

A public data set of spatiotemporal match events in soccer competitions

島崎 圭介

富山県立大学 情報システム工学科

2024 年 7 月 5 日

はじめに

近年サッカー分析は毎試合高精度のデータ取得のおかげで学界や産業で関心が高まっている。

しかし、これらの詳細なデータは専門企業が保有しているため科学的研究のために一般公開されることはほとんどない。その差を埋めるために今回各試合発生したすべての時空間イベントを記録しデータ科学の問題に取り組むためのものを紹介する。

はじまり

サッカーの分析は長い間注目を集めてきました.1950 年代初頭にイギリスでは, ボールをできるだけ早く後ろから前にパスすることによって得点する確率が上がるという統計データを手作業で取得した. これがサッカー分析の始まりともいわれている.

概要

サッカーの統計がさらに発展したのは、すべての試合から抽出されるデータの量が増えたからである。

試合中に発生するイベントを説明し、独自のタグ付けソフトウェア、選手の軌跡を説明し、ビデオ録画するもの、GPS データを通じて収集されるものとこれらの3つは一般的に公開されていないものとなっている。これらによって科学的な研究の足枷になっている。

今回男子サッカーの主要7大会を網羅したオープンサッカーログコレクションについて説明します。このデータは一般に公開されたものとして最大のものとなっており、試合イベントが記述されており、各イベントにはその種類に関する情報が含まれています。そのデータをを用いて今回新たな情報収集を提案します。

提案手法

サッカー業界の大手企業である Wyscout によって収集および提供されているものを使用する。データ収集の手順は、タグ付けツールを使用して実行する、

1. フォーメーションの設定をおこなう。

試合開始時にチームのスターティングフォーメーション、ピッチ上の選手の位置、および選手の背番号を設定します。

2. イベントのタグ付け

試合中のボールタッチごとに選手1人を選択し、タイムライン上に新しいイベントを作成する。

提案手法

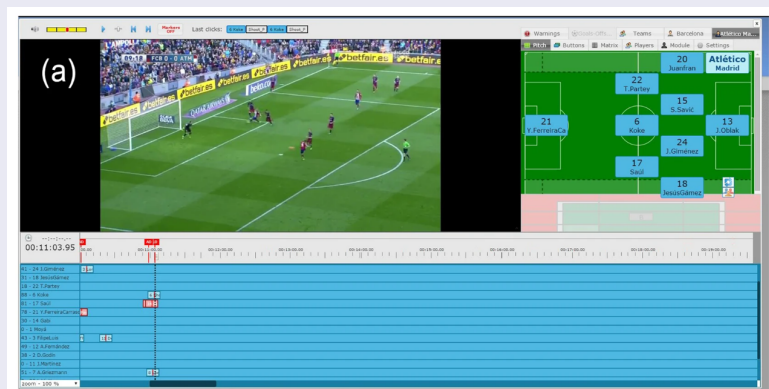


図 1: タグ付け

提案手法



図 2: イベントの追加

提案手法

今回公開するデータはヨーロッパの5つの国内サッカー大会の2017/2018 シーズンに関するものである。さらにワールドカップ2018 とヨーロッパカップ 2016 のデータも公開する。合計で、すべての大会、試合、チーム、選手、イベント、審判、コーチに関する情報に対応する7つのデータセットを提供する。データは、合計で約1941 試合, 3251294 イベント, および 4299 人のプレーヤーの情報がある。

評価

はじめに
提案手法
提案手法

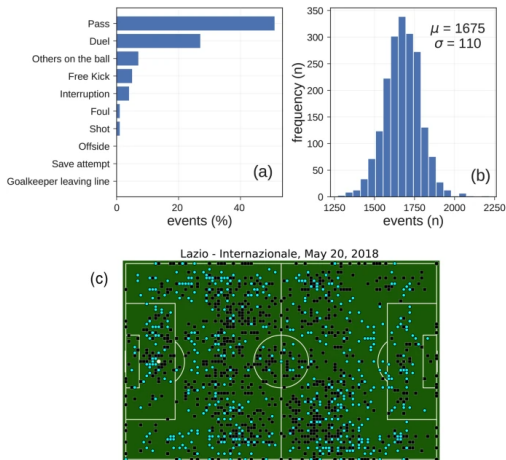


図 3: (a) イベントの発生率 (b) イベント数の分布 (c) 発生したイベントの位置

まとめ

今回一般向けに情報を公開している人の論文を読んでこのデータをどのように取得したかを理解した. このデータを使ってどのように解析すればよいかを考えるという方向に決めた. このデータを取得するプログラムは著者にメールを送って取得するつもりである.