A cool dataset

SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

A public data set of spatio-temporal match events in soccer competitions

Luca Pappalardo[1*], Paolo Cintia[2], Alessio Rossi[2], Emanuele Massucco[3], Paolo Ferragina[2], Dino Pedreschi[2] & Fosca Giannotti[1]

Soccer analytics is attracting increasing interest in academia and industry, thanks to the availability of sensing technologies that provide high-fidelity data streams for every match. Unfortunately, these detailed data are owned by specialized companies and hence are rarely publicly available for scientific research. To fill this gap, this paper describes the largest open collection of soccer-logs ever released, containing all the spatio-temporal events (passes, shots, fouls, etc.) that occured during each match for an entire season of seven prominent soccer competitions. Each match event contains information about its position, time, outcome, player and characteristics. The nature of team sports like soccer, halfway between the abstraction of a game and the reality of complex social systems, combined with the unique size and composition of this dataset, provide an ideal ground for tackling a wide range of data science problems, including the measurement and evaluation of performance, both at individual and at collective level, and the determinants of success and failure.

https://www.nature.com/articles/s41597-019-0247-7