

P3VS：打席結果別の投手の配球パターン比較可視化アプリケーション

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。著者らは、配球を投手が打者 1 人に対して行う投球の組み立てと定義し、視聴者が野球の試合を見ながら、次の配球を対話的に分析し、視聴者自身が次の配球を予測することを支援できるシステムを構築してきた。本論文では、現状の可視化システムを、三振やホームランなどの打席結果ごとに分割して表示可能とするシステムへと改良する。このシステムにより、打席結果ごとの配球パターンの特徴的な傾向を把握することを目指す。

1. はじめに

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。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。これまでの研究では、投手の配球に着目した研究は多く発表されているが、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。また、野手に着目したシステムや大量のビデオから探索するシステムは発表されているが、投手の配球に関して、ユーザーの配球予測を支援するツールはあまり見かけない。

視聴者が配球などの戦術を理解することは試合における戦略や結果などを理解することにつながる。これにより、投手と打者の駆け引きなどを深く理解することが可能になる。

著者らは、投手の配球に焦点を当て、配球を投手が打者 1 人に対して行う投球の組み立てと定義し、視聴者が野球の試合を見ながら、次の配球を対話的に分析し、視聴者自身が次の配球を予測することを支援できるシステムを構築してきた [1]。

本論文では、[1] の可視化システムを、可視化結果の表示画面を打席結果 (三振、ホームランなど) ごとに分割して表示可能とするシステムへと改良する。打席結果ごとに分割して可視化することで、各打席結果に関連性が強いゾーンを特定することが可能となる。このシステムによって、視

聴者が野球観戦をさらに楽しめるような仕組みの構築を目指す。

2. 関連研究

野球の分析においては、多くの研究で投手の配球に着目している [2][3]。辻野ら [2] は、配球をキャッチャーの構えた位置と球種と定義し、位置の割合をヒートマップで可視化することで、投手の成績による差異が出るかを分析した。石橋ら [3] は、配球の予測モデルを作成し、配球を予測した。しかしながら、これらの研究では 1 球ごとの投球場所の順番を考慮した対話的な分析は行っていない。

野球におけるインタラクティブな可視化システムの研究としては、守備位置に着目したものがある [4][5]。しかしながら、投手の配球に関して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研究事例はあまり存在しない。

野球の視聴支援をするための多くのシステムが開発されている。特に動画に着目したものが多く、審判カメラ [6]、2022 年に日本テレビが導入したポリュメトリックビデオ (自由視点映像) [7] がある。これらはリプレイのための仕組みであり、本研究とは目的が異なる。

Video Lens [8] は、大量のビデオから投球位置情報などを視覚化し、インタラクティブに動画を探索できる仕組みを構築した。本研究と異なり、Video Lens は一球ごとの順番を考慮した探索は行っていない。

ESPN の Play-by-Play [9] および Sportsnavi の一球速報 [10] は試合中の各プレーの詳細をリアルタイムで提供するサービスである。これらにより、試合を見れない状況でも一球ごとの詳細を確認できるようになった。しかしながら、試合を視聴しながら次の配球を考える手助けをする仕

¹ 北海道情報大学

^{a)} s2481107@s.do-johodai.ac.jp

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

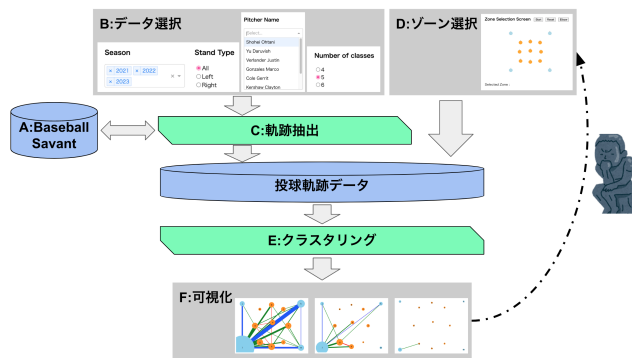


図 1 P3VS の全体像

組みは提供していない。

Baseball Savant の Pitcher-Plinko [11] 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。Pitch-Plinko 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本研究とは異なる。

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を抽出し、その可視化結果をもとにユーザー自身が考え、さらに別のゾーンを選択し、より深い探索を行うことができるシステムを構築する。著者らはこの提案システムを「P3VS」(Pitcher Pitching Patterns Visualization System) と名付けた。加えて、可視化結果の表示画面を打席結果ごとに分割して表示可能とするシステムへと改良する。さらに、P3VS を Web サーバー上で動かし、タブレット端末上で探索および分析できるようにする。P3VS の全体像を図 1 に示す。

まず、最初に投手名、シーズン、打者の左右、クラスタ数を選び、選択項目に応じて該当するデータを取得する(図 1・A および図 1・B)。該当するデータから、1 打席ごとの投球データの軌跡を抽出する(図 1・C)。ゾーン選択画面をクリックすることで、ゾーンを選択し、該当する軌跡のみを抽出する(図 1・D)。抽出された軌跡同士でクラスタリングを行う(図 1・E)。クラスタリング結果ごとに可視化を行い、ユーザー自身が分析を行う(図 1・F)。

システムの構築には Dash^{*1} と Dash Cytoscape^{*2} を使用する。提案システムのインタフェース全体図を図 2 に示す。

以降、この章では図 1 内にふられている番号の順に説明を行っていく。

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 [12] のデータを用いる。なお、本研究における [12] のデータは別ファイルに csv ファイルの形式で保存されている。

Baseball Savant のデータの属性は球速や試合日など 91 項目があるが、今回使用している属性を表 1、Baseball Savant で使用されているゾーン表を図 3・左に示す。表 1 の zone とは、ボールがホームプレートを通過するときのゾーン位置を指す。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、ボールがどこを通過したかを判別できるものである。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、Baseball Savant ではキャッチャー視点でデータを提供しているが、本研究では視聴者が分析しやすくするために、投手視点に反転している。また、表 1 の events は、三振やホームランなどその打席の結果が格納されている。

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
stand	R=右打者, L=左打者
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
events	投球イベントの結果

3.2 データ選択

まず、最初に視聴者が任意の条件で分析できるようにデータを選択する(図 2・A)。投手名 (Pitcher Name) とシーズン (Season) をプルダウンメニューから選択し、右打者あるいは左打者 (Stand Type)、クラスタ数 (Number of classes) をラジオボタンで選択する。なお、シーズンは複数選択することが可能である。各項目が選択されると、3.1 節のデータから選択項目に該当するデータを取得する。

3.3 投球軌跡データ抽出

Miyagi らの軌跡からのパターン抽出方法 [13] を参考に、3.1 節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、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

^{*1} Python の Web アプリケーション開発ツール
<https://dash.plotly.com/>

^{*2} ネットワークを可視化するための Dash のコンポーネント
<https://dash.plotly.com/cytoscape>

MLB Pitcher Pitching Pattern Visualization System

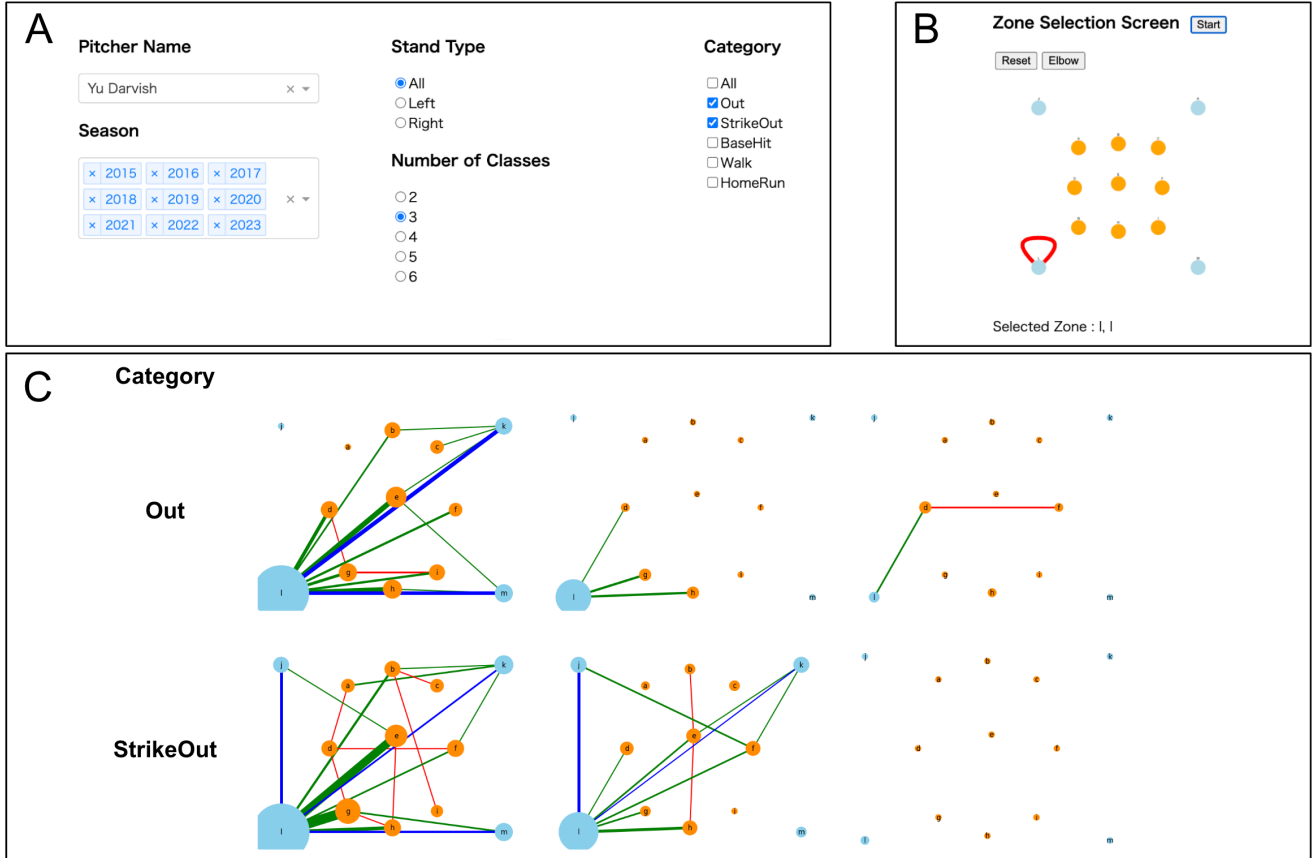


図 2 P3VS のインターフェース全体図

A：データ選択画面，B：ゾーン選択画面，C：可視化結果表示画面

げられた場合、投球軌跡データは mmem となる (図 3・右)。

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 1 以外の投球データを対象に投球軌跡データを抽出する。

3.4 ゾーン選択

ユーザーが任意の投球順番で分析できるようにゾーンを選択する機能を作った (図 2・B)。ゾーンが選択されると、3.3 節で抽出した投球軌跡データから該当する投球軌跡データのみを抽出する。

ゾーン選択画面の実装には Dash Cytoscape を用いる。a~m の 13 個のノードが図 3・左のゾーン表に基づいて配置されている (図 2・B)。ストライクゾーンとボールゾーン

を区別するために、a~i のストライクゾーンのノードの色はオレンジ、j~m のボールゾーンのノードの色は水色に設定する。ノードが選択されたら、3.3 節で抽出した投球軌跡データから選択したゾーンに該当する投球軌跡データのみを抽出する。フィルタリングした後の投球軌跡データの件数を取得し、件数を太さにしたエッジをゾーン選択画面に出力する。

図 2 の場合はノードを 2 つ選択した例で、1 球目および 2 球目にゾーン「l」を選択したパターンである。「ll」で始まる投球軌跡データを抽出し、ゾーン選択画面に該当する投球軌跡データの件数を太さにしたエッジ「l-l」を出力している (図 2・B)。図 4 は最初にゾーン「l」、次にゾーン「g」を選択した場合である、「lg」で始まる投球軌跡データを抽出し、ゾーン選択画面に該当する投球軌跡データの件数を太さにしたエッジ「l-g」を出力する (図 4・右)。

最後に、ゾーン選択画面の横に配置している 3 つのボタンを説明する。Start ボタンは、クリックすることで解析を開始する。Reset ボタンは、クリックすることでゾーン選択画面を選択する前の状態に戻ることができる (図 4・左)。Elbow ボタンは、ユーザーがクラス数に迷ったとき際に、クリックすることでエルボー法の結果を表示

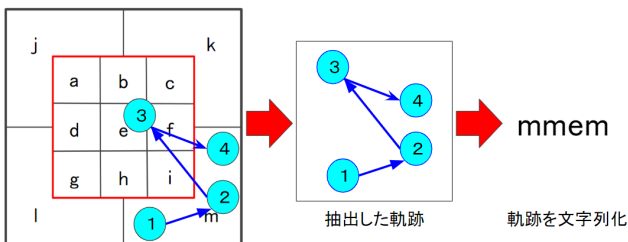


図 3 投球データから投球軌跡データを抽出

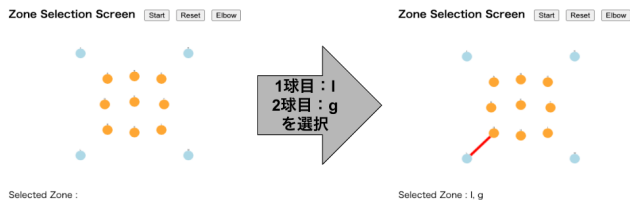


図 4 ゾーン選択画面の具体例

し、ユーザーの探索を支援する (図 8(g)). エルボー法とは、クラスタ数を決定するとき使用される手法である。

3.5 クラスタリング

クラスタリングをするため、抽出した投球軌跡データ同士の距離を求める。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 5 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。

今回、抽出した軌跡を構成する要素数 1 以外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。

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。

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 3・左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。

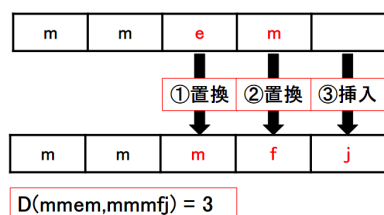


図 5 レーベンシュタイン距離の具体例

よって、レーベンシュタイン距離を改変する必要がある。

3.6 クラスタリング結果をネットワーク図で可視化

3.5 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、3.3 節 図 3 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^{*3}を使用する。可視化の完成図を図 2・C に示す。

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 3.4 節 図 2・B と同様にノードを配置する。

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、ノード間の関係性を明確にするために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 つのノード間で関係性が強い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、ノードの上の自己ループは 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))。なお、以降、本稿の事例では自己ループは見せないこととする。

図 3・左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

^{*3} Python のネットワーク解析ライブラリ
<https://networkx.org>

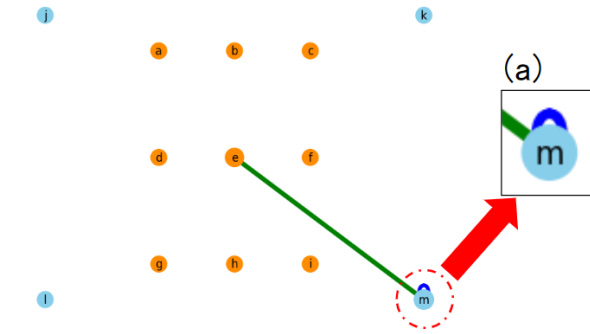


図 6 投球軌跡データ mmem を可視化

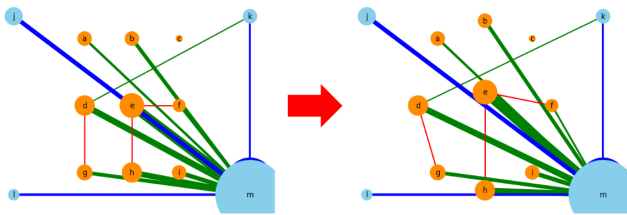


図 7 左: レイアウト変更前 右: レイアウト変更後

青色のエッジ j-m が重なってしまっている．そのため，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．

3.7 打席結果ごとに可視化

P3VS は，打席結果ごとにクラスタリングし，可視化することも可能である．これにより，投球パターンの特徴的な傾向を把握することを目指す．今回は，表 1 の events 属性に格納された打席結果を表 3 のように 5 つの打席カテゴリに分類した．打席結果ごとの可視化も 3.6 節同様ネットワーク図で可視化を行う．

また，打席カテゴリ (Category) は，チェックボックスで追加することが可能である (図 2・A)．図 8 は打席カテゴリを「Out」および「StrikeOut」と選択し，可視化した例である．この場合は，「1」で始まる投球軌跡データの中から，打席カテゴリに該当する投球軌跡データを抽出し，それぞれの打席カテゴリごとにクラスタリングおよび可視化を行なっている．

表 3 表 1 events 属性に含まれる打席結果の分類

打席カテゴリ	含まれる打席結果
StrikeOut	strikeout
Out	field_out, double_play, force_out など
BaseHit	single, double, triple
Walk	walk, hit_by_pitch
HomeRun	home_run

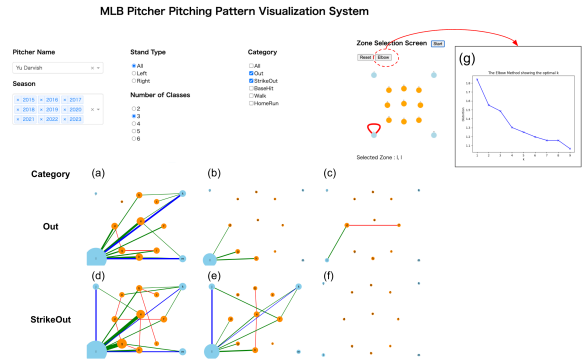


図 8 ダルビッシュ有 (1 球目: ゾーン 1, 2 球目: ゾーン 1, 打席カテゴリ: Out, StrikeOut)

3.8 P3VS を Web サーバー上で動作可能にする

P3VS を広く利用可能にするためには，Web サーバー上での動作環境を整備する必要がある．本研究では，Heroku^{*4}を Web サーバーとして活用し，タブレット端末での探索および分析を実現する (図 9)．これにより，複数人が同時に操作可能な環境が構築され，野球の視聴体験が向上すると考える．

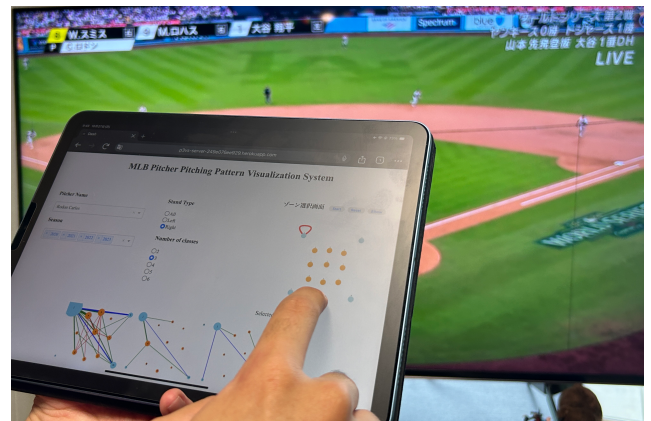


図 9 実際の試合視聴におけるシステム利用例

4. 探索事例

今回は，2018 年から 2023 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，ダルビッシュ有の計 22 人で分析を行った．その中で，2015 年から 2023 年におけるダルビッシュ有^{*5}投手に関し，表 4 に示される最多の組み合わせである 1 球目および 2 球目がゾーン 1 となるパターンを対象に，クラスタ数を 3 に設定し，打席カテゴリ「Out」および「StrikeOut」で分析を行った例を紹介する (図 8)．

今回は説明の都合上，クラスタリング結果の左上に (a)～(f) のラベル付を行っている．

^{*4} <https://jp.heroku.com/home>

^{*5} 現在 San Diego Padres

表 4 今回使用する投球軌跡データの数

投手名	データ数
ダルビッシュ 有	総データ：2901 「1I」で始まるデータ：178 打席カテゴリー「Out」のデータ数：55 打席カテゴリー「StrikeOut」のデータ数：55

4.1 打席カテゴリー：Out(三振以外のアウト)

図 8(a) より，エッジの太さからゾーン 1 に近いストライクゾーン (e, h) に投げることが多いのではないかと読み取れる。また，ゾーン k との関係性もあることから，コーナーを活用した配球をしていることも読み取れる。

このパターンでは，ほとんどがゾーン 1 との関連性を示しており，全体としてシンプルな配球傾向があることが読み取れる。

4.2 打席カテゴリー：StrikeOut(三振)

図 8(d) より，ノードの大きさおよびエッジの太さからゾーン 1 に近いストライクゾーン (e, g) との関係性が強いことが読み取れる。

また，Out の場合の配球パターンと比較して，異なる点がいくつか存在する。まず 1 点目として，Out のときよりもネットワーク図が複雑である点が挙げられる。Out の場合は，ゾーン 1 から出ているエッジが少ないが，StrikeOut の場合はゾーン 1 から出ているエッジが多くなっていることがわかる。さらに，赤や緑のエッジが多く見られることから，StrikeOut の場合には多様な配球パターンが存在するのではないかと推察することができる。

2 点目として，Out の場合はコーナーを活用した配球をしていることが読み取れるが，StrikeOut の場合はゾーン k との関係より，コーナーに加えて高さも活用した配球をしていることが読み取れる。

5. おわりに

本論文では，MLB 公式データサイト Baseball Savant のデータを用いて，投手の配球傾向を可視化するシステムを構築した。このシステムにより，視聴者が野球の試合を見ながら，次の配球を対話的に分析し，視聴者自身が次の配球を予測することが可能となり，野球の試合における視聴体験が向上すると考える。

今後はこのシステムを使用しながら試合を見ることで，「野球の試合における視聴体験が向上するか」や「ユーザー自身の予測において支援できたか」などの観点でユーザー評価を行う。

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，有向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。また，システム

面の課題として，実際の投球間隔に追いつけず，次の配球についてユーザーが考える時間が少ないという問題や球がどこに投げられたかわからないという問題も存在する。

謝辞 本研究は，一般社団法人 新雪の北海道 IT クリエイター発掘・育成事業 (新雪プログラム) に支援をいただき，公立はこだて未来大学名誉教授 美馬義亮先生にアドバイスをいただき実施いたしました。心より感謝申し上げます。

参考文献

- [1] 辻野涼介, 伊藤正彦: P3VS: 野球の視聴者視点で配球戦略を分析するための対話的可視化システムの提案, *EC2024* (2024).
- [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, *情報処理学会 第 85 回全国大会* (2023).
- [3] 石橋克也, 唐惠東, 蔣帥, 俞樺, 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) (2023).
- [4] Dietrich, C., Koop, D., Vo, H. T. and Silva, C. T.: Baseball4D: A Tool for Baseball Game Reconstruction Visualization.
- [5] Lage, M., Ono, J. P., Cervone, D., Chiang, J., Dietrich, C. and Silva, C. T.: StatCast Dashboard: Exploration of Spatiotemporal Baseball Data, *IEEE Computer Graphics and Applications*, Vol. 36(5), pp. 28–37 (2016).
- [6] NiTRO: 更なる進化を遂げた「野球審判カメラ」, https://www.nitro.co.jp/report/baseballcamera_230901.html (参照 2024-12-21).
- [7] 内田 泰: 大リーグの先行く巨人戦、キャノンがカメラ 98 台の自由視点映像を 3 秒で生成, <https://xtech.nikkei.com/atcl/nxt/column/18/00001/07936/> (参照 2024-12-21).
- [8] Matejka, J., Grossman, T. and Fitzmaurice, G.: Video Lens: Rapid Playback and Exploration of Large Video Collections and Associated Metadata, *UIST'14* (2014).
- [9] ESPN Enterprises, Inc.: MLB on ESPN, <https://www.espn.com/mlb/> (参照 2024-12-21).
- [10] スポーツナビ: Sportsnavi, <https://baseball.yahoo.co.jp/npb/> (参照 2024-12-21).
- [11] MLB Advanced Media, LP.: "Pitcher Plinko", "<https://baseballsavant.mlb.com/visuals/pitch-plinko>" (参照 2024-12-21).
- [12] MLB Advanced Media, LP.: Baseball Savant, <https://baseballsavant.mlb.com/> (参照 2024-12-21).
- [13] Miyagi, Y., Onishi, M., Watanabe, C., Itoh, T. and Takatsuka, M.: Classification and visualization for symbolic people flow data, *J. Vis. Lang. Comput.*, Vol. 43, pp. 91–102 (2017).