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

École Normale Supérieure Paris-Saclay - MVA

December 17th 2025

MATHÉMATIQUES
VISION
APPRENTISSAGE

Motivation

Problem. Team formations in soccer are *fluid but structured*: players temporarily swap roles (pressing, overlaps), while genuine tactical changes occur at specific moments.

Solution. We study **SoccerCPD** [1], a change-point detection framework that segments a match into tactically consistent periods by:

1. detecting formation changes from role-adjacency graphs (FormCPD)
2. detecting role switches via permutation dynamics (RoleCPD).

Work Done

Reproduction & Benchmarking

- **Reproduction**: Reproduction of the core algorithms
- **Sensitivity Analysis** : Experiments using the python package on provided data

Our Contributions (Beyond the Paper).

- **Context-aware formations:** separation of attacking vs. defensive shapes using possession.
- **Role stationarity metric:** quantifying player role stability (fixed vs. roaming).
- **Data generalization:** adaptation of the framework to sparse event-based data (StatsBomb).

FormCPD: Formation Change-Point Detection

1. **Role Assignment:** Map players to latent spatial zones via EM.
2. **Graph Encoding:** Build adjacency matrices using Delaunay triangulation.
3. **Change Detection:** Use recursive g-segmentation to find where graph topology shifts.
4. **Pattern Clustering:** Group segments into canonical types via alignment.

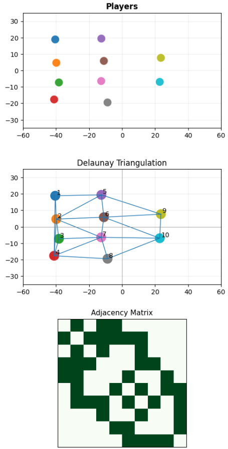

Figure: Pipeline

RoleCPD: Detecting Intraphase Swaps

Goal: Detect permanent swaps within a stable formation.
Metric: Hamming Distance on permutation sequence π_t:

$$d(\pi_t, \pi_{t'}) = \frac{1}{N} \sum_p \mathbb{I}_{\pi_t(X_p) \neq \pi_{t'}(X_p)}$$

Recursive Segmentation Algorithm

- **Filter:** Ignore high-noise frames (Switch Rate > 0.7).
- **Search:** Recursive discrete g-segmentation (graph-based).
- **Stop:** If p-value > 0.01 or segment < 5 min.

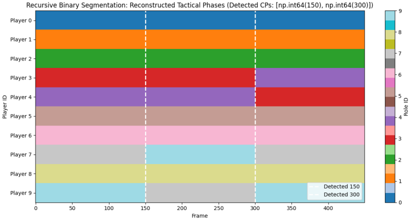

Figure: Visualization of all detected tactical phases after recursive binary segmentation, showing accurate recovery of the ground truth phases.

High-Frequency vs. Sparse Event Data

Tracking Data (GPS)

- **Source:** Fitogether (10Hz), SkillCorner (10Hz) [2].
- **Quality:** Dense, clean signals.
- **Prep:** Normalized to centered metric pitch.

Event Data (StatsBomb)

- **Source:** On-ball events only.
- **Challenge:** Extreme sparsity (< 10 pts/min).
- **Bias:** Spatial bias toward the ball.

Novel Adaptation for Event Data:

- **Pseudo-Trajectories:** Aggregated events over 5-minute rolling windows.
- **Proxy Graph:** Average adjacency matrices instead of frame-by-frame Delaunay.

Density Comparison

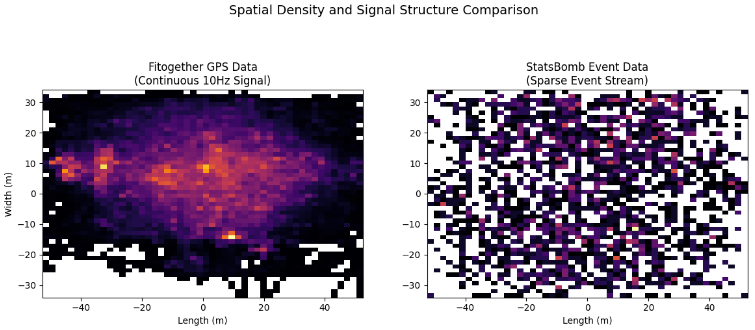

Figure: Spatial density comparison between continuous Fitogether GPS tracking data and StatsBomb event-based data (right). The GPS data shows continuous coverage of player positions, while the event data is sparse and concentrated around ball actions.

Experiments on SoccerCPD : Robustness Analysis

Comparison across CPD backends (single match).

- Four backends tested: `gseg_avg`, `gseg_union`, `kernel_rbf`, `kernel_linear`.
- Identical preprocessing and hyperparameters.
- All methods detect formation changes at similar times.
- Differences correspond to sensitivity (extra/missing CPs), not different segmentations.

Sensitivity to hyperparameters (`gseg_avg`).

- Increasing `min_fdist` reduces the number of segments while preserving CP locations.
- `min_fdist` acts as a pruning threshold for minor structural variations.
- Varying `min_pdur` has no effect, indicating well-separated formation changes.

Contextualizing Formations: Attack vs. Defense

Problem

A single "average" centroid formation ignores game state.

Method:

- Filter frames by possession status.
- Compute formations separately.

Insight:

- Defense: Compact block.
- Attack: New attacking shape.

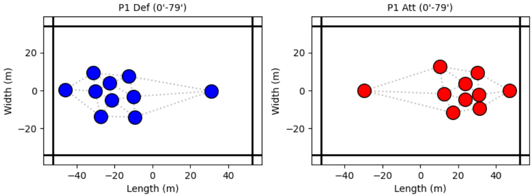

Figure: Detected formations for match "1925299" (A-League, SkillCorner) separated by phase of play: defending (left) vs. attacking (right).

Quantifying Role Fluidity via Stationarity

Motivation

Beyond average position, how consistently does a player occupy their role's spatial zone?

Definition: Standard deviation of Euclidean distance from the role center.

- **Fixed Roles:** CBs show low Stationarity.
- **Roaming Roles:** Wingers/Midfielders show high Stationarity.
- Validates tactical roles beyond simple coordinates.

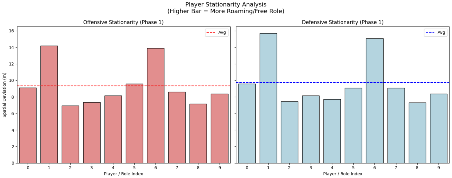

Figure: Player stationarity during Phase 1 of match "1925299" (A-League, SkillCorner) separated by phase of play: attacking (red) vs. defending (blue).

Offensive vs. Defensive Structures (FC Barcelona)

Challenge

Sparse event data lacks continuous player trajectories. Does the SoccerCPD framework still reveal tactical shapes?

- Despite low data density, distinct shapes emerge when aggregated over time.
- **Defense:** High density in some zones (winning back possession).
- **Attack:** More spread out, reflects offensive maneuvers.

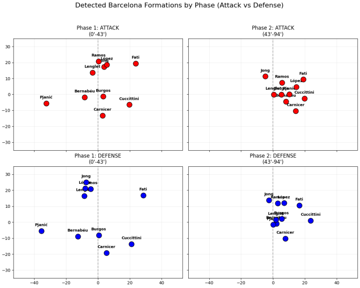

Figure: Detected formations for FC Barcelona using StatsBomb data separated by different tactical phases within the match.

Quantifying Fluidity in Event Data

Challenge

Sparse event data may not capture fine-grained movement.

Consistency: The "Fixed" vs. "Fluid" distinction persists.

Phase Analysis (Barcelona):

- Similar average stationarity across phases.
- But different players show role fluidity shifts between phases.

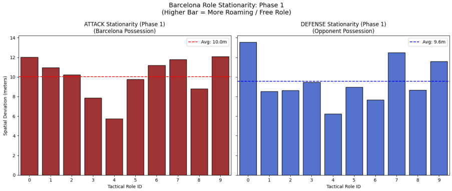

Figure: Player stationarity during Phase 1 of FC Barcelona match (StatsBomb).

Conclusion

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

Bibliography I

[1] 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.

[2] SkillCorner. *SkillCorner Open Data: Broadcast tracking data.* https://github.com/SkillCorner/opendata. Accessed: 2025-01-04. 2020.

RoleCPD: Hamming Distance Switch Rate

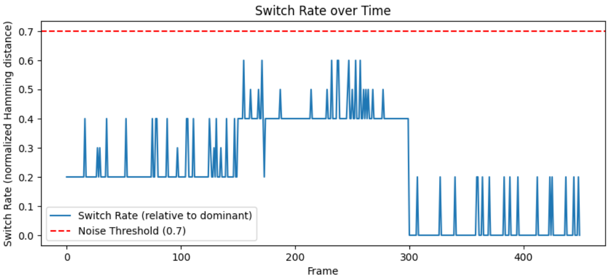

Figure: Hamming distance-based switch rate time series used for change point detection.

RoleCPD: Formation Ground Truth

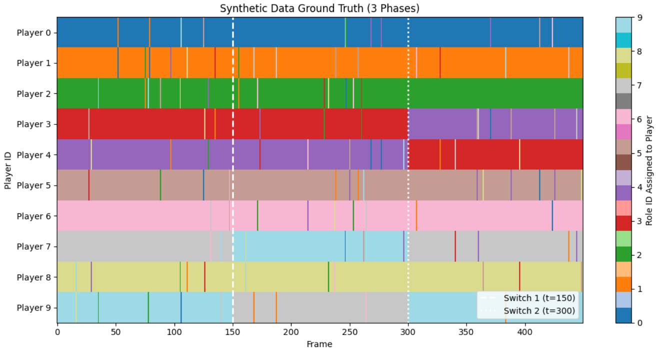

Figure: Ground truth role assignments over time with two tactical switches at frames 150 and 300.

RoleCPD: Need for Recursive Segmentation

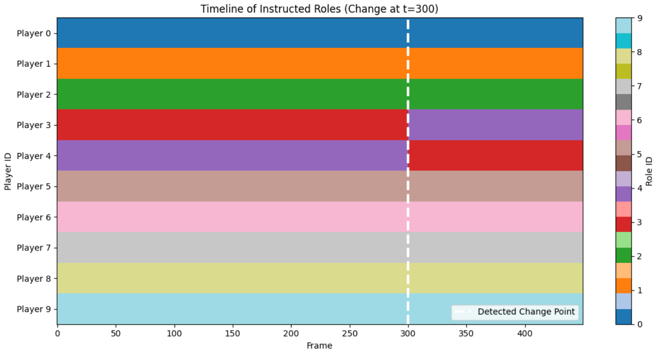

Figure: Visualization of one detected tactical phase with dominant role assignments. This shows the need for recursive segmentation to capture multiple phases.

Stationarity: FC Barcelona Phase 2 (Event Data)

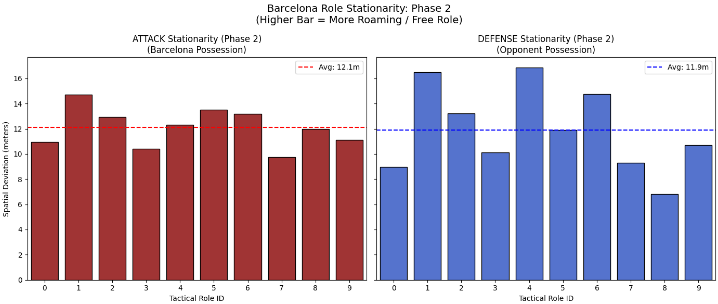

Figure: Player stationarity during Phase 2 of FC Barcelona match, illustrating the consistency of player positions within the tactical phase, even with event-stream data.