

Detecting tactical patterns in football with fast search and density peak clustering

Michael Ståhle*

February 2025

Abstract

Sports analytics has seen rapid growth, supported by advancements in data acquisition and computational methodologies. This thesis focuses on clustering football possession chains using on-ball event data to detect and interpret tactical patterns. We adopt the Fast Search and Density Peaks (FSDP) algorithm (a distribution-free, density-based clustering method) to group similar ball trajectories. Two distance measures, the Fréchet distance and the Longest Common Subsequence (LCSS) distance, are assessed based on their theoretical properties and alignment with the assumptions of the FSDP algorithm. Our empirical analysis demonstrates that the Fréchet distance, which is metric-based and sensitive to continuous shape variations, provides more coherent clustering results than the LCSS distance. Using the Fréchet distance, four distinct tactical patterns emerge from the possession data of Arsenal F.C. during the 2017/18 Premier League season. Notably, attacks along the left flank produce nearly twice as many goals as those on the right flank, suggesting a clear advantage when exploiting that side of the pitch. These findings highlight the importance of selecting distance measures that both adhere to metric properties and capture the nuances of the underlying data structure. They also illustrate the potential of density-based methods to uncover meaningful tactical insights, paving the way for further methodological refinements (such as alternative distance measures, fuzzy cluster assignment, and inclusion of player tracking data) that could deepen our understanding of football strategy.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden.
E-mail: michael.staahle@gmail.com. Supervisor: Chun-Biu Li.