Classifying Styles of Attacks for the FC Barcelona Soccer Club

Eric Bradford, Alex Cauneac, Emma Chesley, Sean Ko, Kevin Lee, Garrett Souza, and Tim Yang

Introduction

Through incorporating spatial, contextual, and statistical information into the clustering framework, we developed a holistic assessment of team attacking tendencies. We have created and refined a model which clusters opponent attacks into homogeneous subsets. This model is capable of assessing the degree of similarity across team attacking tendencies, providing information on which teams in La Liga have the most comparable attacks. Our deliverables to FC Barcelona include a scouting report and python library.

Clustering Methodology and Results

We trained a model using the Expectation Maximization Clustering method, which computes the probability of a data point belonging to a certain cluster. We determined the optimal number of clusters to be 16 using the Bayesian Information Criterion Score, which minimizes the variance within each cluster while penalizing the total number of clusters.

Feature Engineering

We were given 506 events data sets. For the clustering model, we represented each attack as a 21-dimensional feature vector. The 21 features fell into 3 categories: macro-events, passing sequence flow motifs, and spatial region features.

CLUSTER	CHARACTERISTICS	LOCATION	RESULT
1	Many Passes	Defensive and Middle Third	Long Ball
2	Short, sloppy	Defensive Third	Long Clearance
3	Quick	Right Flank	Cross
4	Quick	Left Flank	Cross
5	Short passes	Defensive Third	Dispossessed in Midfield
6	Patient, Many Passes	Midfield	Never Reaches Final Third
7	Long, Many Passes	Midfield	Enters Final Third
8	Patient, Many Passes	Center and Left of Defensive and Middle Thirds	Never Reaches Final Third
9	Long Distribution From Goalkeeper/Defender	Final Third	Fail to Reach Second Ball
10	Long Distribution From Goalkeeper/Defender	Right Side	Shot or Cross
11	Clearance to Sideline	Defensive Third	Ball Goes Out Before Half Field
12	Dribble through lines	Left Flank	Cross
13	Set Pieces, Throws, Corners	N/A	Cross
14	Patient, Pass Through Lines	Defensive Third	Cross
15	Patient, Back Passes	Midfield	Breaking Lines Into Final Third
16	Patient, Break Through Lines	Defensive and Middle Third	Rarely Reach Final Third

- **Macro-events** give a high-level overview of the attack. Features in this component describe attack characteristics such as duration, number of passes, speed, directness, and success.
- **Passing sequence flow motifs** describe how players interact in an attack. Each attack was divided into overlapping 3 pass sequences and player specific information was removed.
- **Spatial region features** provides details on the areas of the field that a team uses in the attack. We divided the field into a 3x3 grid of equally sized cells and calculated the percentage of time the attack spends in each cell.

Scouting Report

The scouting report aggregates team-specific conclusions from the clustering model into a concise format. The graphic provides a visual representation of a team's attacking tendencies, highlighting clusters which were deemed most relevant for that team. The section listing teams within La Liga with similar attacking tendencies adds context to the clustering breakdown of the chart, and is useful information for coaches and team members who know the attacking styles of various teams in La Liga.