Machine Learning for Time Series

Presentation - Mini-Project

Adonis JAMAL Fotios KAPOTOS

SoccerCPD: Formation and Role Change-Point Detection in Soccer Matches Using Spatiotemporal Tracking Data

Ecole Normale Supérieure Paris-Saclay - MVA

December 17th 2025

Context and Problem Statement

- **Context:** Analyzing team formation is crucial for interpreting tactics in fluid sports like soccer.
- **The Challenge:** Tracking data is noisy; players frequently switch positions temporarily.
- **Limitation of existing methods:** Often assume constant formations or react too sensitively to frame-by-frame noise.
- **Objective:** Reproduce and extend **SoccerCPD** [Kim+22], an unsupervised framework that distinguishes:
  - **FormCPD:** Shifts in global spatial configuration.
  - **RoleCPD:** Long-term tactical changes in individual roles.

Work Repartition and Contributions

Work Repartition

- **FormCPD:** Implementation of Delaunay-based adjacency and formation change detection.
- **RoleCPD:** Synthetic role data generation, permutation analysis, and recursive segmentation.

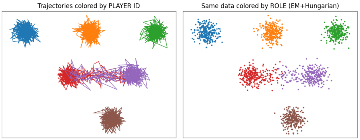

Figure: Role vs Player coloring

Implementation & Extensions

- Python package with R-backend for g-segmentation.
- **Novel Experiment:** "Player Stationarity" metric.

Methodology: FormCPD (1/2)

Role Assignment and Spatial Structure

1. Role Assignment

- Players assigned latent spatial roles via constrained EM.

2. Delaunay Triangulation

- Encodes local spatial relationships.
- Produces binary role-adjacency matrix $A(t)$ invariant to permutations.
- Emphasizes *topological structure*.

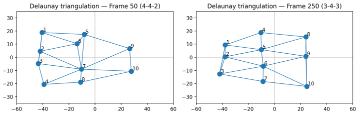

Figure: Delaunay Triangulation reconfiguration

Methodology: FormCPD (2/2)

Change-Point Detection and Adjacency

3. Discrete g-segmentation

- Applied to adjacency sequence $\{A(t)\}$.
- Distance: $d_M(A(t), A(t')) = \|A(t) - A(t')\|_{1,1}$.
- Constraints: $p < 0.01$, min duration 5 min.

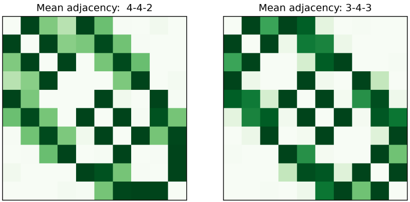

Figure: Mean Adjacency Matrices

Methodology: RoleCPD

Detecting Intraphase Tactical Shifts

- **Goal:** Detect permanent role swaps within stable formation.
- **Metric:** Normalized Hamming Distance (Switch Rate):

$$d(\pi_t, \pi_{t'}) = \frac{1}{N} \sum_{p \in P} \mathbb{1}_{\pi_t(X_p) \neq \pi_{t'}(X_p)}$$

- **Process:**
  1. Preprocessing: Exclude Switch Rate > 0.7.
  2. Recursive g-segmentation.

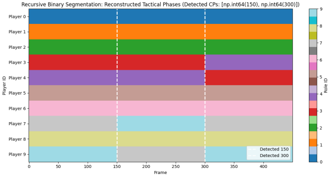

Figure: Recovered Tactical Phases

Data Sources and Diagnosis

High-Frequency Tracking (GPS)

- SkillCorner (10Hz) & Last Row (20Hz).
- Clean signals, suitable for Gaussian assumptions.

Event-Stream (StatsBomb)

- **Challenge:** Extreme sparsity (< 10 pts/min).
- **Adaptation:** Pseudo-trajectories via temporal aggregation (5-min average).

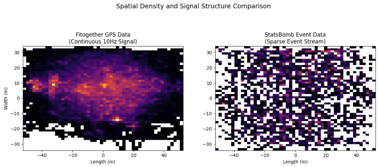

Figure: GPS vs Event Data Density

Results: Sensitivity Analysis

Sensitivity to `min_fdist`

- `min_fdist`: Threshold for formation distance.
- **Observation:** Increasing threshold reduces segment count but preserves stable change-points.
- **Conclusion:** Acts as a pruning mechanism for minor structural variations.

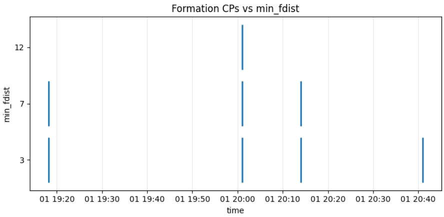

Figure: Timeline stability under varying thresholds

Extension 1: Possession Context

Refining Formation Analysis

- **Limitation:** Global average ignores game state.
- **Method:** Filter frames (Attack vs. Defense).
- **Result (Match 1925299):**
  - **Defending (Left):** Compact block.
  - **Attacking (Right):** Wider, advanced shape.

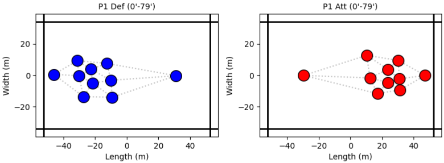

Figure: Defending vs Attacking Formations

Extension 2: Player Stationarity Metric

Quantifying Role Fluidity

- **Metric:** Std dev of Euclidean distance between player position and role center.
- **Findings (A-League):**
  - Captures variance in movement within a "stable" phase.
  - Distinguishes "Fixed" roles (CB) from "Roaming" roles (Wingers/Midfield).

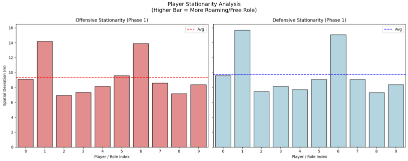

Figure: Player Stationarity (Phase 1)

Conclusion

- **Successful Reproduction:** Implemented FormCPD and RoleCPD, validating the framework on synthetic and real matches.
- **Robustness:** Confirmed method stability via sensitivity analysis.
- **Key Extensions:**
  1. **Possession Context:** Revealed distinct attacking/defensive shapes.
  2. **Stationarity Metric:** Provided quantitative insights into player role adherence.
- **Versatility:** Demonstrated applicability to sparse event-stream data.

Bibliography

[Kim+22] Hyunsung Kim et al. "SoccerCPD: Formation and Role Change-Point Detection in Soccer Matches Using Spatiotemporal Tracking Data". In: *The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, 2022, pp. 3146–3156. DOI: 10.1145/3534678.3539150. URL: https://doi.org/10.1145/3534678.3539150.