Automated Play Segmentation of Continuous GPS Data from Collegiate Football Players

Wu-Tsai Undergraduate Fellowship

Eli Swagel

Wu Tsai Human Performance Alliance
UNIVERSITY OF OREGON

Understanding of game demands at the play level can inform training methodologies and load monitoring

Plays must currently be tagged by hand, which is inefficient and time costly.

Develop an automated method of identifying start and stop times for plays during a football game

An automated system of play labeling would be efficient and allow for quicker analysis of game demands and player loads

Methodology

Use Catapult data to identify patterns in player movement indicating when a play occurs

Data:
Collected by sensors worn by players during 2021-2022 football season
Football team manually labeled start and stop times for plays

Variables:
Acceleration, velocity, and metabolic power

stumpy

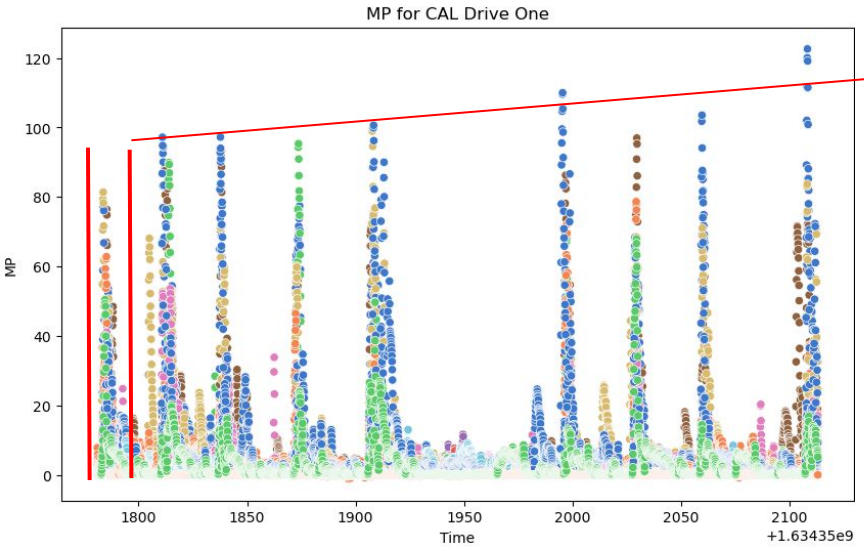
T. S. Learn
Transport Spedycja e-learning

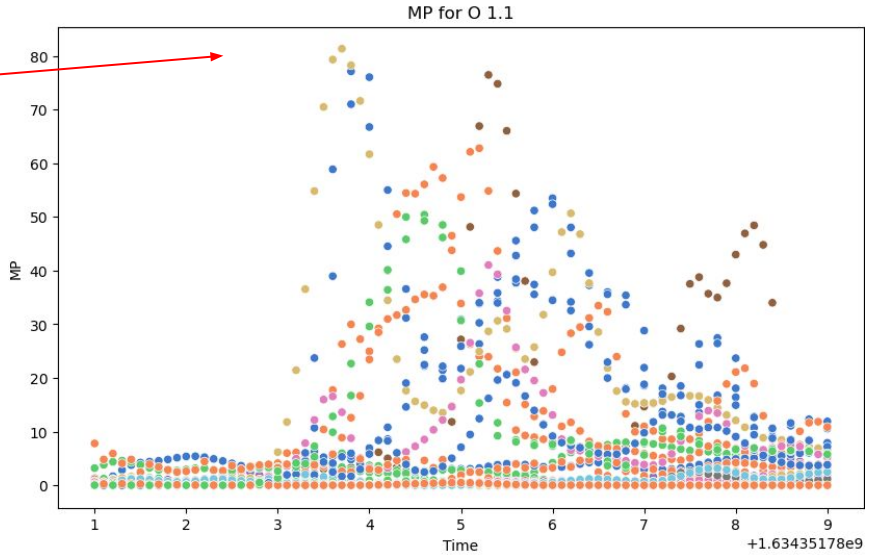
CATAPULT

Metabolic Power

Estimate of the energy demands of acceleration and deceleration events

Metabolic power is collected and calculated by Catapult





Template Matching Using Metabolic Power

Tslearn:

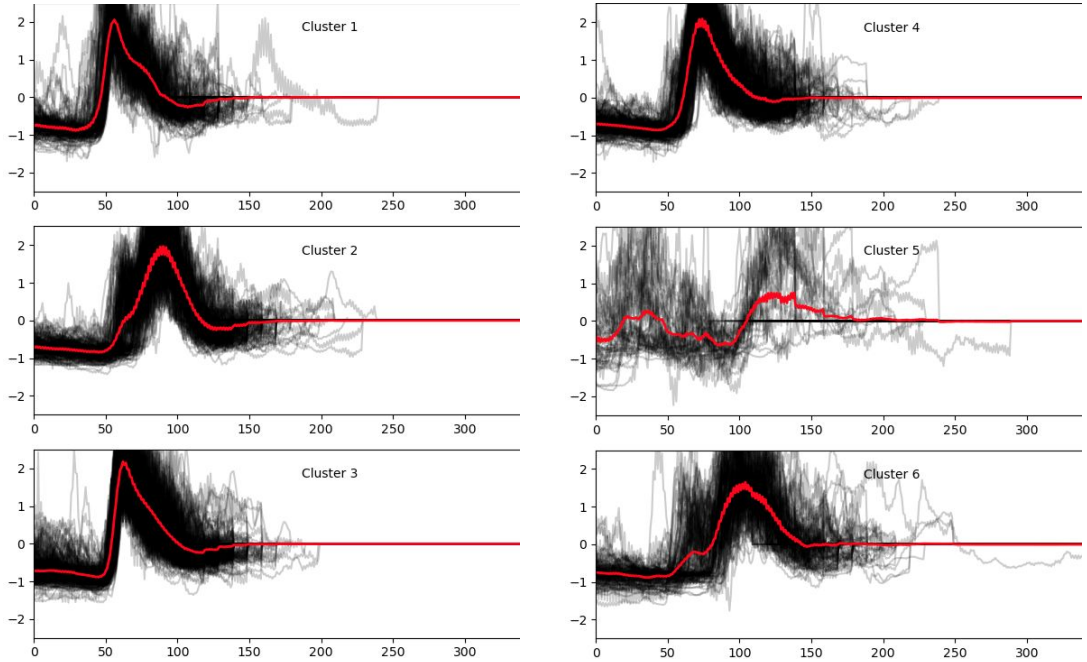
Python package providing ML tools for time-series analysis

Cluster plays from the 2021-2022 season into 6 templates

Stumpy:

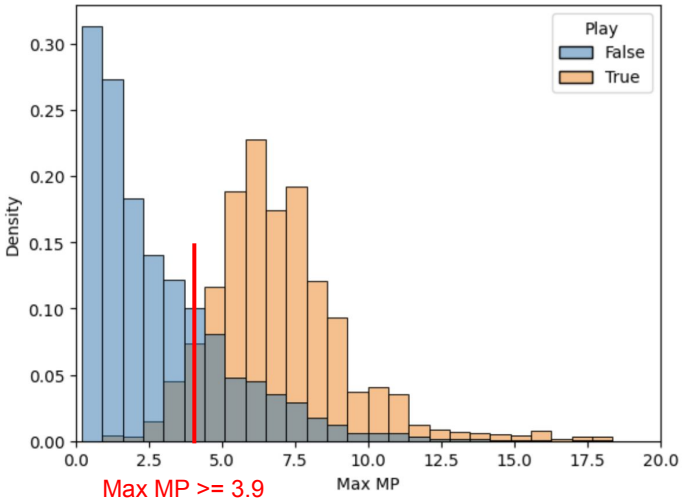
Template matching algorithm that Scans time series to find matching patterns to a subsequence

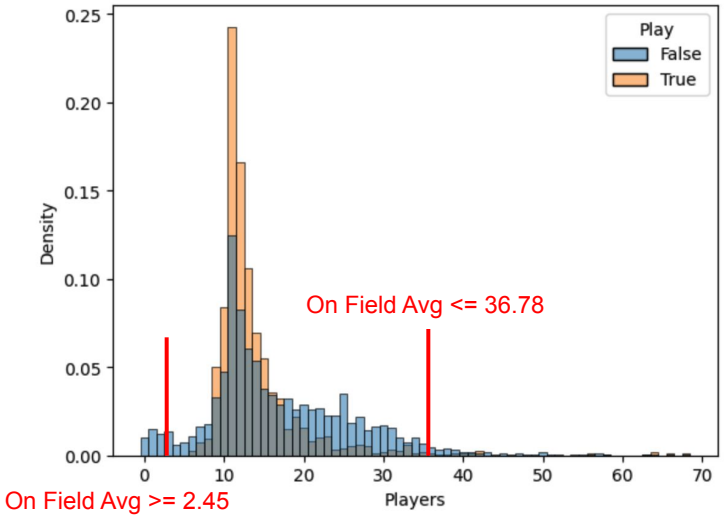
Find patterns of metabolic power that match templates for the metabolic power of plays



Thresholding MP and Players on Field

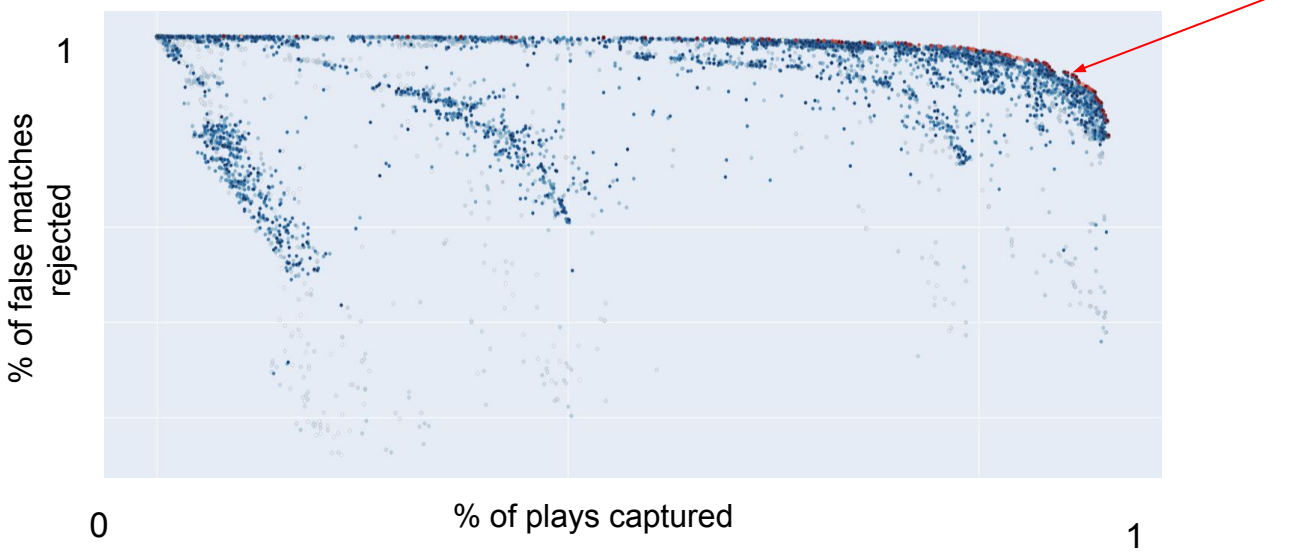
We found pattern matches to each template for all home games during the 2021-2022 season
Stumpy gives normalized distance measure for match and start index for when it occurs
Set thresholds for players on the field and maximum metabolic power to filter out noise





Results

Optimization algorithm runs through combinations of parameter values, searching for the optimal combination

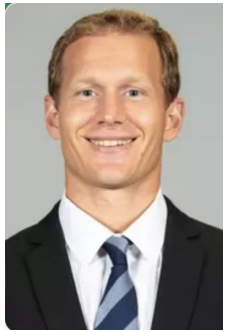


Selected Working Point:
89% of true plays captured
98% of false matches rejected

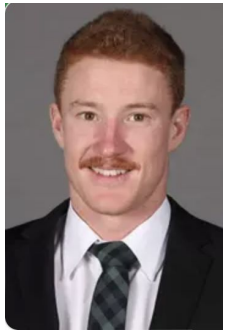
Our model accurately identifies the start and end times for 86% of plays, and there are some errors due to inconsistencies in labeling

Acknowledgements:

Jake Searcy: Assistant Research Professor of Data Science

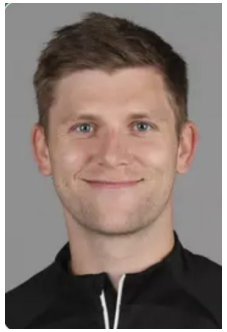


Jordan Troester: Director of Performance and Sports Science



Ben Mckay: Sports Science Coordinator

Anish Dulal: Graduate Research Assistant



Ethan Martin: Sports Science Coordinator