

投球順番および投球場所を考慮した投手の配球傾向を可視化するシステムの構築

辻野 涼介^{1,a)} 伊藤 正彦^{1,b)}

概要: 野球には様々な戦術が存在し、中でも配球は重要な戦術である。配球は、試合の展開および勝敗に直結する要素である。つまり、投手の配球に焦点を当てた研究が必要となるが、現時点では、投球順番および投球場所を考慮した投手の配球に関する研究はあまり見かけない。そこで、本論文では、配球を投手が打者 1 人に対する投球の組み立てと定義し、投球順番および投球場所を考慮した投手の分析を行う。データから抽出した軌跡データ同士の距離を求め、クラスタリングすることで、投手の配球パターンを可視化するシステムを構築する。結果として、投手ごとの配球傾向の違いを見つけることができた。

1. はじめに

野球には様々な戦術が存在し、その中でも走塁、守備位置、バッティングなどが重要な要素として挙げられる。特に、試合結果や選手の成績に大きな影響を与えるのが「配球」である。配球とは、投手の打者に対する投球の組み立てを指す。投球の組み立てとは、投手が投球する際に使用する投球の戦略やパターンのことを指す。投手がどのような球をどの順番で投げるかは、打者を打ち取るか否か、試合の展開および勝敗に直結する要素であり、その戦術の選択が重要である。

投手の配球が試合の展開および勝敗、選手の成績に影響を与えることを理解した上で、投手の配球に焦点を当てた研究が必要となる。これまでの研究では、投手の成績ごとの配球を分析した研究や特定の捕手に注目した研究は発表されているが、投球順番および投球場所を考慮した投手の配球とクラスタリングを組み合わせた研究はあまり見かけない。

本論文では、投手の配球に焦点を当てる。本論文における配球は投手が打者 1 人に対する投球の組み立てと定義づける。選手、年およびクラスタ数を選択し、可視化した結果を提示するシステムの構築を提案する。データサイト Baseball Savant[1] のデータを用いて、MLB 投手の配球パターンを可視化し、可視化結果を提示するシステムを構築する。

2. 関連研究

辻野ら [2] は、投手の成績に着目し研究を行った。配球を各球種ごとのキャッチャーの構え位置と球種と定義し、キャッチャーが構えた位置の割合をヒートマップで可視化した。ただ、投球順番を考慮していません。結果にあまり差が出なかった。

石橋ら [3] は、配球を球種 (速い球 or 遅い球) × 高さ (高め or 低め) × コース (内 or 外) = 8 分類と定義し、特定のキャッチャーに注目した研究を行った。配球の予測モデルを作成し、モデル予測することで配球を予測した。さらに、モデル予測の解釈によって、各キャッチャーがどのような意図をもって配球していたかを推定した。

Baseball Savant の Pitch-Plinko[4] は、インタラクティブなシステムであり、カウント毎に、球種の割合を円グラフ、ストライクまたはボールをエッジの太さで可視化し、投手のカウントごとの球種の割合、ストライクまたはボールの割合が視覚的にわかるようにしている。

野球におけるインタラクティブな可視化システムとして、守備位置に着目したものがある [5][6]。[5] は各プレイを視覚的に探索し、ゲームの新しい指標を導き出した。[6] は収集されたトラッキングデータをクエリ、フィルタリングにより分析し、ユーザを支援するビジュアル・インターフェイスを作成した。

関連研究との違いとして、[2] は投手に注目した研究という点は共通しているが、配球の定義、1 球ごとの分析をしている点および投球順番を考慮していない点で異なる。本研究では、配球を投手が打者 1 人に対する投球の組み立て

¹ 北海道情報大学

^{a)} s2021167@s.do-johodai.ac.jp

^{b)} imash@do-johodai.ac.jp

と定義し、投球順番および投球場所を考慮した上で、1打席ごとの分析を行うことでより詳細な分析を行う。

[3]は[2]同様、配球の定義、投球順番および投球場所を考慮していない分析という点で異なる。また、特定のキャッチャーに注目した研究という点でも異なる。本研究では、投手に着目した分析を行う。[4]は投球順番は考慮しているが、ボールが投げられた場所は考慮していない点で異なる。本研究では、ボールが投げられた場所も考慮した分析を行う。

3. 提案システム

本研究では、投球順番を軌跡とみなして、クラスタリングにより軌跡のパターン抽出を行う。今回、使用するデータから抽出した軌跡同士の距離を求めて、クラスタリングを行う。クラスタリング結果に基づいて可視化を行い、分析する。さらに、インタラクティブな可視化システムとして、投手名、年およびクラス数を選択項目として設定し、各項目が選択されると投球分析結果を提示するシステムを構築する。

3.1 使用するデータ

データはMLB機構が運営している公式データサイト Baseball Savant[1]のデータを用いる。今回は、2022年シーズンにおける投球数が多いピッチャー20人+大谷翔平の計21人で分析を行う。Baseball Savantのデータの属性は球速や試合日など91項目あるが、今回使用している属性を表1、[1]から引用したゾーン表を図1に示す。ゾーン表とは、打者がボールを打つ際に、投手が投球するゾーンを示したものであり、投手と打者の戦術や相手の弱点を考慮して、投球のバリエーションを持たせるために利用される。[1]のゾーン表は、a~iの9個のストライクゾーン、j~mの4つのボールゾーン計13個のゾーンで構成される。なお、図1のゾーン表はキャッチャー視点であることを注意する必要がある。

表1 使用するデータの属性

属性名	解説
game_date	試合日
batter	打者のMLB Player ID
pitcher	投手のMLB Player ID
zone	ゾーン位置(捕手視点)
p_throws	R=右投手, L=左投手
inning	投球回
inning_topbot	Top=オモテ, Bot=ウラ
game_pk	ゲームID
at_bat_number	試合内の通し打順
pitch_number	打席内の投球数

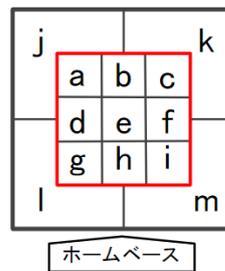


図1 使用するゾーン表([1]より引用)

3.2 投球軌跡データ抽出

宮城らの軌跡からのパターン抽出方法[7]を参考に、[1]のデータから投球データの抽出を行う。これを本論文では、投球軌跡データと名付ける。1打席分の投球データの軌跡を抽出し、図1に基づき、軌跡を文字列化することで投球軌跡データの抽出を行う。例えば、1球目および2球目はゾーンmに投げられ、3球目はゾーンe、4球目ゾーンmの順で投げられた場合、投球軌跡データはmmemとなる(図2)。

今回は軌跡を構成する要素数(1つの打席が完結するまでの投球数)が4または5の投球データを対象に投球軌跡データを抽出する。

投球軌跡データが4または5のデータを使用する理由として、1打席あたりの投球数の平均と分散を算出したとき、平均が4.67、分散が1.67だったため、頻度の高い4または5のデータで分析するのが適切と判断したためである。

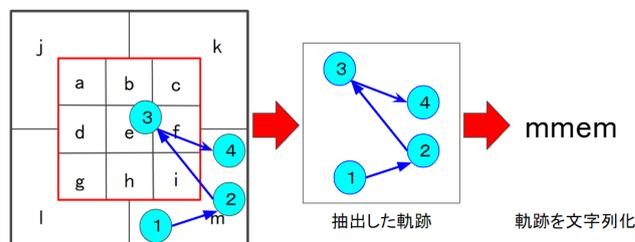


図2 投球データから投球軌跡データを抽出

3.3 投球軌跡データ同士の距離を求め、距離行列を作成

抽出した投球軌跡データ同士の距離はレーベンシュタイン距離を用いて求める。レーベンシュタイン距離とは、とある記号列 S_1 に対して、1文字挿入、削除および置換の3種類の操作をそれぞれ何回行えば記号列 S_2 に変換できるかを調べる手法であり、操作の回数が少ないほど、類似した記号列とみなすことができる。

図3のような投球軌跡データmmemとmmmfjの距離 D を求めることとする。今回の場合1文字目、2文字目は一致していることから、操作は行わない。3文字目、4文字目および5文字目は一致していないため置換、挿入の操作を行う。操作を行った回数は3回であることから、投球軌

跡データ mmem と mmmfj の距離は $D=3$ と求めることができる。

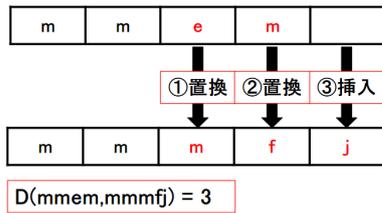


図 3 レーベンシュタイン距離の具体例

今回、抽出した軌跡を構成する要素数 4 または 5 の投球軌跡データ同士の距離をレーベンシュタイン距離で求め、距離行列を作成する。

今回は、レーベンシュタイン距離を何も改変せずに分析するが、今後はゾーン表の位置ごとに操作コストを改変する。例えば、a と m の距離を求めることとする。本来ならば距離は 1 であるが、図 1 のゾーン表に当てはめたときに、ゾーン a とゾーン m 間には物理的な距離があるため、距離は 1 として求めてよいのかという問題が発生する。よって、レーベンシュタイン距離を改変する必要がある。

3.4 3.3 節で求めた距離の値を基にクラスタリング

3.3 節で作成した距離行列を基にクラスタリングを行う。クラスタリングには k-medoids 法を用いる。k-medoids 法は非階層型クラスタリングの 1 つであり、クラスタの中心が点そのものとなり、要素間の距離が与えられれば実行できる。本研究では、距離行列を用いているため k-medoids 法を適用することができる。

この際、最適なクラスタ数の決定に関する問題が発生する。クラスタ数の決定に関して本稿では、エルボー法を用いる。エルボー法とは、クラスタリングの際に最適なクラスタ数を決定するための手法の一つであり、クラスタ数 (k) を変化させながらクラスタリングを行い、その際のクラスタ内誤差 (クラスタ内のデータ点とクラスタ中心との距離の総和) を評価する。クラスタ数に対するクラスタ内誤差をプロットし、エルボーと呼ばれるクラスタ数が増えてもクラスタ内誤差の改善が鈍化するポイントを見つけ、最適なクラスタ数として解釈する。

図 4 は大谷翔平の 2022 年の投球軌跡データにエルボー法を適用したものである。図 4 より、 $k=4$ および $k=5$ にかけて、クラスタ内誤差の値の変化が鈍くなっていることがわかる。よって、大谷翔平の 2022 年の投球軌跡データにおける適切なクラスタ数は 4 と判断することが可能である。

また、他の投手でエルボー法を実施した際、適切なクラスタ数は 5 または 6 と示す結果が出た。よって、本論文ではクラスタ数 4、5 および 6 の 3 種類で分析を行う。

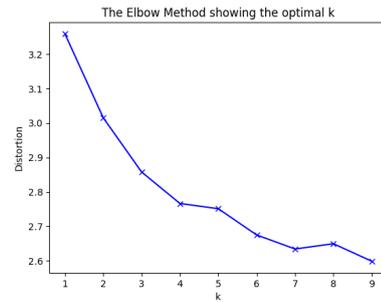


図 4 大谷翔平の 2022 年の投球軌跡データにおけるエルボー法の結果

3.5 クラスタリング結果をネットワーク図で可視化

3.4 節のクラスタリング結果をネットワーク図で可視化する。ネットワーク図で可視化するにあたって、投球軌跡データからノード、エッジ、ノードの頻度およびエッジの頻度を抽出する。

例として、3.2 節 図 2 の投球軌跡データ mmem を用いて説明する。投球軌跡データ mmem からノード、エッジ、ノードの頻度およびエッジの頻度を抽出したものを表 2 に示す。抽出するにあたって、今回はエッジの関係性を見るため無向グラフで分析を行い、「e-m」「m-e」など 2 つのノード間で双方向のエッジが存在する場合は「e-m」としてカウントする。

表 2 投球軌跡データ mmem から抽出されたデータ

ノード	e, m
エッジ	m-m, m-e, e-m
頻度 (ノード)	e:1 回 m:3 回
頻度 (エッジ)	e-m:2 回 m-m:1 回

ノード、エッジなどの抽出が完了したら可視化を行う。ネットワークの可視化には NetworkX^{*1}を使用する。可視化の完成図を図 5 に示す。

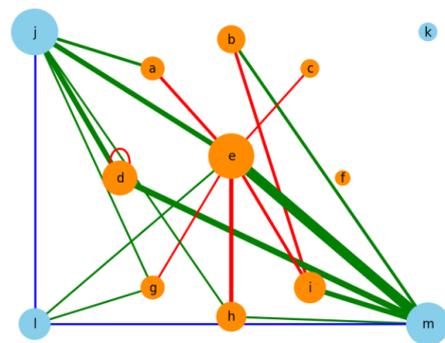


図 5 可視化の完成図

本研究の可視化は、クラスタリング結果ごとに可視化を行う。最初に a~m のノード 13 個を 3.1 節 図 1 のゾーン

^{*1} Python のネットワーク解析ライブラリ <https://networkx.org>

表に基づき、ノードを配置する。ストライクゾーンとボールゾーンを判別するために、今回は a~i のストライクゾーンのノードの色はオレンジ、j~m のボールゾーンは水色に設定した。

ノードを配置したら、可視化を行っていく。可視化するにあたって、頻度によって、ノードの大きさ、エッジの太さ、エッジの色を変える。ノードの大きさに関して、ノードの大きさは $100 \times (\text{ノードの頻度})$ で計算する。NetworkX におけるノードの大きさのデフォルト値は 300 で、ノードの計算式に当てはめて可視化したとき、ノードが大きくなりすぎてしまったため、今回は 100 で設定した。

また、エッジをすべて同じ色で描画したとき、見づらくなってしまったためエッジの色も変更する。ストライクゾーンからストライクゾーンのエッジは赤色、ストライクゾーンからボールゾーンは緑色、ボールゾーンからボールゾーンは青色で描画する。

今回の可視化は 2 つのノード間に関係性が強いのをを見つけることを目的としているため、エッジの頻度が 2 以上のものを描画する。また、ノードの上の輪っかには 2 球以上連続で同じ場所に投げられたことを示す (図 6 内 (a))。

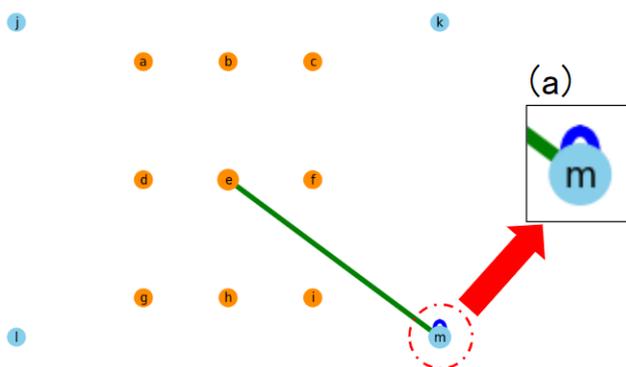


図 6 投球軌跡データ mmem を可視化

図 1 のゾーン表のレイアウトで可視化すると、図 7・左のようにノードとエッジ、エッジとエッジが重なり視認性に欠ける問題が発生する。例えば、緑色のエッジ e-m と青色のエッジ j-m が重なってしまっている。そのため、レイアウトにも気を付ける必要がある。解決策として、図 7・右のようにストライクゾーンのノード b,d,e,f,h をそれぞれ上下左右に動かし、できるだけ重ならないように配置し、視認性の向上を図っている。

4. 探索システムの実装

探索システムは選択項目として、投手名、年およびクラスタ数を用意した。各項目が設定されると、投球分析結果を提示するシステムを構築する (図 8)。

探索システム構築には Dash^{*2}を使用する。まず、投球分

^{*2} Python の Web アプリケーション開発ツール
<https://dash.plotly.com/>

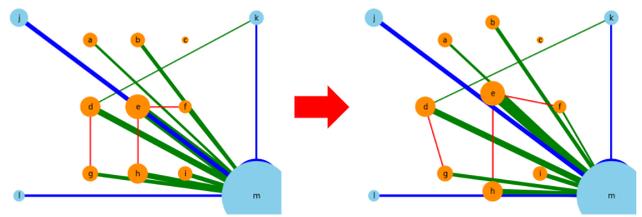


図 7 左:レイアウト変更前 右:レイアウト変更後

析したい投手をプルダウンメニューから選択し、年およびクラスタ数をそれぞれラジオボタンから選択する (図 9)。

3 つの項目が選択されたら、選択された選手、年およびクラスタ数をもとに 3 章の一連の処理を実行した結果を表示させるシステムを構築した。

5. 分析・考察

大谷翔平投手、ダルビッシュ有投手の 2 投手で分析を行い、比較した。今回は投手ごとに左打者、右打者および右左のデータすべての計 3 パターン、クラスタ数は 4 で分析を行った。2 投手のデータ数は表 3、それぞれのパターンで可視化したものを図 10、図 11 および図 12 に示す。以下、左、右および両方の順で分析を行っていく。また、可視化したネットワーク図はキャッチャー目線であることに注意する必要がある。

表 3 今回使用する投球軌跡データの数	
投手名	データ数
大谷 翔平	総データ：253
	左打者のデータ：111
	右打者のデータ：142
ダルビッシュ 有	総データ：274
	左打者のデータ：159
	右打者のデータ：115

5.1 左打者

大谷 (図 10・左) の class0 から左打者から遠い場所 (以下、外角) を攻めていることがわかり、特にゾーン j とストライクゾーンとの関係性が高いことが読み取れる。class1 からはゾーン m を絡めた配球をしていること、ゾーン m とボールゾーンとの関係性が高いことが読み取れる。

一方、ダルビッシュ (図 10・右) の class0 から大谷とは少し違う傾向が見えた。大谷は、赤いエッジがあまり無いが、ダルビッシュは、赤いエッジが多い、かつ太いことからストライクゾーンからストライクゾーンとの関係性が高いことがわかる。よって、ダルビッシュは左打者に対して、高低をうまく使った配球をしていることがわかる。また、ゾーン j およびゾーン m のノードの大きさが似ているため、コーナーもうまく使った配球をしていることもわかる。class1 は大谷の class0 と似た左打者の外角を攻める配球傾向を得られた。

MLB Pitcher Pitching Pattern Visualization System

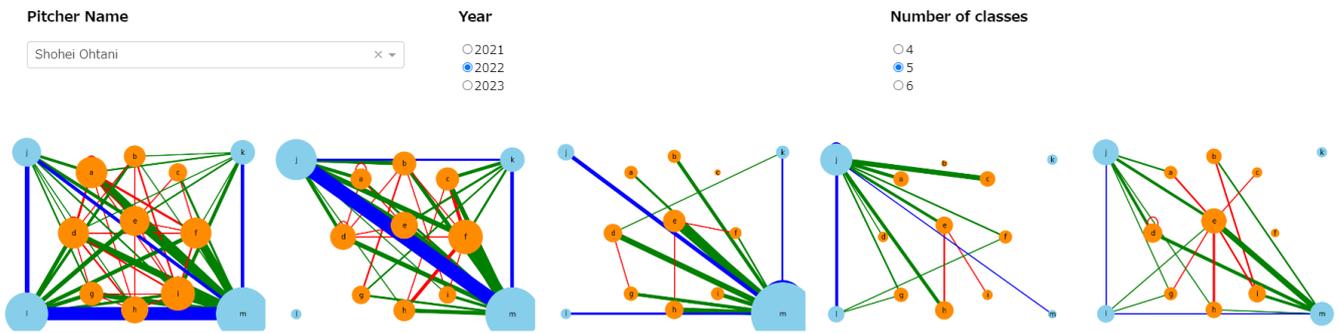


図 8 探索システム全体像

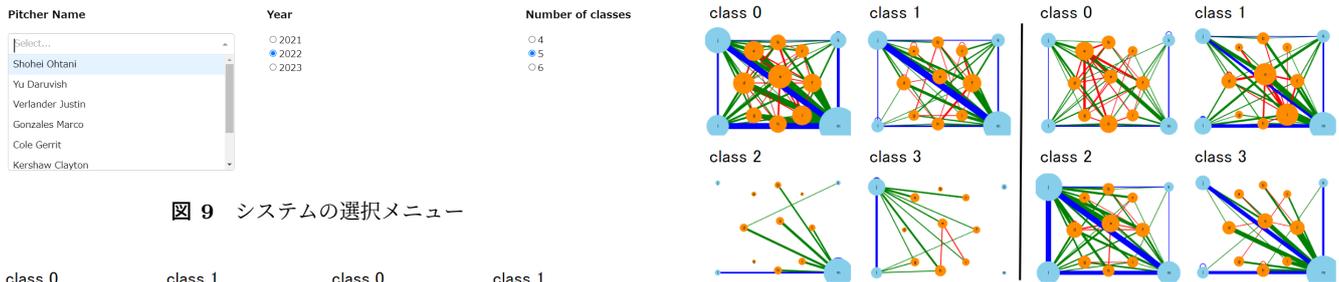


図 9 システムの選択メニュー

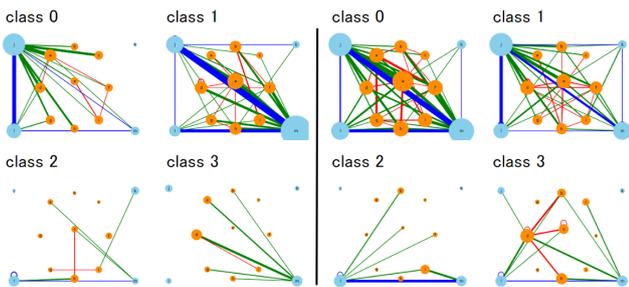


図 10 左打者のデータ (左: 大谷翔平, 右: ダルビッシュ有)

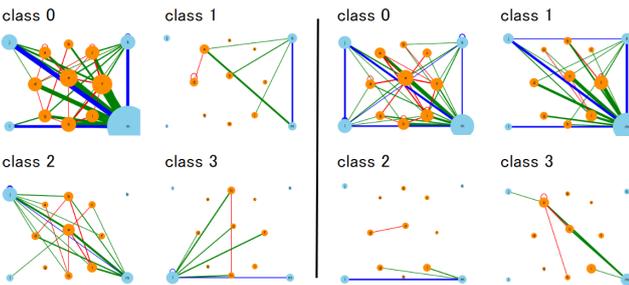


図 11 右打者のデータ (左: 大谷翔平, 右: ダルビッシュ有)

5.2 右打者

大谷 (図 11・左) は class0 より右打者の外角低め中心の配球をしていることがわかる。また、ゾーン M との関係性がエッジの太さから非常に強いことも読み取れる。これは、大谷の代名詞ともいわれるスリーパー*3という変化球が関係していると考えられる。

一方、ダルビッシュ (図 11・右) の class0 および class1 と大谷の class0 を比べて、右打者の内角の 2 つのボール

*3 スライダーの一種で、大きく横に変化する変化球

図 12 右左合わせたデータ (左: 大谷翔平, 右: ダルビッシュ有)

ゾーン間のエッジがあるか、否か以外はそこまで差がない。よって、大谷とダルビッシュの右打者に対する配球傾向は似ているといえる。

5.3 すべてのデータ

大谷 (図 12・左) は class0 よりゾーン d,g,h,i,j,l および m の左下のゾーンを使う投球パターン、class1 よりゾーン a,b,c,f,j,k および m の右上のゾーンを使う投球パターンの主に 2 種類の投球パターンが存在する。大谷は本格派ピッチャーと呼ばれ、威力のある速球と優れた変化球、豊富なスタミナとコントロールを高いレベルで兼ね備えた投手である。彼の武器である 160km を超えるストレートと、スプリット*4・スリーパーなどの変化球があれば、「わかっていても打てない」という打者心理を生み出している可能性があると考えられる。

一方、ダルビッシュ (図 12・右) は class0~class3 の結果を通して、さまざまな投球パターンが存在することがわかる。ダルビッシュは技巧派のピッチャーと言われており、多彩な球種、緩急、内外、タテヨコ、出し入れを自在に使いこなすピッチャーである。これらによって、ダルビッシュの投球パターンに幅を生み、様々な投球パターンが存在していると考えられる。

*4 打者の手元で落ちる変化球

6. おわりに

本論文では、MLB の Baseball Savant のデータを用いて、投球順番と場所考慮した投球分析を行い、可視化するシステムを構築した。これらの結果を用いて、今後アマチュア向けのテンプレートが作成できるのではないかと考えている。

今後の課題として、ノードの配置などのレイアウトの問題、今回は無向グラフで分析したが、有効グラフにした場合の見せ方、球種、球速および結果などの属性を加えてパターン抽出をしたときの可視化方法の検討、レーベンシュタイン距離を改変するという課題がある。

参考文献

- [1] Baseball Savant, <https://baseballsavant.mlb.com/>
- [2] 辻野涼介, 豊田響希, 伊藤正彦: 投手と打者の成績区分ごとの配球位置の傾向に関する視覚的分析, 情処全国大会, 2022.
- [3] 石橋克也, 唐恵東, 蔣帥, 兪樺, 亀井清華, 森本康彦: SHAP を用いた MLB の配球分析、第 15 回データ工学と情報マネジメントに関するフォーラム (DEIM2023).
- [4] <https://baseballsavant.mlb.com/visuals/pitch-plinko>
- [5] Carlos Dietrich, David Koop, Huy T. Vo, and Claudio T. Silva: Baseball4D: A Tool for Baseball Game Reconstruction & Visualization, IEEE VAST 2014, 2014.
- [6] Marcos Lage, et al.: StatCast Dashboard : Exploration of Spatiotemporal Baseball Data, IEEE Computer Graphics and Applications 36(5), pp.28–37, 2016
- [7] 宮城優里, 大西正輝, 渡辺知恵美, 伊藤貴之, 高塚正浩: 記号化された歩行経路群のパターン抽出と可視化, 可視化情報学会論文誌, 38(3), 2018.