツや物価予測等にも適用可能であると考えている。機械学習の多くの研究の題材となっている画像データは入力が 1 画素であり、意味を保有していないため適用できない。

本論文は、第 2 節で既存手法である SHAP の説明を行い、第 3 節で木構造型ニューラルネットワークモデルを提案し、第 4 節で今回行った実験の概要に関して述べ、第 5 節で結果・考察、第 6 節で提案モデルによって可能となった抽象度の違う解釈を行うことで、このモデルの有用性を主張する。

2 SHAP

本研究で解釈に利用する SHAP に関して説明する。これは、協力ゲーム理論のシャープレイ値を機械学習に応用したもので、一つ一つの特徴量がモデル予測値に与える貢献度を表現する。

まず、協力ゲーム理論における考え方について説明する。以下の表 1 の条件を計算例とする。この時、それ

表 1: 参加者の組み合わせごとの獲得賞金

参加者	賞金
A さん	10 万円
B さん	6 万円
C さん	4 万円
A さん B さん	30 万円
A さん C さん	22 万円
B さん C さん	16 万円
A さん B さん C さん	60 万円

それぞれの貢献度にしたがって分配金を決める際に、シャープレイ値を使用する。

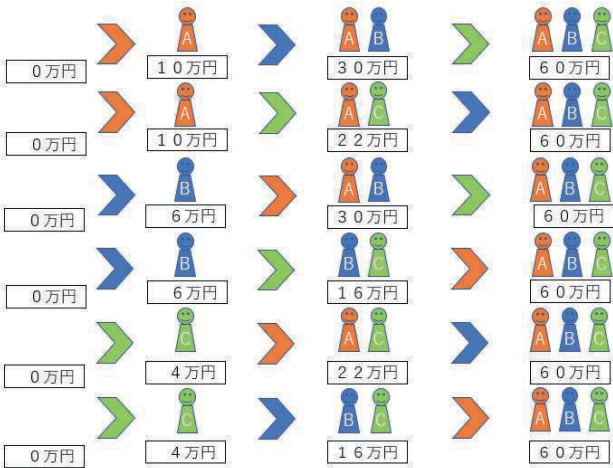


図 1: ゲーム理論における例

計算方法は、3 人を順番に加えたときの賞金の変化額の平均をとる。今回のケースだと、3 人の加え方は 6 通りであり、それぞれの変化は図 1 の通りである。よって、A さんの貢献度は $(10 + 10 + 24 + 44 + 18 + 44)/6 = 25$ 万円、B さんの貢献度は $(20 + 38 + 6 + 6 + 38 + 12)/6 = 20$ 万円、C さんの貢献度は $(30 + 12 + 30 + 10 + 4 + 4)/6 = 15$ 万円となる。この 3 人の合計金額が 60 万円になるため、この値を分配金とすることができる。

協力ゲーム理論においては、誰もいない状態から人を追加していく方法をとっているが、機械学習において入力を与えないことはできないため、いない状態を予測モデルの期待値という考え方を導入することで解決する。すなわち平均的な予測値をベースに、ある特徴量が加わったときの予測値の変動量を、その特徴量の限界貢献度としていく考え方の導入である。今回は、それぞれの特徴の平均値を利用して平均的な予測値を算出した。機械学習モデルでの SHAP 利用時の限界貢献度のイメージについて説明する。このモデルでは、気温・湿度・入場者数という 3 つの特徴からアイスクリームの売り上げ予測を行うことを想定している。図 2 はこの時の気温 → 湿度 → 入場者数の順に確定した特徴量を加えたときの限界貢献度のイメージを示したものである。予測の期待値から、気温の値を変更した際の変化、湿度の値を変更した際の変化、入場者数の値を変更した際の変化の大きさと予測に対する影響を矢印の大きさと向きで表現している。この例では、気温・湿度は売上を増加させる影響を与え、入場者数は売上を減少させる影響を与えており、その大きさは湿度・入場者数・気温の順に大きいことがわかる。

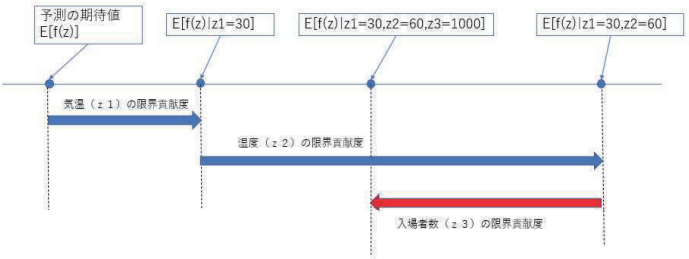


図 2: 限界貢献度のイメージ

同様にしてすべての順序での限界貢献度を求め、その平均を取ればシャープレイ値が求まる。このシャープレイ値によってある予測における各特長の寄与が定量的に表現できる。本研究では、この手法を利用して解釈を行っていく。

3 提案モデル

今回提案するモデルに関して説明する。今回は抽象度の違う解釈を可能にするために事前に分岐を作成し、その分岐の間に従来のネットワーク構造を組み込む。これにより、ネットワークの途中で意味を持ったノードを生み出すことができ、その地点での解釈を付与可能となり、任意の抽象度での解釈を可能とする。

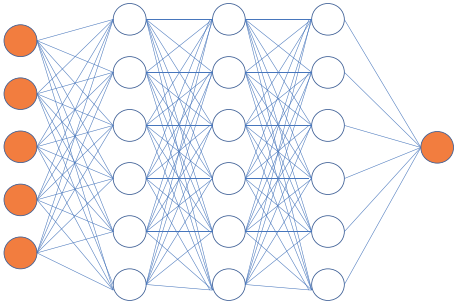


図 3: 従来のニューラルネットワーク

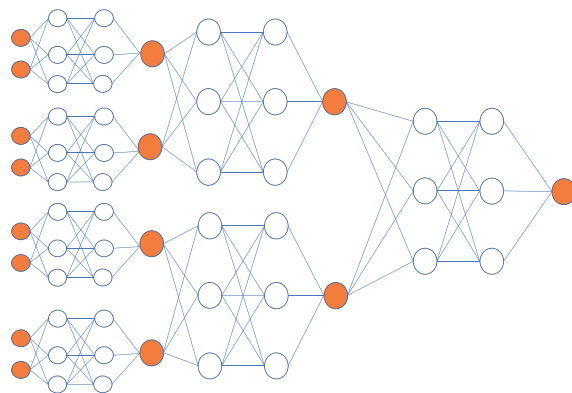


図 4: 提案モデル

図 3 は、従来のニューラルネットワークを、図 4 は今回の提案モデルを模したものである。意味のあるノードに色を付けて表現している。この色付きのノードの追加によって、抽象度の違う解釈を可能とする。既存の手法においては、両端のノードのみが意味のあるノードとして、SHAP を適用することで解釈が可能であるが、今回の提案手法では両端以外の色付きノードによる SHAP の適用も可能になる。これにより、抽象的な概念の影響度も推し量ることが可能となる。意思決定において、具体的な方針（例：練習メニューの決定）を決める際には両端のノードを利用した解釈を利用し、抽象的な方針（例：チームとしての方向性）を決める際には今回のモデルで可能となった解釈を利用することを可能とした。この任意の抽象度による複数の解釈を一つのモデル内で可能としていることがこのモデルの優れている点である。

4 実験概要

野球のデータの特徴をチーム属性と攻守属性の 2 段階に分割（4 分割）したネットワークについて簡単な実験を行う。ここで、チーム属性とは勝率計算における対象のチーム（A）とその対戦相手のチーム（B）で分ける属性で、攻守属性とは攻撃関連の特徴と守備関連の特徴に分ける属性である。今回は、攻撃属性として本塁打数・単打数・2 塁打数・四球数・盗塁数を与え、守備属性として、被安打数・被本塁打数・奪三振数・敬遠数・失策数を採用し、それぞれ A・B のチームの特徴を入力として与えた。実験は python を用いて行い、提案モデルの構築には pytorch を利用し、データは『データで楽しむプロ野球』[11] 内記載の日本プロ野球のデータを利用する。また、比較として、sklearn ライブラリの MLPRegressor に 20 特徴を入力として与えた結果を提示する。

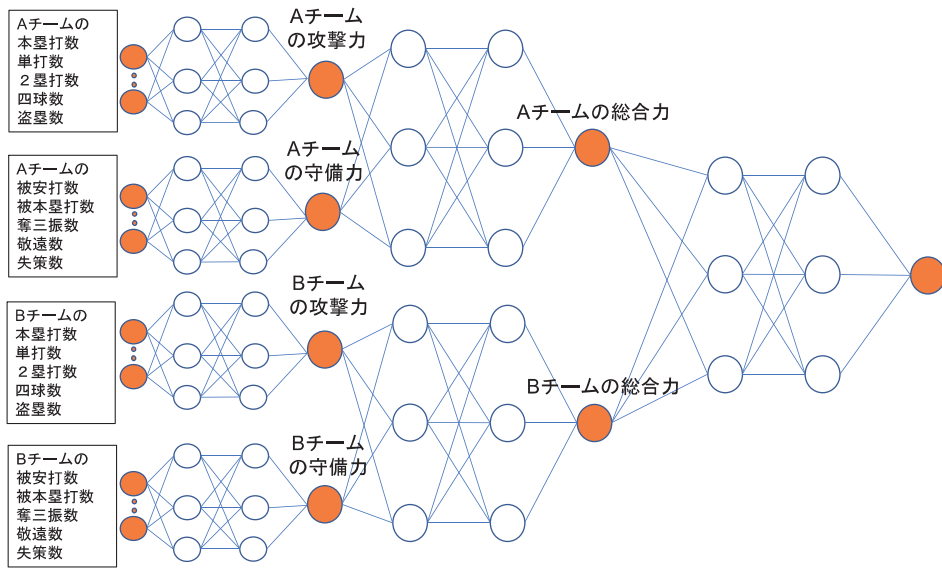


図 5: 実験モデル

モデルのイメージは図 5 の通りである。説明変数である 20 特徴と目的変数である勝率の他に、中間地点に A チーム攻撃力・守備力、B チーム攻撃力・守備力、A チーム総合力、B チーム総合力のノードが追加されている。

また、データは 2011-2021 の各対戦相手ごとのシーズン成績を利用し、各年の A チーム対 B チームの予測には、その年のその組み合わせでの成績をもとに予測する。入力値はそれぞれの値のチームに所属する選手の合計値をもとに全チームの成績で各項目ごとに正規化したものとする。それぞれの特徴の分布は以下の図 6 で示す。

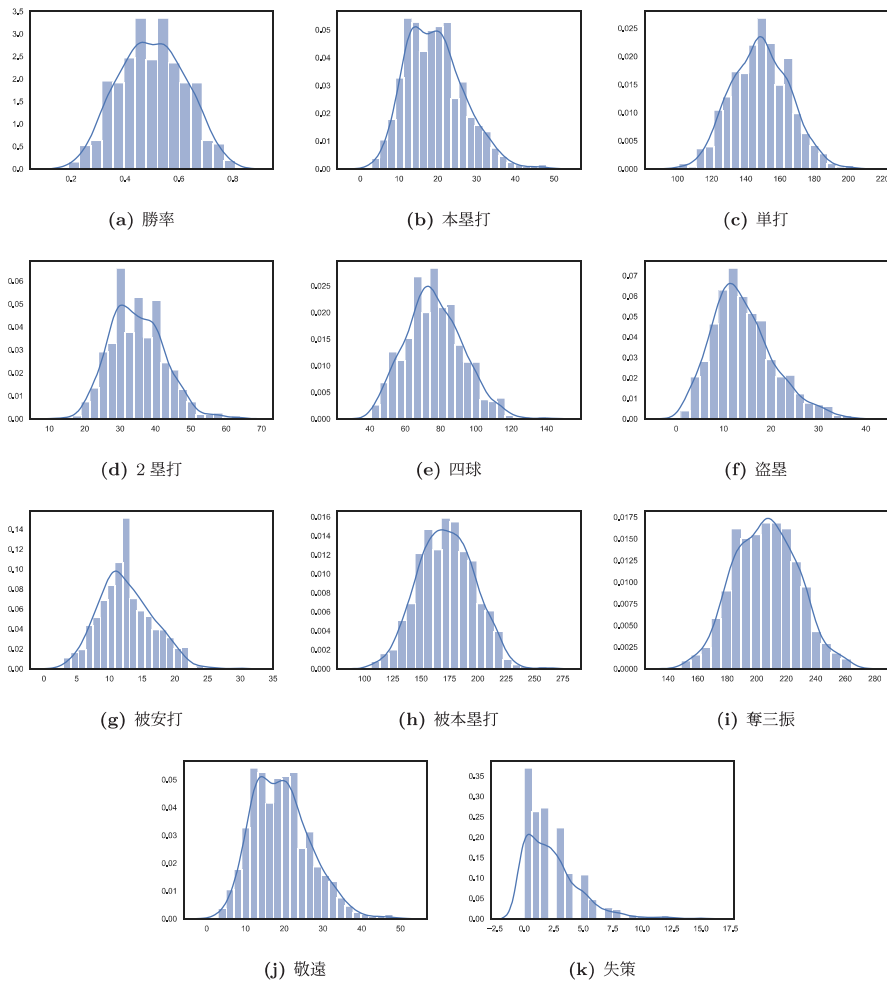


図 6: 分布

図より、ほとんどの特徴は偏りのない正規分布のような分布となっているが、失策という偶発的に起こるものに関しては偏りのある分布となっている。

5 結果

回帰問題としての結果を表 2、分類問題としての結果を表 3 にまとめた。

回帰問題としては、RMSE を利用して精度を比較したが、提案モデルは比較モデルよりも多少精度が悪い結果となった。また、分類問題としての精度は、提案モデルは比較モデルの同等以上の精度となっている。このことから、提案モデルは著しく精度を落とすことなく予測を行えているといえる。分類問題として扱った理

表 2: 回帰問題としての精度

	提案モデル	比較モデル
RMSE	0.09398	0.09306

表 3: 分類問題としての精度

	提案モデル	比較モデル
ACC	0.7636	0.7455
pre	0.7595	0.7125
recall	0.7500	0.7500
spe	0.7765	0.7416
F1	0.7547	0.7308

由としては、解釈として意思決定の支援を考えた際に、シーズンを通しての勝ち越しか負け越しかの予測が間違っているモデルへの信頼は薄くなってしまうことからである。次に、縦軸に正解データ横軸にモデルの予測値をとった場合のプロットを図 7 に示す。

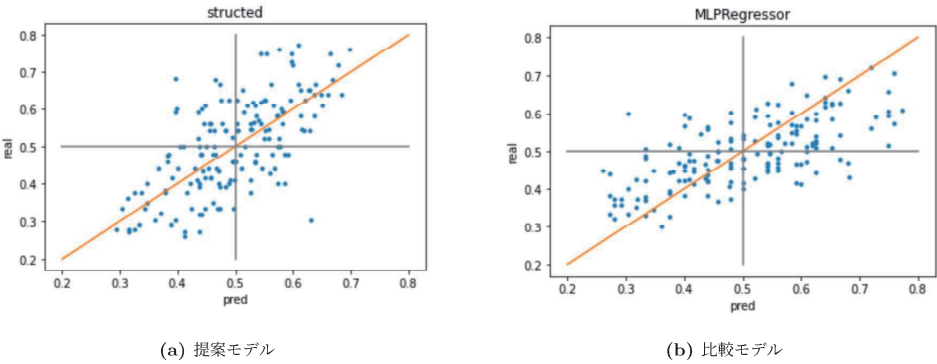


図 7: 結果のプロット

この図より、提案モデルは比較モデルに比べて、相関が小さく、予測値を表現している横軸の幅が小さいことがわかる。これは、予測値が 0.5 によっており、極端な予測をしていないことが原因と考えられる。そのため、エラーなどによる点数の考慮やより特徴を捉える必要がある。

今回導入した途中にある意味のあるノードの結果についてまとめる。このノードの値はネットワーク上にあり、シグモイド関数を活性化関数として利用しているため、0 から 1 の範囲の値をとる。

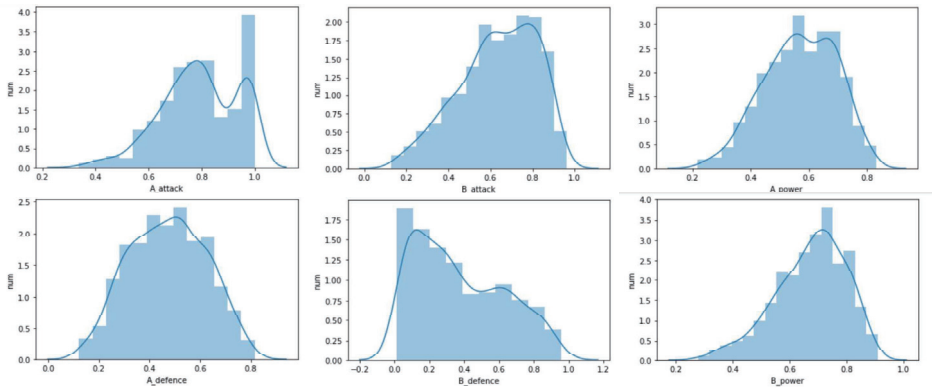


図 8: 中間ノードの結果

図 8 は、左から A チームの攻撃・A チームの守備・B チームの攻撃・B チームの守備・A チームの総合力・B チームの総合力を指すノードの数値データを可視化したものである。この結果より、片側によっているノードが存在していることが判明した。また、A チーム守備と B チーム守備は、順番は異なるが同一データであるにもかかわらず、振る舞いが大きく異なっている。この現象の理由は初期値の違いや学習不足など多々考えられるため、現状不明である。

6 解釈

今回は、3つのケースにおいて、抽象度の違う解釈を SHAP に基づいて行う。

- ケース 1：A チーム vsB チームの試合においてある説明変数の影響を見る
- ケース 2：A チーム vsB チームの試合において各チームの攻撃・守備の影響を見る
- ケース 3：1 シーズンの A チームの他のチームとの試合での平均的な影響を見る

利用するデータは、表 4 の通りである。

表 4: データ詳細（正規化後）

チーム	本塁打	単打	2 塁打	四球	盗塁	失策数	奪三振	被安打	被本塁打	敬遠
A	0.2444	0.4608	0.3000	0.5051	0.3611	0.2593	0.3871	0.2609	0.1333	0.0667
B	0.1333	0.3333	0.3800	0.2828	0.3889	0.3333	0.4065	0.3913	0.2444	0.0000

6.1 ケース 1

まず、A チーム vsB チームの試合においてある説明変数の影響を見る。A チーム攻撃の特徴の影響を観測すると、図 9 のように表現することができる。

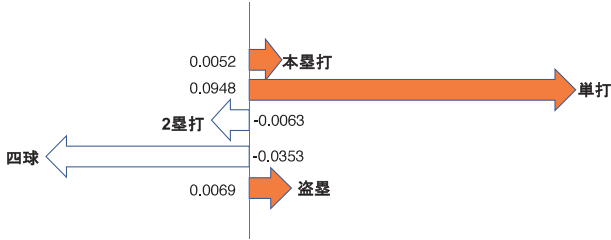


図 9: 貢献度の大きさ

単打の影響力が最も大きく本塁打の影響が最も少ないことが見て取れる。このことは A チームのデータの単打の成績が良く、本塁打の成績が良くないことと一致している。また、ほかのチームと比較して多くなっている四球数は負の影響を与えていることが見て取れる。このことは、四球が多くすること（選球眼を鍛えること）は勝率を向上させるための戦略とすべきではないことがわかる。

6.2 ケース 2

次に、A チーム vs B チームの試合において各チームの攻撃・守備の影響を見る。影響を観測すると、図 10 のように表現することができる。

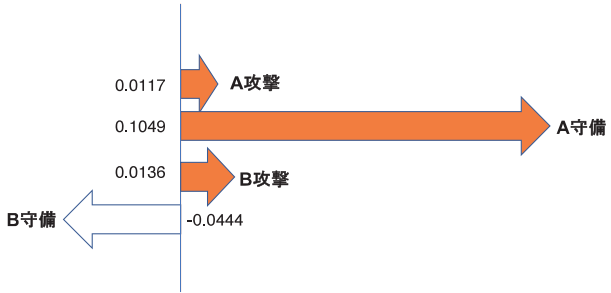


図 10: 貢献度の大きさ

各チーム守備の影響が大きいことがわかるが、B 攻撃が A チームの勝率に正の影響を与えてしまっている。これは常識とは相反する現象である。B チームの攻撃力が高いほど、A チームの勝率が高くなることを示しているためである。実際 B チームの攻撃力は、平均以上の力があり、これが勝率に正の影響を与えることはおかしい。このような現象は、精度が低いためにおこる可能性があり、モデルの性能向上が必要である。

6.3 ケース 3

最後に、1 シーズンの A チームの他のチームとの試合での平均的な影響を見る。表 5、図 11 の通りの分布になる。

表 5: それぞれの貢献度

	本塁打	単打	2 塁打	四球	盗塁	A 守備
最大値	0.0113	0.1091	0.0007	-0.0337	0.0164	0.0097
最小値	-0.0067	0.0948	-0.0019	-0.0551	-0.0048	-0.1364
平均値	-0.0002	0.0982	-0.0108	-0.0461	0.0032	-0.0915

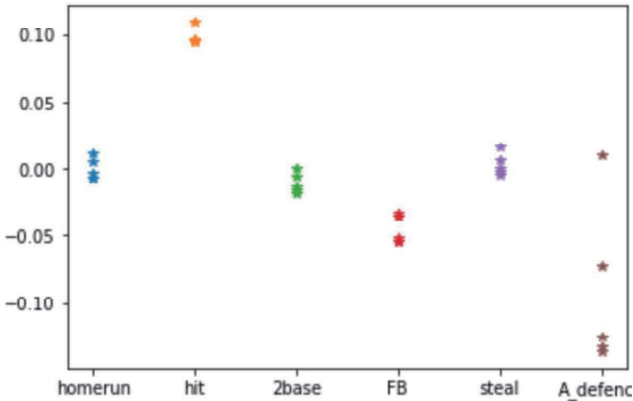


図 11: 各変数の平均的な影響

上記の結果より、A チームの守備を除く A チーム攻撃の特徴（本塁打・単打・2 塁打・四球・盗塁）は、どのチームとの予測においても似た値になっていることが見て取れる。つまり、これらの指標はどのチームに対しても同様の影響を与えているといえる。特徴単位で見ると、単打の影響力が最も大きく、長打（本塁打・2 塁打）の影響はほとんどない。やはり、データから見て取れるようにこのチームはパワーよりもミートのチームであり、それが勝率に貢献していることがわかる。また、四球数はやはり正の影響とはなっていないため、この解釈でも四球が多くすること（選球眼を鍛えること）は勝率を向上させるための戦略とすべきではないこととケース 1 と一致する結果が得られた。

7 まとめと今後の課題

本研究では、事前に分岐構造を作成することによって、従来モデルに抽象度の違う解釈を与えることに成功した。従来モデルではできなかった視点での解釈が可能となり、より意思決定の支援を可能になる成果である。単純な精度（RMSE）は従来モデルの方が少し良いという結果になったが、分類問題としての精度は提案手法の方が良いものとなった。解釈によって、勝率向上のための指針を作ることができた。しかし、解釈自体に常識と離れたものが出てきている部分がある。異なる分解方法・構造での検証や分割の段階ごとにモデルを変更する等を行っていき、精度の高く解釈のできるモデル考案を進め、意思決定の支援が可能なレベルまでの解釈を付与させたい。

参考文献

- [1] Angelino E. Larus-Stone N. Alabi D. Seltzer M. and Rudin C., Learning certifiably optimal rule lists for categorical data, 00A0The Journal of Machine Learning Research 18(1)(2017)
- [2] DataRobot, SHAP を用いて機械学習モデルを説明する, <https://www.datarobot.com/jp/blog/explain-machine-learning-models-using-shap/>, (閲覧日 2022 年 12 月 2 日)
- [3] Gunning D., Explainable artificial intelligence (xAI), Technical Report, Defense Advanced Research Projects Agency (DARPA) (2017)
- [4] Haykin S, Neural networks and learning machines (3rd ed.), New Jersey:Prentice Hall(2008)
- [5] Lundberg S. and SuIn L., A unified approach to interpreting model predictions, 31st Conference on Neural Information Processing Systems (2017)
- [6] McCabe A and Trevathan J, Artificial intelligence in sports prediction, Information Technology: New Generations ITNG 2008 Fifth International Conference on IEEE(2008)
- [7] Purucker M.C., Neural network quarterbacking, IEEE Potentials(1996)
- [8] Ribeiro M.T. Singh S. and Guestrin C., Why should I trust you? : explaining the predictions of any classifier, arXiv:1602.04938 , (閲覧日 2022 年 12 月 2 日)
- [9] Valero S, Predictiong win-loss outcomes in MLB regular season games - A comparative study using data mining methods, International Journal of Computer Science in Sport 15(2)(2016)
- [10] Vidal T. Pacheco T. and Schiffer M. , Born-Again tree ensembles, the 37th International Conference on Machine Learning (ICML). Vol. 119, pp. 9743-9753 (2020)
- [11] データで楽しむプロ野球,<https://baseballdata.jp>, (閲覧日 2022 年 12 月 30 日)