

Project

Gustav Sternelöv
25 april 2016

The data - WhoScored and Opta

Definition of the metrics (shots, key passes etc.)
Time period for the data. 2015/16 season up to round...
Limitations such as the number of games played, only time when used as a striker etc.
How data is collected
All metrics are per 90 minutes. Average contribution/performance per 90 minutes.

Purpose

Why only strikers (narrow down the project), why only top leagues (availability of data)
Investigate if the strikers can be divided into different kinds of strikers.
Use for analysis of strikers, in the context of scouting, as well as for player development. "This player is supposed to be a target player but only wins ... aerials per 90 minutes."

Types of strikers

<http://www.guidetofootballmanager.com/tactics/strikers>

Target man
Complete forward
Trequartista
False nine
Defensive forward
Deep-lying forward

<http://businessdayonline.com/2014/12/five-dynamics-of-modern-soccer-strikers/>

Target man
Complete forward
Trequartista
Poacher
Advanced forward

Mine would be *target man*, *complete forward*, *trequartista*, *poacher* and *false nine*. Some more link could be useful.

Would like to say that there are at least three different kinds of forwards and at most six different types.
Does then imply that the k-means should be prespecified with k equal to either of 3, 4, 5 or 6.

Characteristics of methods

Actual methods are k-means, hierarchical clustering and DBScan. Is of interest to look more closely at the characteristics of each of these methods.

K-means:

Hierarchical clustering:

DBScan: