

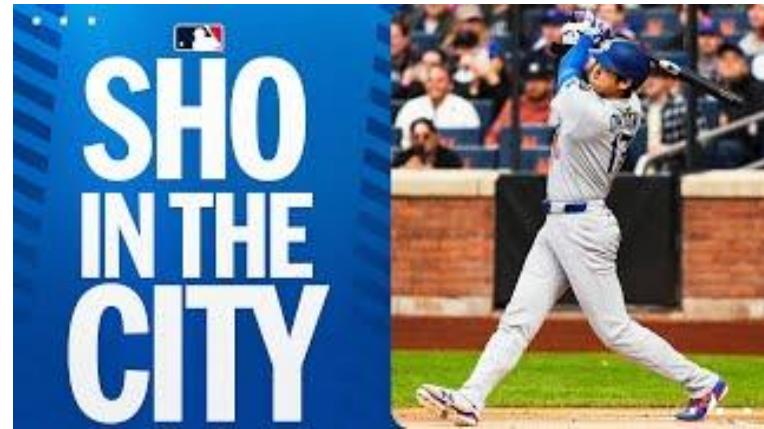


P3VS: A Visual Analytics System for Structural Analysis of Pitching Sequence Trajectories

Ryosuke Tsujino, Masahiko Itoh
Hokkaido Information University, Japan

What is Baseball?

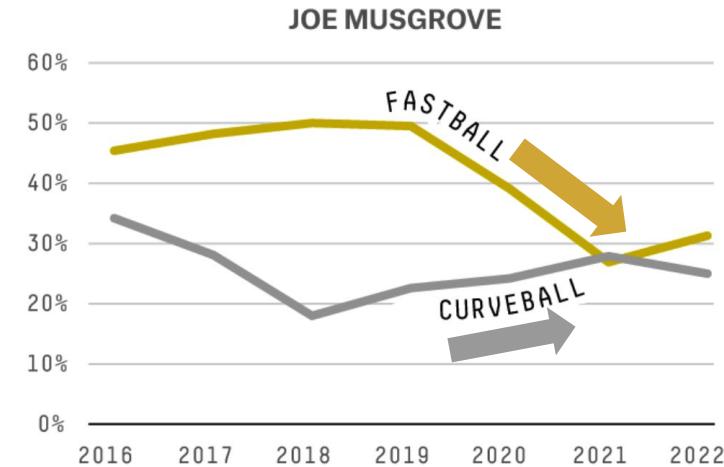
- One of the most popular sports worldwide
 - Especially in the US, Japan, Korea, and Latin America
- A pitcher throws a ball, a batter hits it
- Strategy is important
 - Pitching order is the most important



Shohei Ohtani goes deep off Kodai Senga 🎉
<https://youtu.be/afDsQahOEYM?feature=shared>

Background

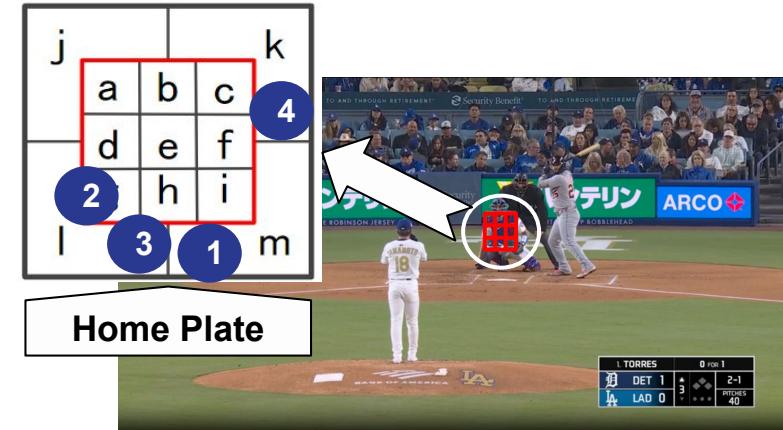
- Importance of **Pitch Sequence**
 - Deciding what pitches to throw and in what order is crucial for pitchers.
- The evolution of modern baseball
 - New pitch types are emerging
 - Fastballs are getting faster
 - Strategies becoming more complex



R. O'Connell, "Why Some MLB Pitchers Are Abandoning the Fastball," FiveThirtyEight, 2022. Available:
[https://fivethirtyeight.com/features/why-some-mlb-pitchers-a
re-abandoning-the-fastball/](https://fivethirtyeight.com/features/why-some-mlb-pitchers-are-abandoning-the-fastball/)

Audience Understanding of Tactics

- Audience understanding of tactics
 - helps the audience see the game more deeply



Pitch No	Pitch	Type	MPH
1	Ball	Curveball	77.8
2	Called Strike	4-Seam Fastball	94.6
3	Ball	Split-Finger	91.1
4	HomeRun!!	Sinker	94.4

What is Pitch Sequence?

- Pitch Sequence
 - A strategic plan of what pitches to throw and in what order

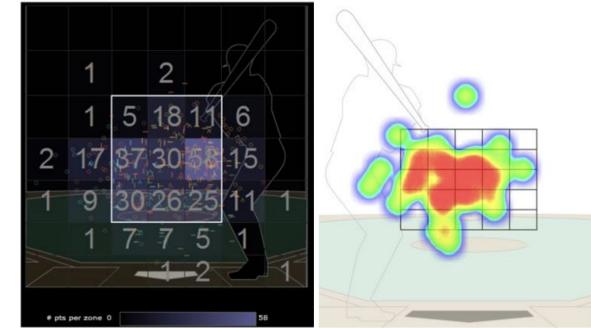
PITCH	TYPE	MPH		
1 Foul Ball	Four-seam FB	94		
2 Foul Ball	Sinker	94		
3 Strike Swinging	Slider	86		
▼ Tauchman fouled out to left.				1 6
				0 RUNS, 0 HITS, 0 ERRORS

ESPN Yoshinobu Yamamoto Play-by-Play

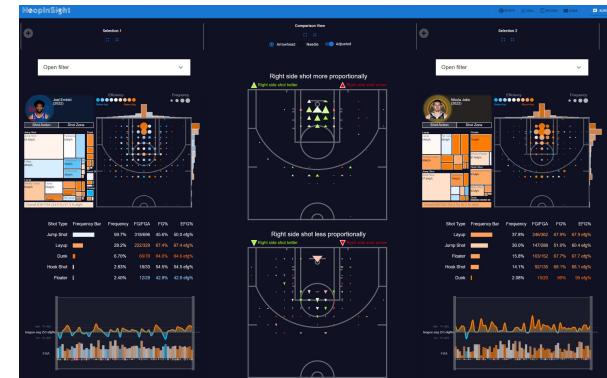
https://www.espn.com/mlb/playbyplay/_gamelid/401696190

Related Works

1. Pitch Analysis Tools
 - Focus on pitch frequency and location
2. Pitch Sequence Analysis
 - Show importance of order,
but often short or simple
3. Tactical Visualization in Other Sports
 - Time series and structural pattern analysis, not yet applied to baseball



B. Moon and R. Brath. Bloomberg sports visualization for pitch analysis. IEEE The 1st Workshop on Sports Data Visualization, 2013. 2

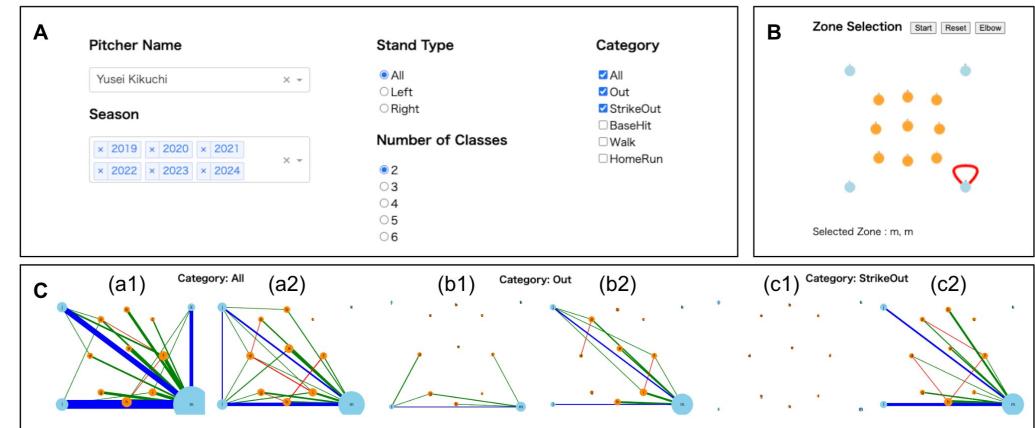
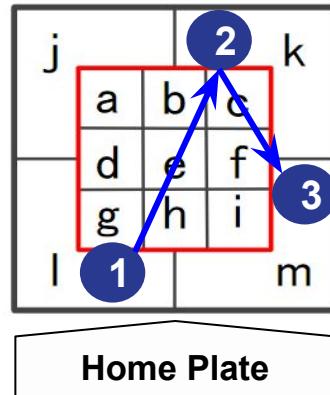


Y. Fu and J. Stasko, "HoopInsight: Analyzing and comparing basketball shooting performance through visualization," in Proc. VIS2023, Oct. 2023.

Research Purpose

P3VS (*Pitcher Pitching Patterns Visualization System*)

- Treats pitch sequences as trajectories to extract data
- Uses clustering and visualization to analyze and compare strategic patterns



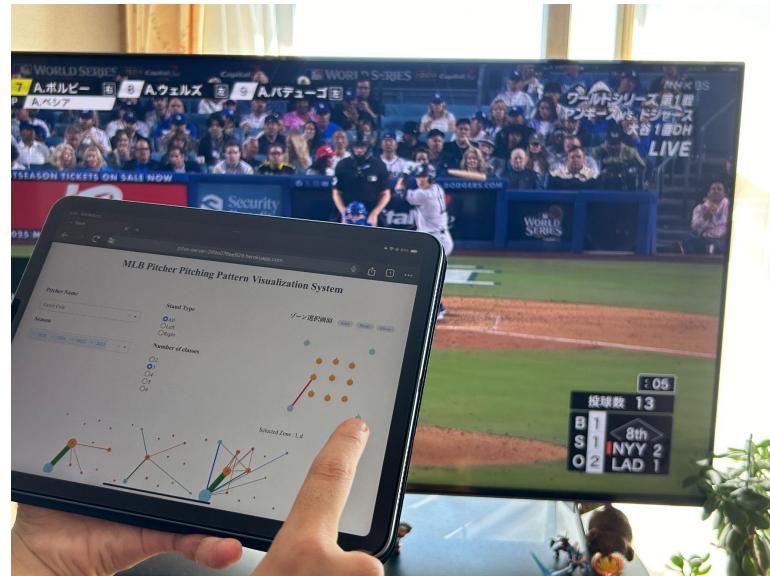
Proposed System - Requirements -

8

Designed for audience, analysts,
and baseball professionals

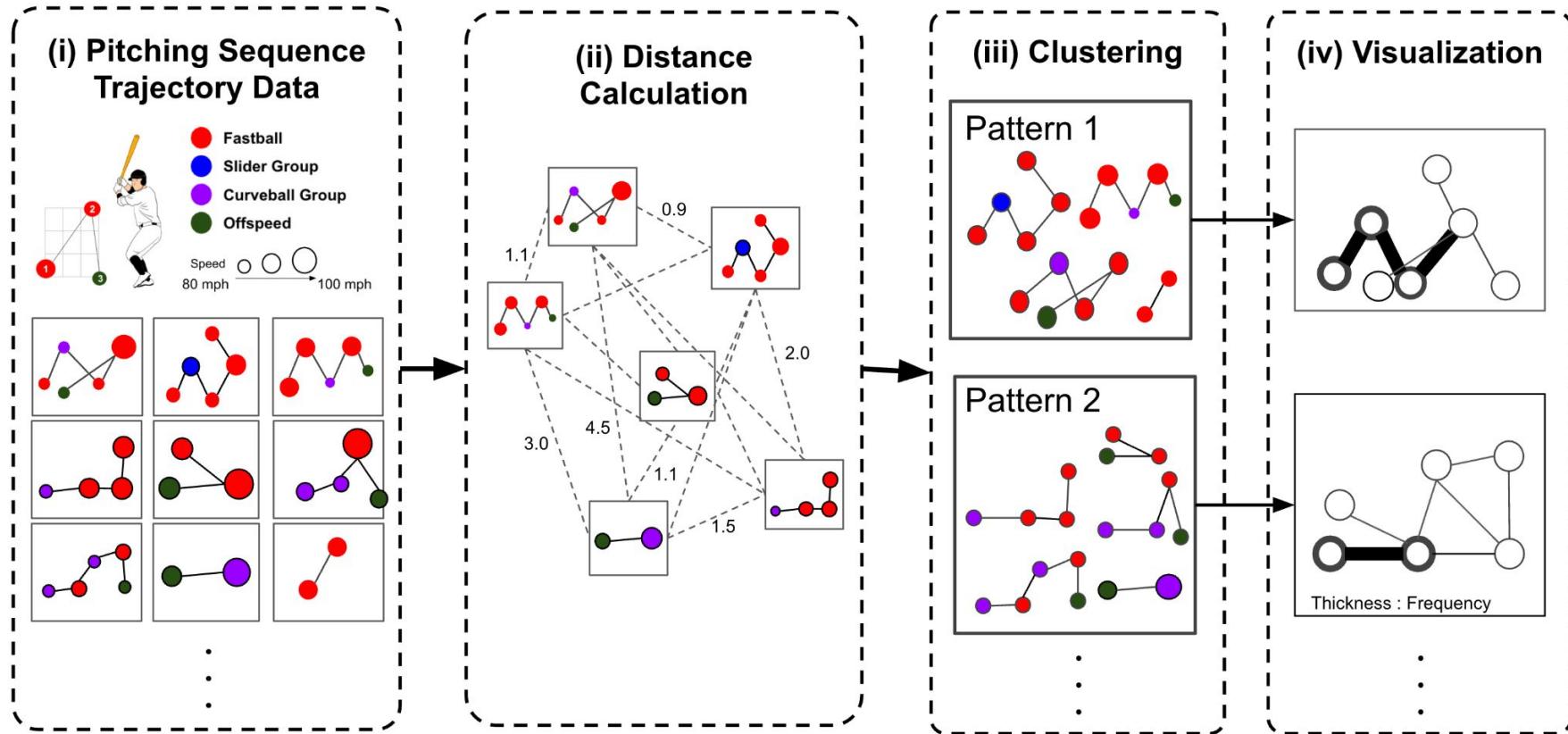
Requirements

- **R1:** Zone-based pitch filtering
- **R2:** Pattern discovery from sequences
- **R3:** Outcome-based comparison
- **R4:** Real-time interaction and updates
- **R5:** Advanced filtering options



Overview of P3VS

9



Pitch Focused Dataset

- This study focuses on pitch level data provided by Baseball Savant*
- We use 15 key attributes



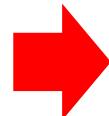
SELECTED ATTRIBUTES FROM THE BASEBALL SAVANT DATASET.

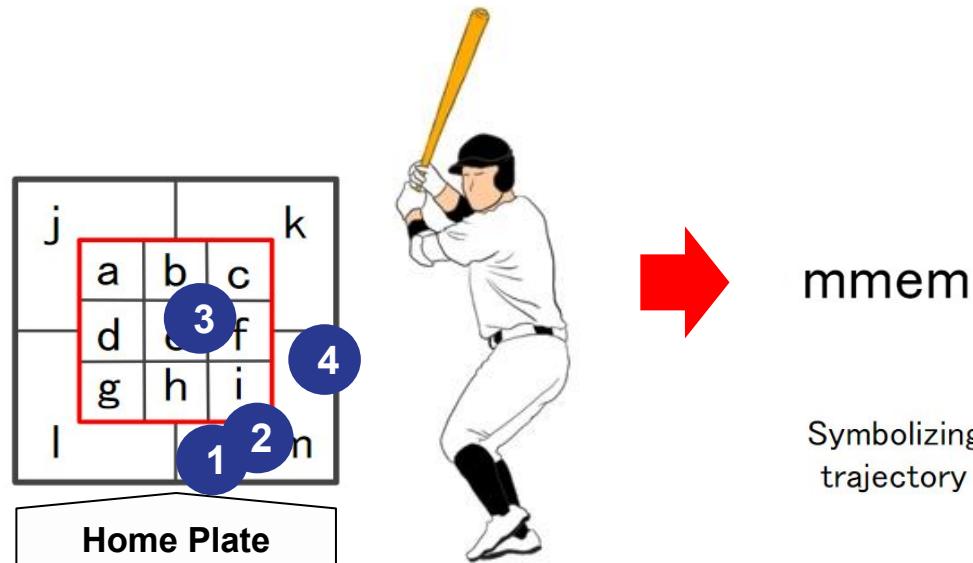
Attribute	Description
game_date	Date of the game
batter	MLB player ID of the batter
pitcher	MLB player ID of the pitcher
zone	Zone where the pitch crossed home plate
stand	Batter's stance (R = Right-handed, L = Left-handed)
inning	Inning number
inning_topbot	Half-inning indicator (Top or Bottom)
effective_speed	Pitch speed at the plate (mph), adjusted for extension
pitch_name	Pitch type (e.g., Four-Seam Fastball, Slider)
px	Horizontal movement (inches)
py	Vertical movement (inches)
game_pk	Unique game identifier
at_bat_number	At-bat number within the game
pitch_number	Pitch number within the at-bat
events	At-bat result (e.g., strikeout, single)

* An official data site operated by MLB.

Pitching Sequence Trajectory Data Extraction 11

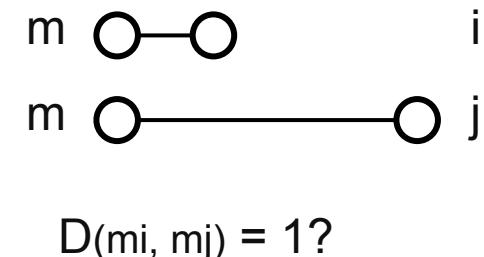
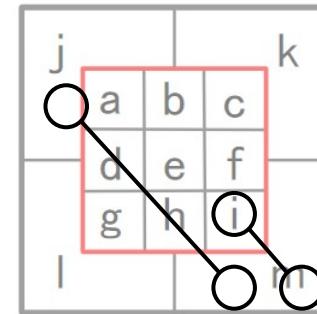
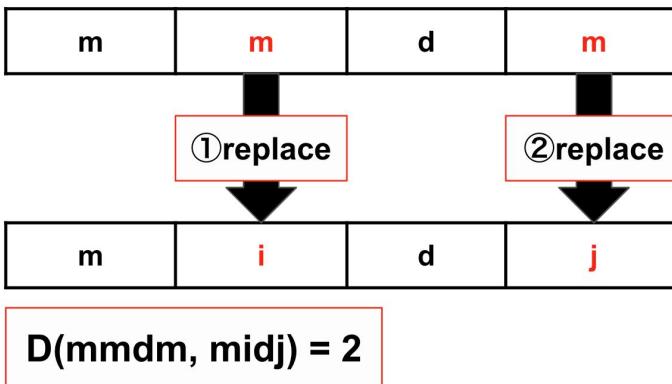
- Treats pitch sequences as trajectories to extract data

 Named “**Pitch Sequence Trajectory Data**”



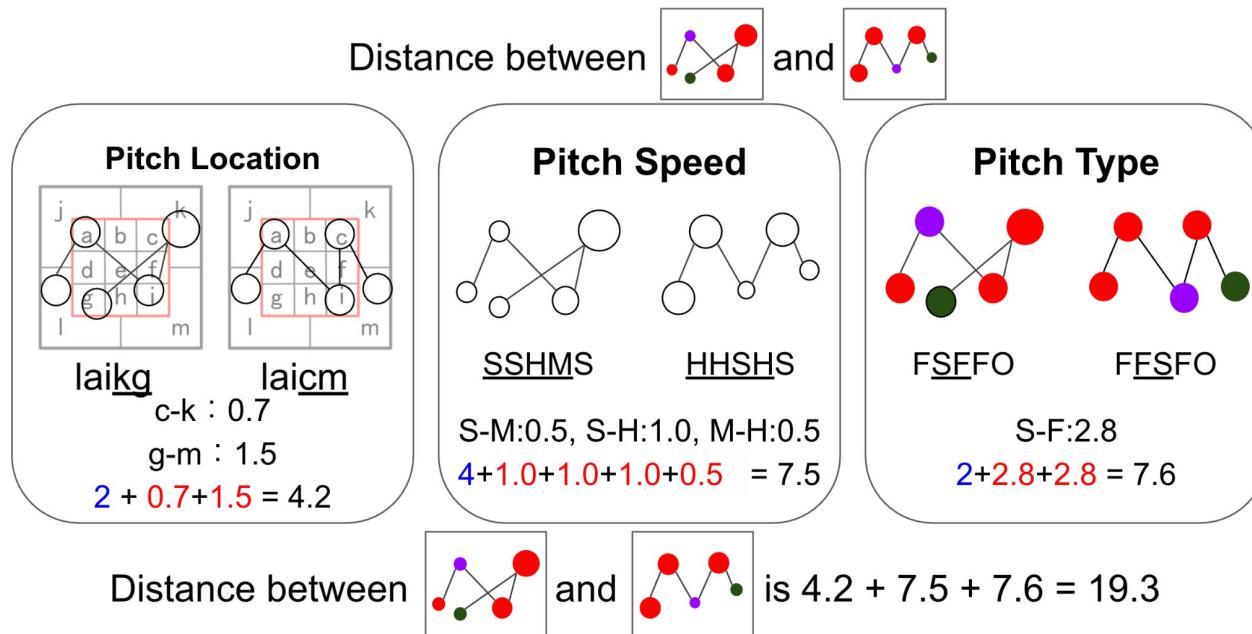
Distance Calculation - Levenshtein Distance -

- The distance between pitch trajectories is calculated using the Levenshtein Distance
- Pitch zone transitions consider physical distance
 - Need to modify the Levenshtein distance



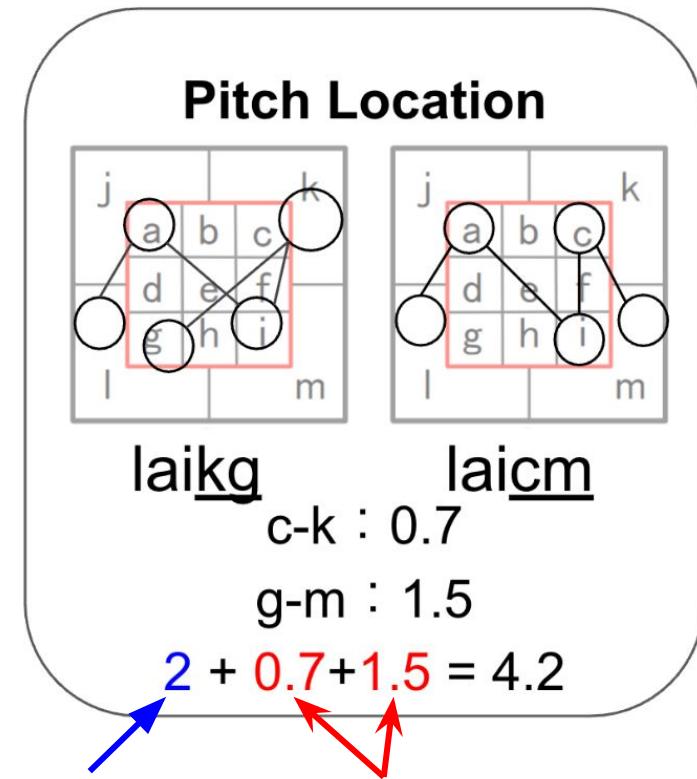
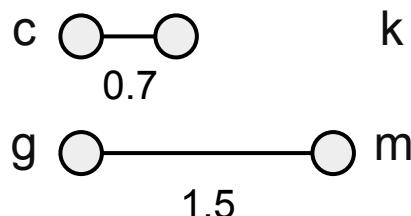
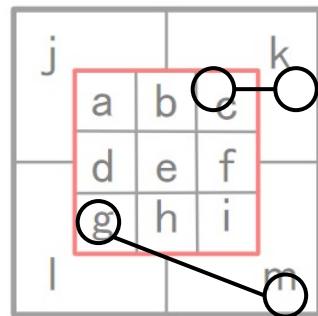
Distance Calculation - Modified Levenshtein Distance - 13

Weights are calculated based on Pitch Location, Pitch Speed, and Pitch Type, and added to **the base distance** between trajectories



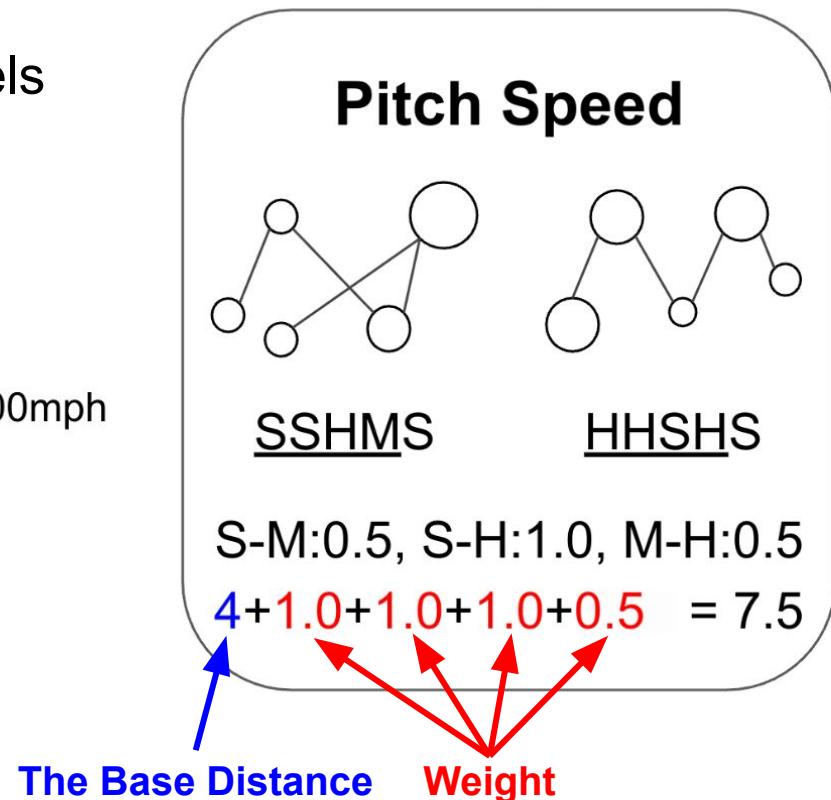
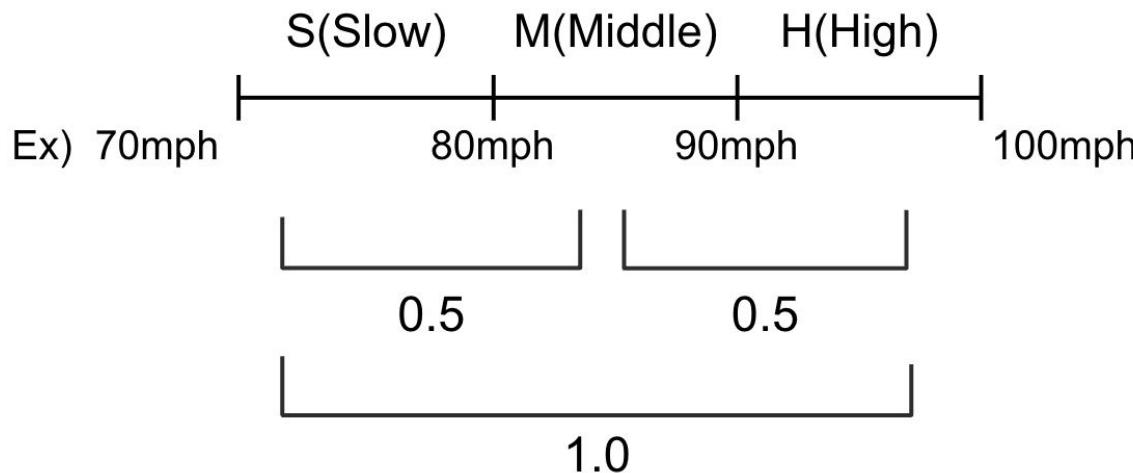
Distance Calculation - Pitch Location -

- Assign coordinates to each zone, and compute weights using Euclidean distance between zones



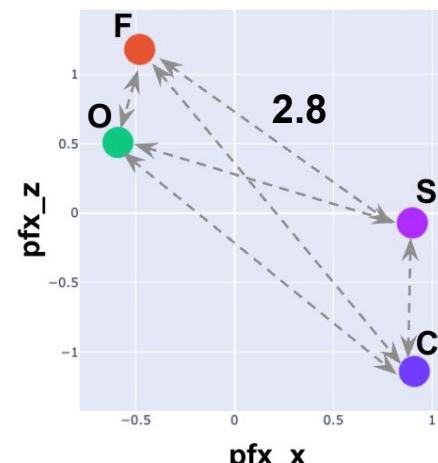
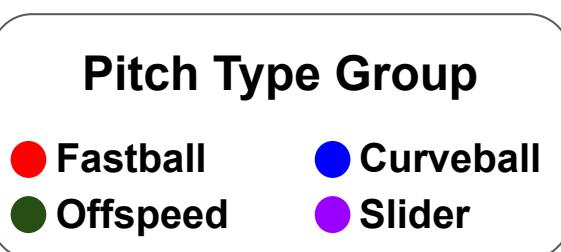
Distance Calculation - Pitch Speed -

- Pitch speed is classified into three levels
- A weight is assigned

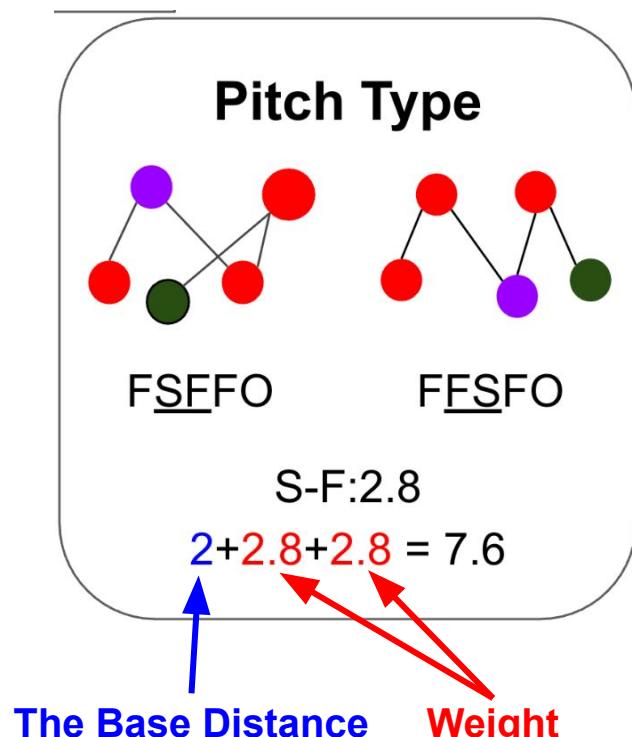


Distance Calculation - Pitch Type -

- We grouped pitch types into 4 categories
- A weight is assigned based on differences in pitch movement data

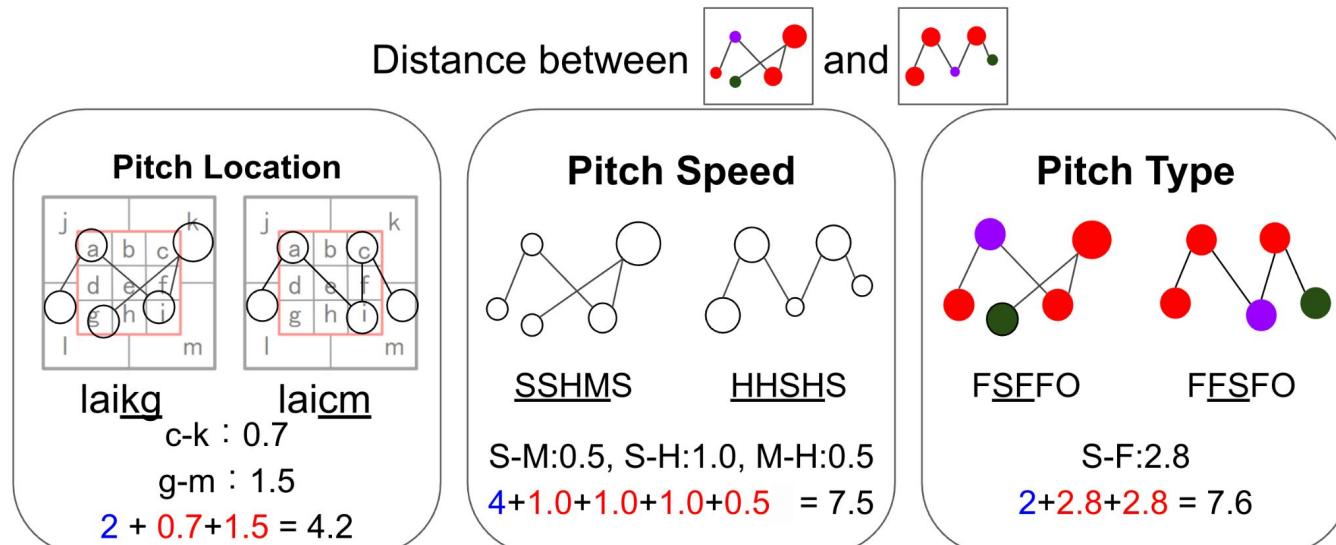


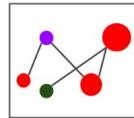
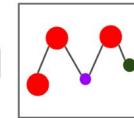
pxf_x ·· Horizontal movement (inches)
pxf_z ·· Vertical movement (inches)



Distance Calculation - Modified Levenshtein distance - 17

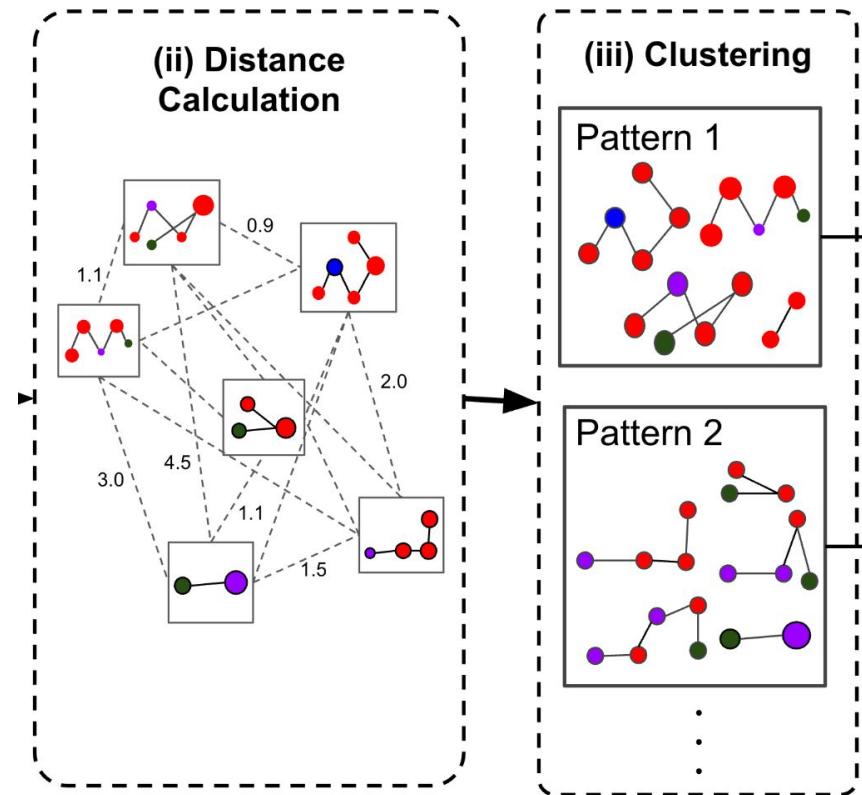
Calculate each individual cost, then sum them up to get the total distance.



Distance between  and  is $4.2 + 7.5 + 7.6 = 19.3$

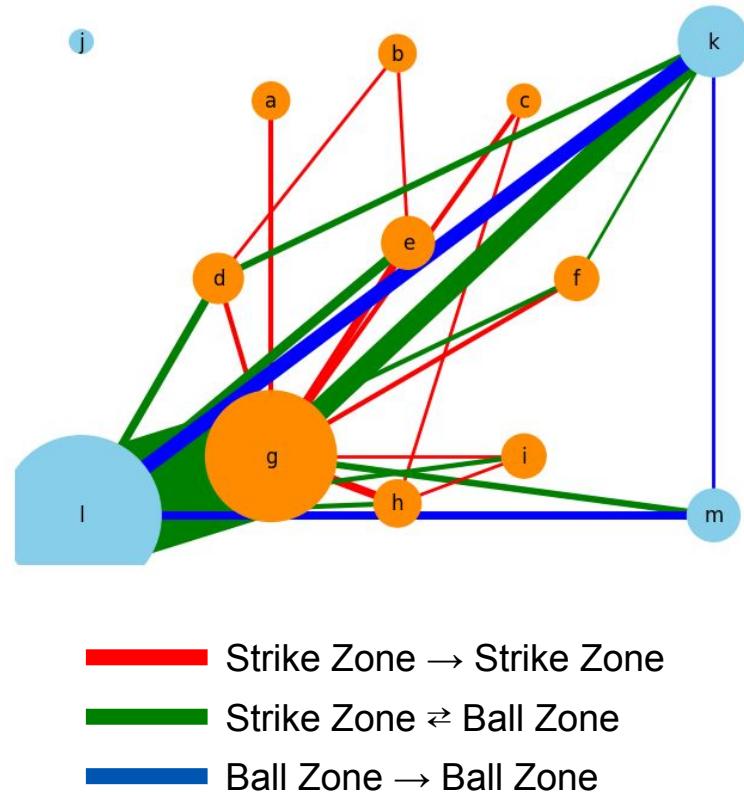
Trajectory Clustering

- K-medoids is a clustering method
 - Requires a distance matrix between data points
- Applied to pitch trajectory distances
- Used to extract trajectory patterns



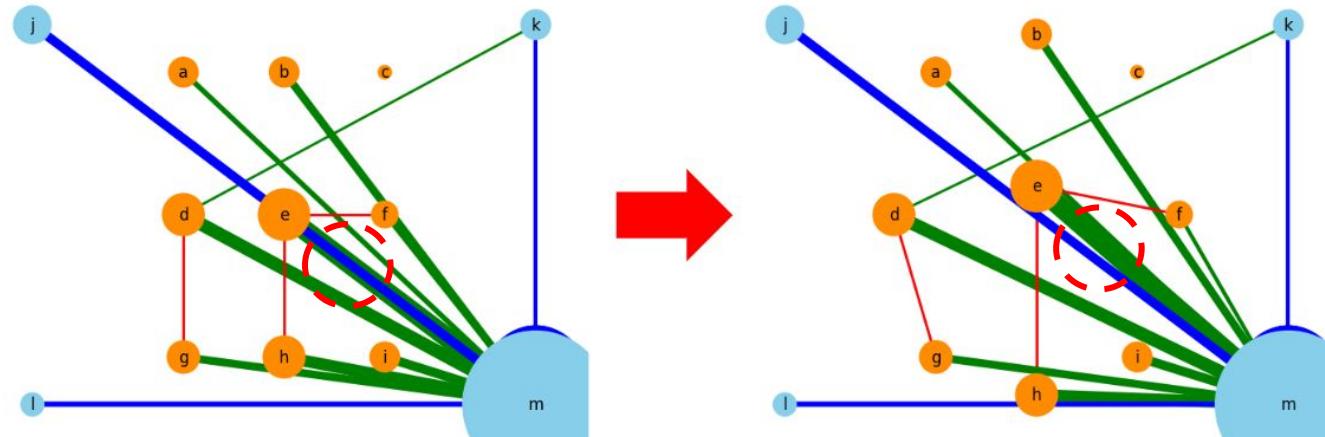
Visualization of Pitching Sequence Trajectory Data 19

- Visualization with Network Diagram
- Nodes = Pitch zones
 - Node size = Pitch frequency
- Edges = Transitions between zones
 - Line width = Transition frequency



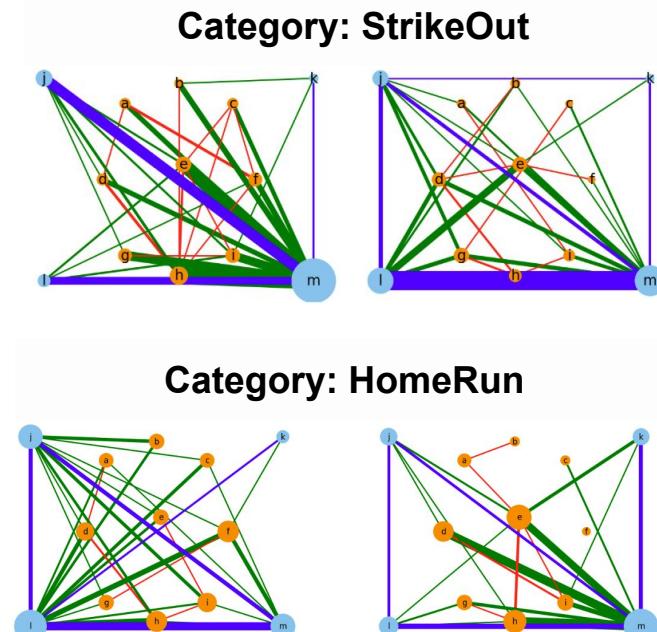
Network Diagram - Layout -

- Visualized using the conventional zone layout
 - Overlapping edges reduced visibility in the network diagram
 - Nodes within the strike zone (*b, d, e, f, h*) were slightly shifted to improve clarity (moved vertically or horizontally)



Network Diagram - Category -

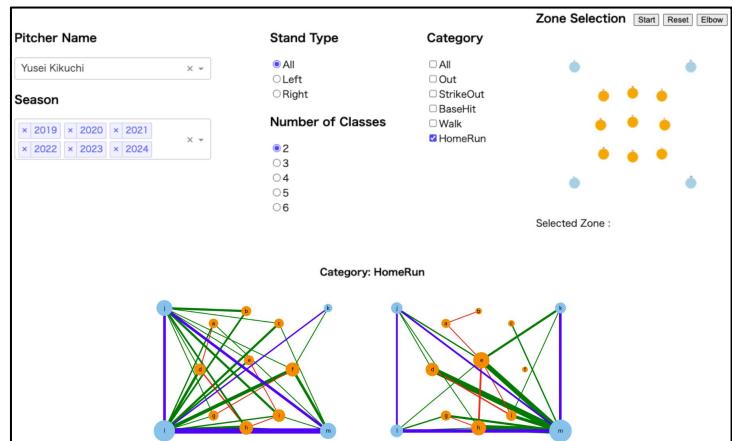
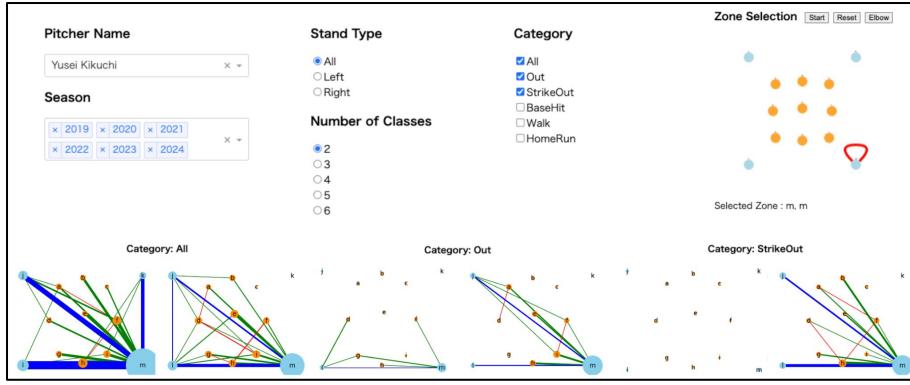
- A user are interested in outcome-based pitching trends
- We grouped at-bat outcomes into five types
 - StrikeOut
 - Out
 - BaseHit
 - Walk
 - HomeRun



Case Studies - Overview -

22

- Pitcher
 - Yusei Kikuchi *
- Season
 - 2019 - 2024

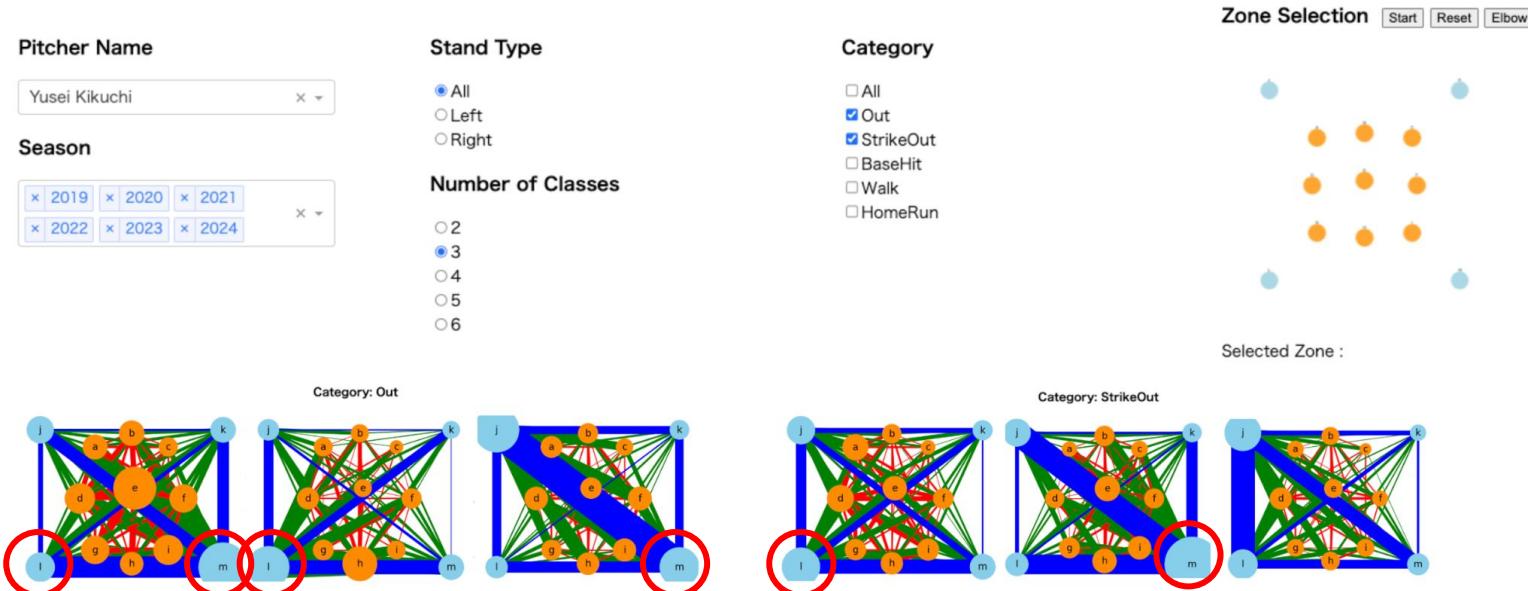


*Los Angeles Angels

Case Studies - Zone: Without No Selection -

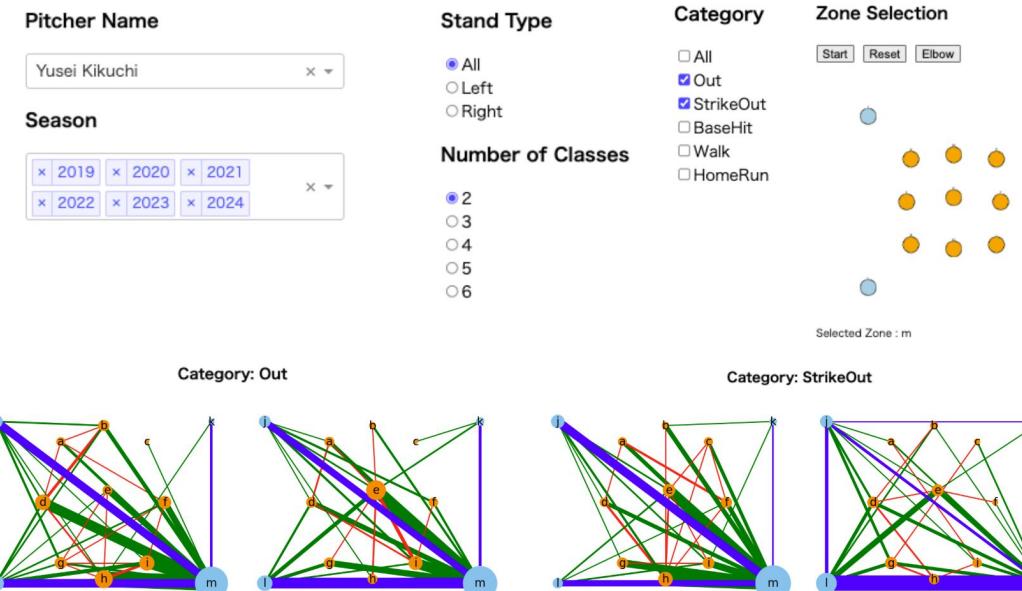
23

Users infer that zones m and l are frequently targeted (large nodes), and are motivated to explore sequences starting from these zones.

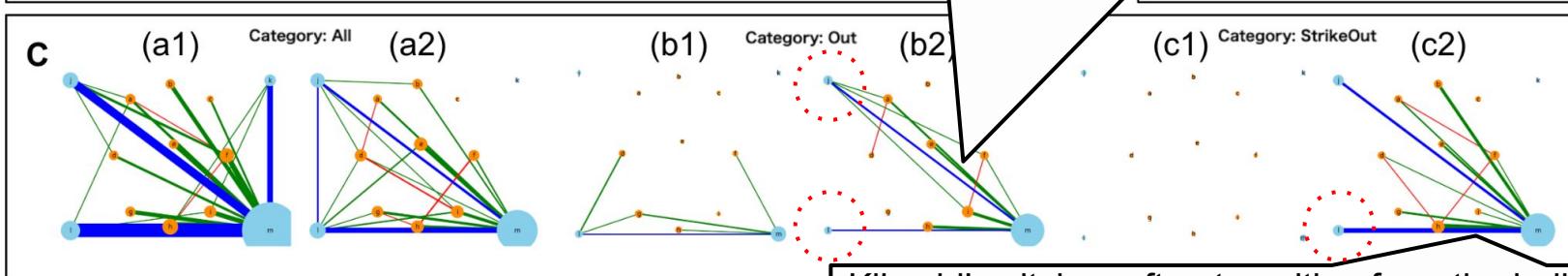
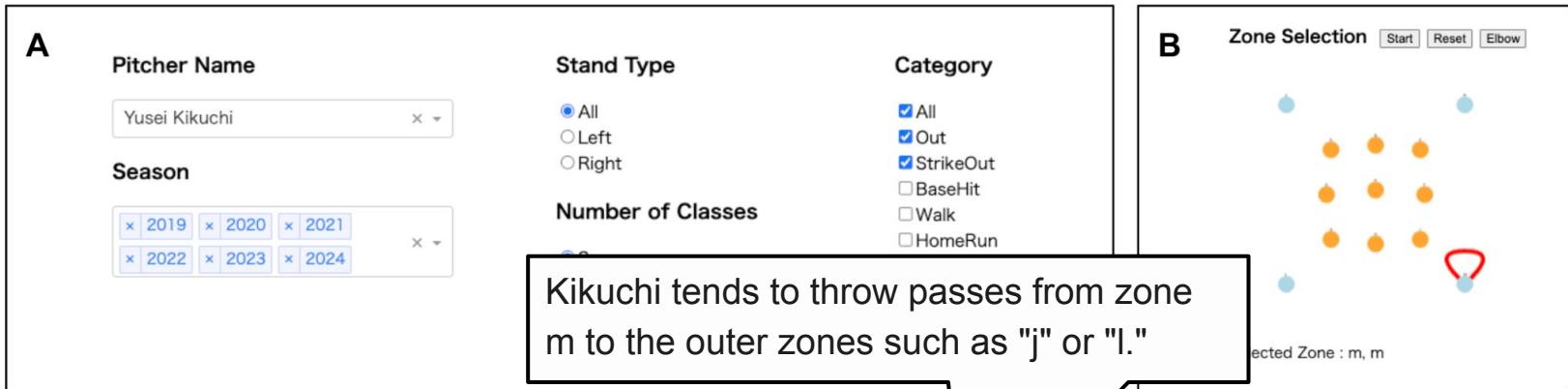


Case Studies - Zone:m -

- Selecting zone 'm' reveals similar visualizations, suggesting common patterns in pitch sequences

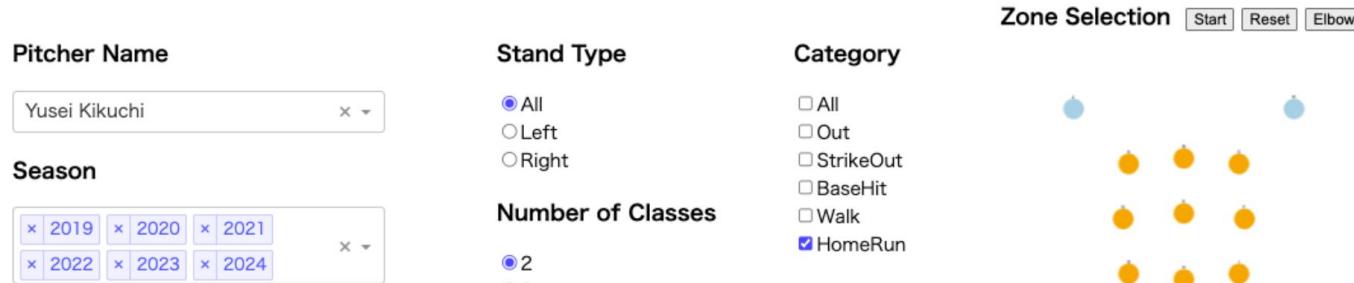


Case Studies - Zone:m-m -

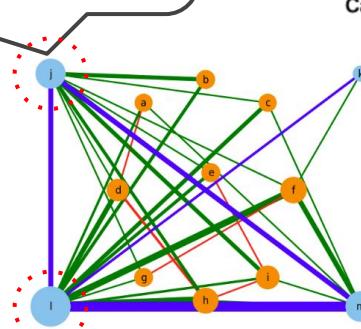


Kikuchi's pitches often transition from the ball zone I to m.
He also tends to spread his pitches from zone m into various strike zones.

Case Studies -Zone:Without Zone Selection, Category:HomeRun-

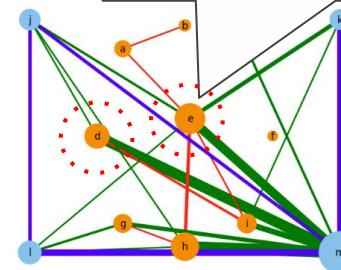


Many sequences start from zones j and l (out-of-zone)
→ Transition into the strike zone



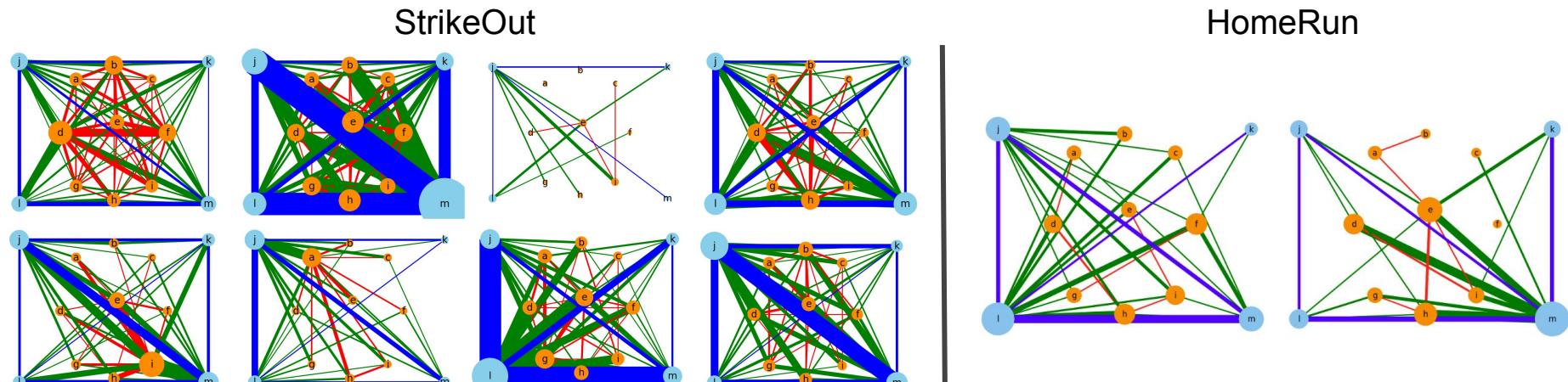
Category: HomeRun

Selected Zone :
Pitches are concentrated in central zones d and e



Case Studies - Compare StrikeOut with HomeRun - 27

- Compare StrikeOut with HomeRun
- Might say that, compared to home runs, strikeouts make more effective use of the ball zone



How Each Requirement is Fulfilled

28

As this case study shows, our system meets the initial requirements

R4 Real-time interaction and updates

R5
Advanced
filtering options

A

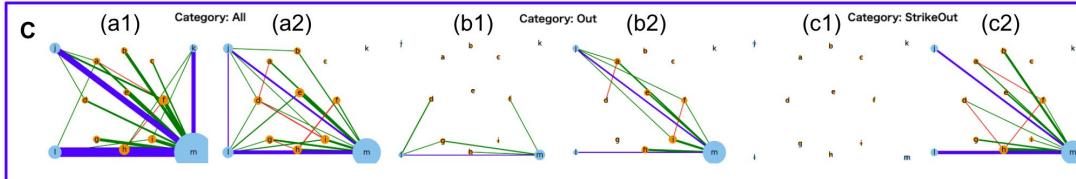
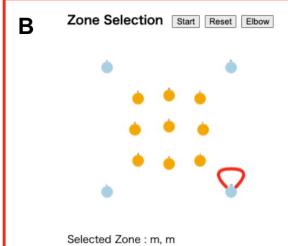
Pitcher Name: Yusei Kikuchi

Season: 2019, 2020, 2021, 2022, 2023, 2024

Stand Type: All, Left, Right

Number of Classes: 2, 3, 4, 5, 6

Category: All, Out, StrikeOut, BaseHit, Walk, HomeRun



R1
Zone-based
pitch filtering

R2
Pattern discovery
from sequences

R3
Outcome-based
comparison

- We built a system called "P3VS" that allows real-time analysis of pitching strategies, even during a game.
 - Makes it possible to analyze pitching strategies of individual pitchers

Future Works

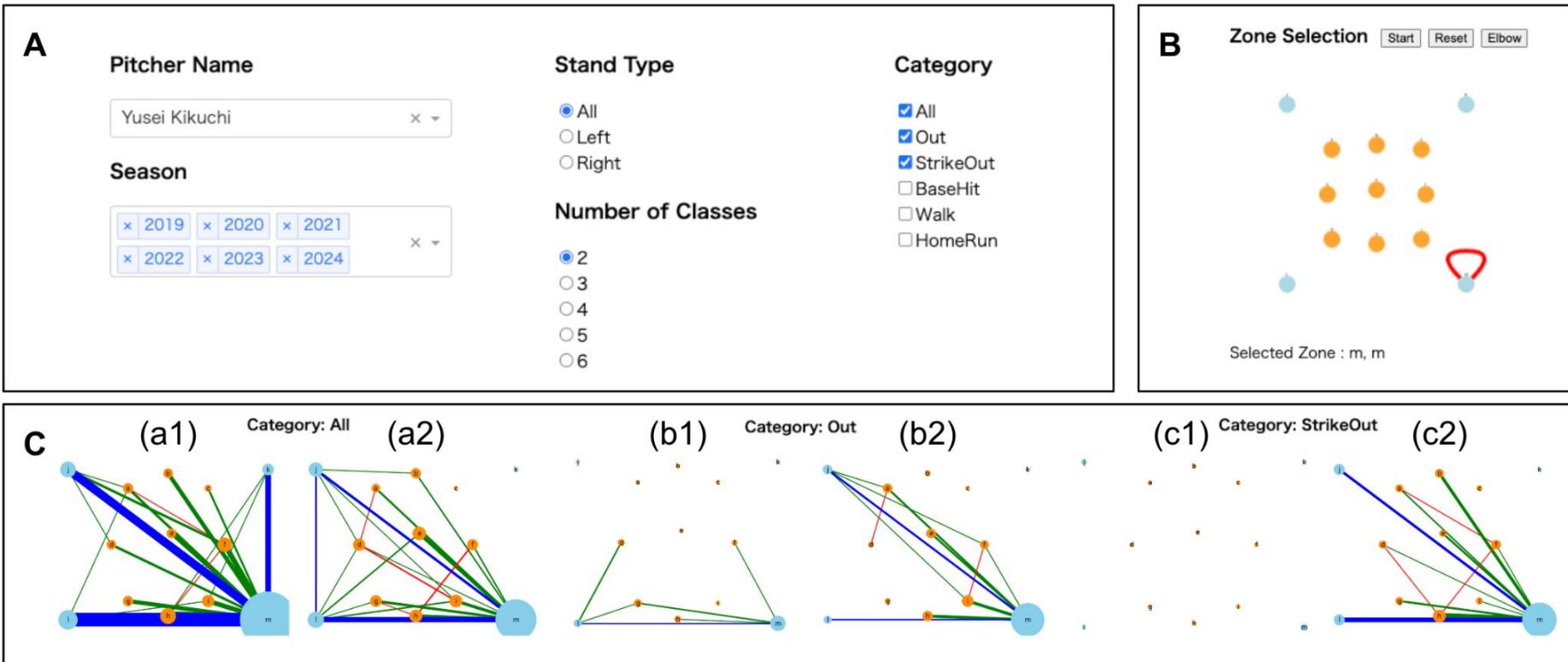
- Conduct user evaluations to assess usefulness
- Incorporate pitch type and speed into visualizations
- Add directionality to edges in the network

ACKNOWLEDGMENT

30

- Supported by:
 - Hokkaido IT Creator Discovery and Development Project
(Shinsetsu Program)

Thank You for Your Attention

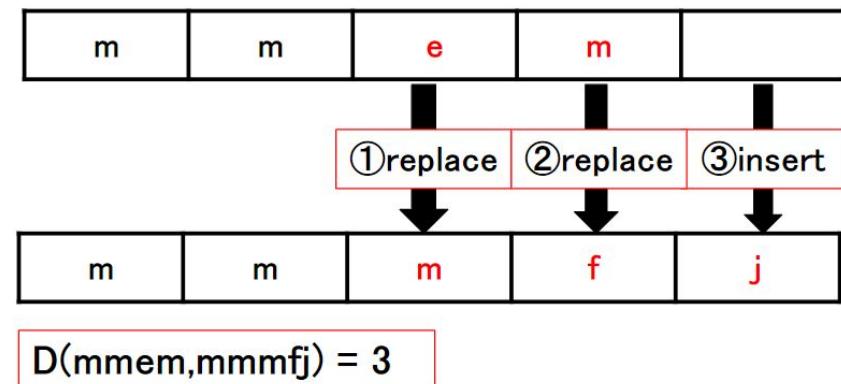


Ryosuke Tsujino, Hokkaido Information University, Address : s2481107@s.do-johodai.ac.jp



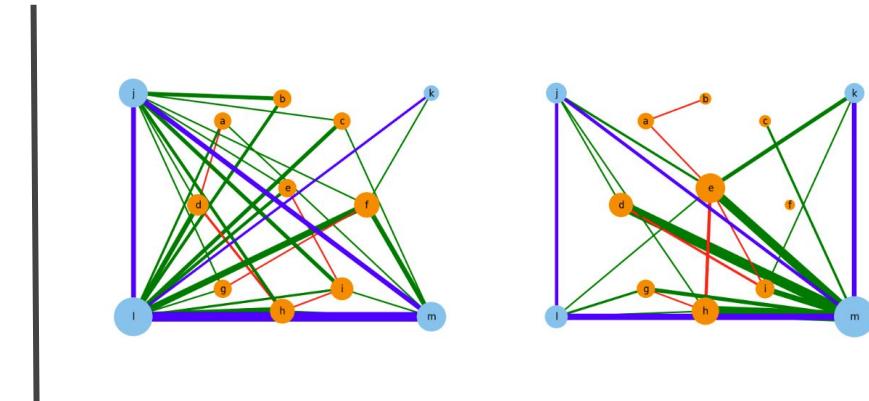
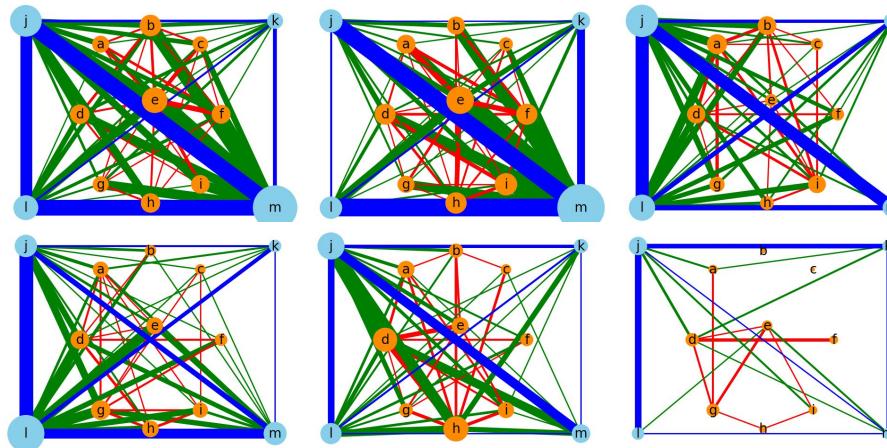
Distance Calculation - Levenshtein distance -

- The distance between pitch trajectories is calculated using the Levenshtein distance.
- Transform S_1 into S_2 using insertion, deletion, and substitution operations
- Count the operations required to convert S_1 to S_2



Case Studies - Compare StrikeOut with HomeRun - 34

- Compare StrikeOut(Left) with HomeRun(Right)
-



Case Studies - Compare StrikeOut with HomeRun - 35

- Compare StrikeOut(Left) with HomeRun(Right)
-

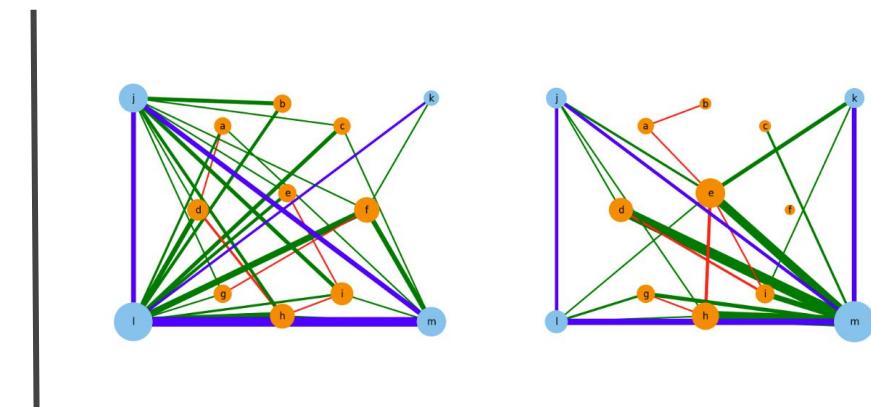
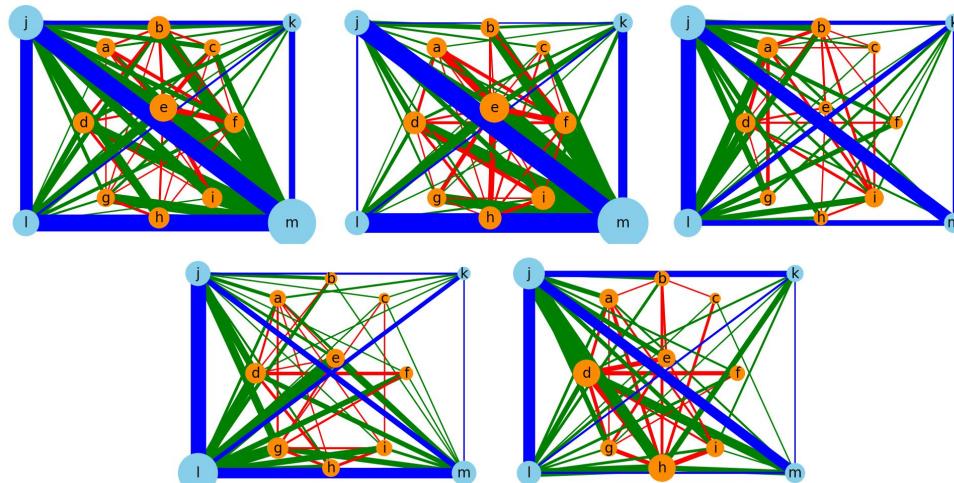


TABLE III
AT-BAT RESULT CATEGORIES BASED ON THE EVENTS ATTRIBUTE.

Category	Included results
StrikeOut	strikeout
Out	field_out, double_play, force_out, etc.
BaseHit	single, double, triple
Walk	walk, hit_by_pitch
HomeRun	home_run

Case Studies - Compare StrikeOut with HomeRun - 37

- Compare StrikeOut(Left) with HomeRun(Right)
-

