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** [Kim+22], 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.

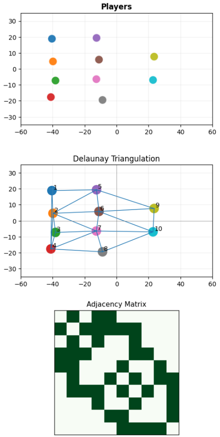

Figure: Pipeline

RoleCPD: Detecting Intraphase Swaps

Objective: Detect permanent role swaps (e.g., Wingers switching sides) within a stable formation.

The Metric: Switch Rate

- Input: Sequence of role permutations π_t.
- **Hamming Distance:**

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

- Measures deviation from the "Dominant Permutation".

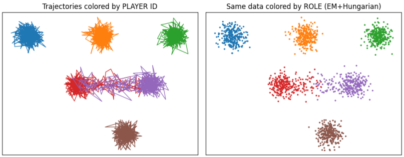

Figure: Synthetic Role Swap Data

RoleCPD: Recursive Segmentation

Discrete g-segmentation on Permutations

1. **Pre-process:** Filter frames with Switch Rate > 0.7 (temporary noise).
2. **Recursive Search:**
   - Apply graph-based change-point detection on the sequence.
   - **Stopping Criteria:** p-value > 0.01 or segment length < 5 min.
3. **Result:** Identifies distinct tactical role configurations.

High-Frequency vs. Sparse Event Data

1. **Tracking Data (GPS)**
   - **Source:** SkillCorner (10Hz) [Ski20], Last Row (20Hz) [TF20].
   - **Quality:** Dense, clean signals.
   - **Prep:** Normalized to centered metric pitch.

2. **Event Data (StatsBomb)**
   - **Source:** On-ball events only.
   - **Challenge:** Extreme sparsity (<10 pts/min vs $600+$ for GPS).
   - **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

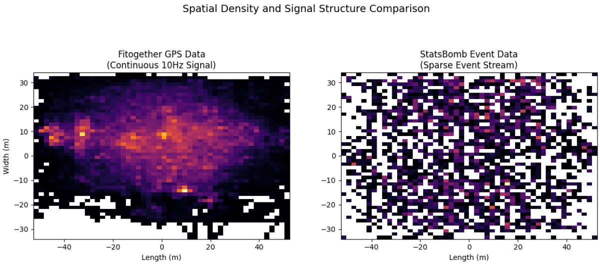

Figure: Heatmap: Dense GPS coverage vs. Sparse/Biased Event Data.

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 (GPS)

Problem: A single "average" centroid formation ignores game state.

Method:

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

Insight:

- Defense: Compact block.
- Attack: Wide, expansive shape.

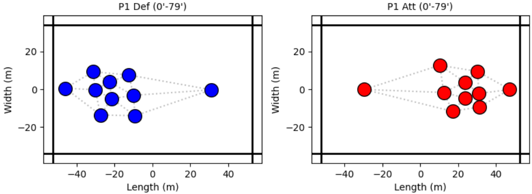

Figure: Match 1925299: Distinct shapes detected.

Quantifying Role Fluidity (σ)

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

GPS Data Findings (A-League)

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

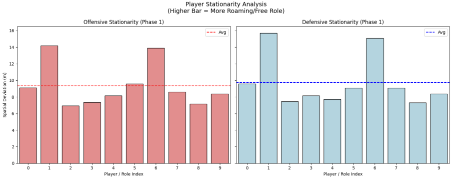

Figure: Stationarity profile (GPS Data).

Offensive vs. Defensive Structures (FC Barcelona)

We extended the "Possession Context" analysis to sparse StatsBomb data.

Tactical Shapes in Sparse Data

- Despite low data density, distinct shapes emerge when aggregated over time.
- **Defense:** High density in central zones (low width).
- **Attack:** Wingers push high and wide; full-backs overlap.

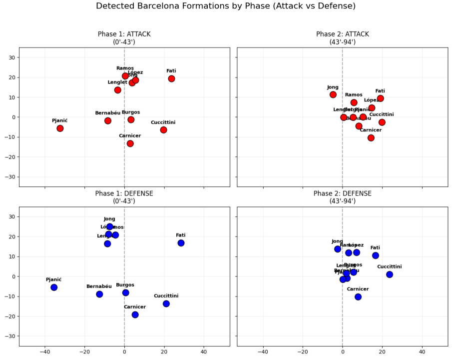

Figure: Barcelona: Attack (Red) vs Defense (Blue).

Quantifying Fluidity in Event Data

Does the "Stationarity Metric" hold up with sparse event streams?

Comparison Results

- **Consistency:** The "Fixed" vs. "Fluid" distinction persists.
- **Phase Analysis (Barcelona):**
  - *Attack Phase:* High deviation (Avg 12m) → Fluidity.
  - *Defense Phase:* Lower deviation (Avg 9.6m) → Rigidity.

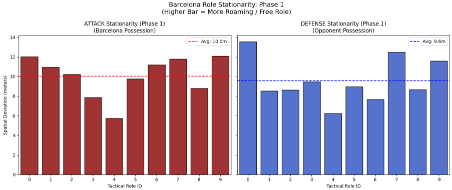

Figure: Stationarity: FC Barcelona (Event Data).

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 I

[Ski20] SkillCorner. *SkillCorner Open Data: Broadcast tracking data*. `https://github.com/SkillCorner/opendata`. Accessed: 2025-01-04. 2020.

[TF20] Ricardo Tavares and Friends of Tracking Data. *Last-Row: Sample tracking data and code*. `https://github.com/Friends-of-Tracking-Data-FoTD/Last-Row`. Accessed: 2025-01-04. 2020.

Bibliography II

[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.