Visualization System for Pitching Sequence Patterns based on Pitching Locations' Order

Ryosuke Tsujino*
Hokkaido Information University

Masahiko Itoh†
Hokkaido Information University

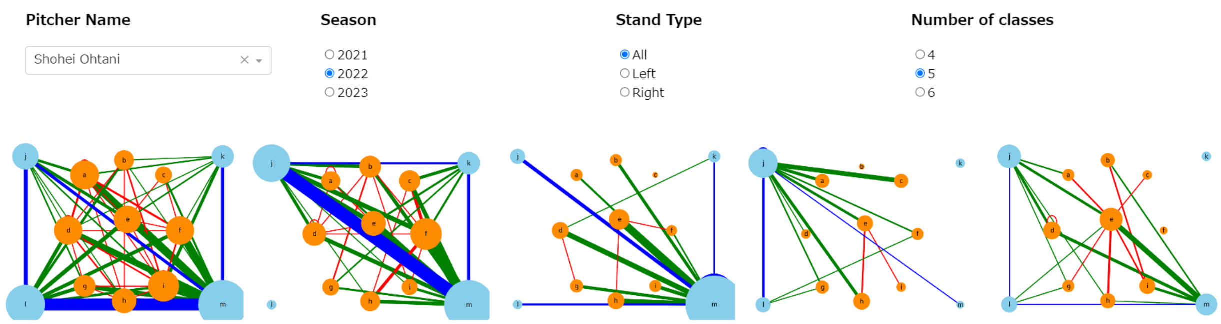

Figure 1: MLB Pitcher's Pitching Sequence Pattern Visualization System: user can select pitcher name, season, stand type (all, left-handed batter, or right-handed batter), and number of clusters in visualization system. After user selects them, system shows network diagram of pitching sequence patterns.

Abstract

Baseball has various tactics, and the combination of pitches is one of the most important tactics. This combination of pitches directly affects the development of the baseball game, wins and losses, and individual players' performance. In other words, there is a need for research that focuses on pitchers' combinations of pitches. However, there is not much research on pitchers' combinations of pitches considering pitching the order of locations.

In this paper, we define the combination of pitches as the order of pitchers' pitch locations to one batter, and we analyze pitchers by considering them. We propose a system for visualizeing the combination of pitches for each pitcher by clustering the orders of pitchers' pitch locations. For this purpose, we treat the order of pitchers' pitch locations extracted from pitching data as a trajectory, calculate the distance between trajectories, perform trajectory clustering, and visualize resulting patterns. Using the proposed visualization system, we can analyze each pitcher's pitching sequence patterns and find differences in pitching tendencies among pitchers.

1 Introduction

In baseball, there are various important factors in the tactics used, such as running, fielding positions, and batting. In particular, the combination of pitches has a significant impact on game results and player performance. The combination of pitches means how a pitcher constructs his pitches to the batter. It is the pitching strategy or pitching pattern used by a pitcher when pitching. The choice of tactics is important because what pitches a pitcher throws and in what order directly affects whether a batter will hit the ball, the development of the game, and the outcome of the game.

Research that focuses on pitchers' pitching is needed. Tsujino et al. [2] focused their research on pitcher performance. They defined the combination of pitches as pitch distribution that is defined by the catcher mitt locations and pitch types. Then, they visualized the percentage of the catcher mitt locations for each pitch type using a heat map. However, the pitch order was not taken into account. In this paper, we define the combination of pitches as the order of pitchers' pitch locations to one batter.

We propose a system for visualizing pitching sequence patterns that are clustering results of the orders of pitch location. The system visualizes clusters based on a selected player, season, stand type, and the number of clusters. We utilize the data from Baseball Savant to analyze MLB pitchers. Visualized results show the pitching patterns of each pitcher and differences in pitching tendencies among pitchers.

2 Proposed System

In this paper, we consider the pitching order as a trajectory, and we use clustering to extract trajectory patterns. We perform clustering by finding the distance between the trajectories extracted from the pitching data. Then, we visualize and analyze clustering results. Moreover, we construct an interactive system to visualize pitchers' pitching tendencies.

2.1 Pitching Data

We use data from Baseball Savant[1]. We analyze a total of 21 pitchers, including the top 20 pitchers with the most pitches in 2022 and Shohei Ohtani. We use 10 attributes such as game date, zone location, plate appearance number of the game, total pitch number of the plate

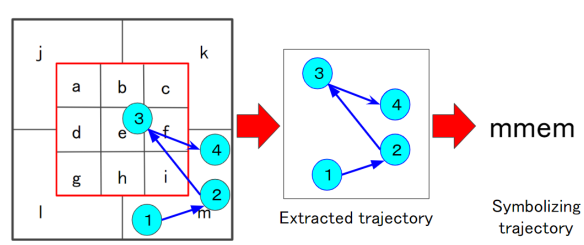

Figure 2: Extraction of pitching trajectory data from pitching data

[1] https://baseballsavant.mlb.com/

*e-mail: s2021167@s.do-johodai.ac.jp
†e-mail: imash@do-johodai.ac.jp

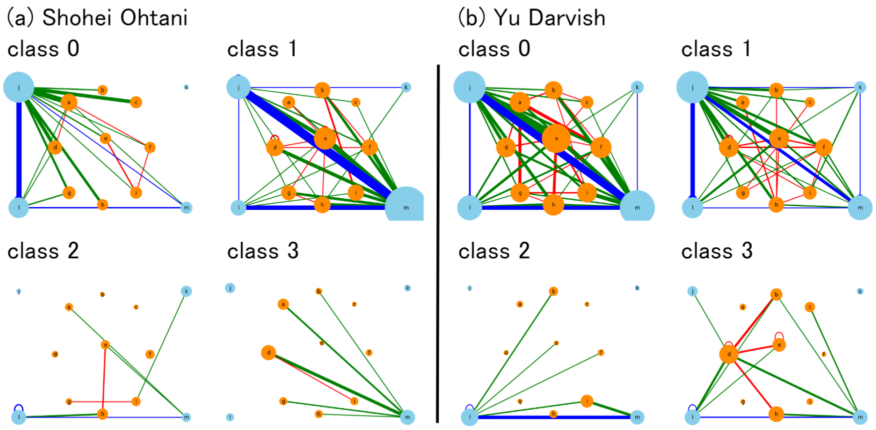

Figure 3: Visualization results for left-handed batters: (a) Shohei Ohtani, and (b) Yu Darvish.

appearance,and left-handed/right-handed batter. We use the zone table obtained from the data site (Figure 2, Left).

2.2 Pitching Trajectory Data Extraction

To extract pitching trajectory data, we used the method proposed by Miyagi et al. [1] that extracts patterns from trajectories. We extract the trajectory of one batter's pitching data, and convert it to a string as shown in Figure 2. We call this pitching trajectory data. For example, if the first and second pitches were thrown in zone m, the third pitch was thrown in zone e, and the fourth pitch was thrown in zone m, the pitch trajectory data is "mmem" (Figure 2).

2.3 Trajectory Clustering

The distance between the extracted pitching trajectories is determined using the Levenshtein Distance. This method calculates the number of single character insertions, deletions, and substitutions required to convert a specified string S_1 to string S_2. The smaller the distance value, the more similar character strings can be determined.

This time, we use pitching trajectory data with 4 or 5 elements (characters) that make up the trajectory. We calculate the distance between pitching trajectories using Levenshtein Distance and create a distance matrix.

Clustering uses k-medoids, which is performed on the basis of a distance matrix. K-medoids is a form of non-hierarchical clustering that can be performed when the distances between the elements are given.

2.4 Visualization System

There are four items that can be selected in the visualization system: pitcher name, season, stand type (all, left-handed batter or right-handed batter), and number of clusters. When we select each item, a network diagram of the pitching analysis results is displayed as shown in Figure 1.

Nodes represent pitching zones in the zone table. The edges represent the path of the trajectory. We place 13 nodes (from a to m) on the basis of the zone table in Figure 2. To distinguish between the strike zone and the ball zone, the nodes in the strike zone (from a to i) are set to orange, and the nodes in the ball zone (from j to m) are set to light blue. We move the position of strike zone nodes to avoid overlapping edges.

The size of a node represents the frequency of pitches to that zone. The width of an edge represents the frequency of passing through that path. Edge colors represent strike zone to strike zone, strike zone to ball zone, and ball zone to ball zone as red, green, and blue respectively.

3 Case Studies

We analyzed the pitching trends of Shohei Ohtani and Yu Darvish. The number of clusters was set to 4. Here, we introduce the results of comparing the pitching tendencies of both pitchers using the patterns with left-handed batters as an example. Note that the visualized network diagram is from the catcher's perspective.

In Ohtani's class 0 (Figure 3 (a)), it can be seen that he attacks the area far from the left-handed batter (hereinafter referred to as the outside corner). In particular, the green edges are concentrated in zone j, which shows that there is a strong relationship between zone j and the strike zone. In class 1, his pitching uses zone m a lot. Additionally, it can be seen that the blue edges that represent the relationships between the ball zones are thicker. This means that there is a strong relationship between zone m and the ball zone.

Darvish's class 0 shows a slightly different trend from Ohtani (Figure 3 (b)). Ohtani's results have less red edges overall, but Darvish's results have more red edges and are thicker. It shows that the relationship between the strike zones is strong. Also, from the relationship between zone a and zone g, it can be seen that Darvish uses high and low pitches to distribute pitches to left-handed batters. Furthermore, the sizes of the nodes in zone j and zone m are similar, and the relationship between them is strong. This indicates that Darvish's pitching takes advantage of corners. Darvish's class 1 pitches, like Ohtani's class 0 pitches, tended to attack the outside corners of left-handed batters.

4 Conclusion

In this paper, we proposed a system that visualizes and analyzes MLB pitchers' pitching sequence patterns taking into account the order of pitching locations. We believe that these results can be used to create templates for amateur players in the future.

We plan to consider how to extract and visualize patterns using additional attributes such as pitch type, speed, and outcome in future work.

References

[1] Y. Miyagi, M. Onishi, C. Watanabe, T. Itoh, and M. Takatsuka. Classification and visualization for symbolic people flow data. *J. Vis. Lang. Comput.*, 43:91–102, 2017.

[2] R. Tsujino, H. Toyota, and M. Itoh. Visual analysis of pitch location trends by performance category for pitchers and batters. In *Proc.85th National Convention of IPSJ*, 2023.