Football-Analytics-101

Jun 30, 2020

Contents:

1 Introdution **1**

2 Lessons & Papers **3**
- 2.1 Lessons . . . 3
- 2.2 Selected Papers . . . 3
- 2.3 People . . . 4

3 Data Sources & Competitions **5**
- 3.1 Data Sources . . . 5
- 3.2 Training Environments . . . 5
- 3.3 Competitions . . . 6

4 Football Analytics Companies & Products **7**

5 Sports Analytics Reference **9**
- 5.1 Basketball (Mainly on NBA) . . . 9
- 5.2 Amerinan Football (Mainly on NFL) . . . 9
- 5.3 Baseball . . . 10
- 5.4 Lacrosse . . . 10

CHAPTER 1

Introdution

I create this doc in order to provide ordered, collected and detailed, information about football (soccer) analytics for research works and interesting stuff, activated by NBAStuffer-Analytics-101. Collected information contains statistical analysis as well as AI research, which will be updated step-by-step. If you have any idea or new findings, feel free to commit to this doc or contact me via the email shown on my github.

So far, I have collected informations from the following aspects:

- *Data Sources & Competitions*: Useful public data source, some are free while others must be bought from data companies. There are also sports analytics challenge and competitions included, some of which offers off-line data for testing insights on.
- *Lessons & Papers*: Helpful lessons of Sports Analytics and interesting research papers published on important AI conference and MIT Slogan Sports Analytics Conference.
- *Football Analytics Companies & Products*: Football analytics companies, including general analytics companies that cover other sports. These companies are mainly data collectors, statistical analyzers and special solution providers. Companies using AI (Artificial Intelligence) are valued much more. Related product demo vediosa are also contained.
- *Sports Analytics Reference*: Some useful reference webpages that provide interesting analytics information on other sports, like basketball, American football, lacrosse, baseball and others.

CHAPTER 2

Lessons & Papers

Helpful lessons of Sports Analytics and interesting research papers published on important AI conference and MIT Slogan Sports Analytics Conference. I also add some researchers who are pioneers to sports analytics.

2.1 Lessons

CSC2541 CSC2541 is a graduate course in machine learning for Sport Analytics from Toronto University, which contains useful resources.

MIT Slogan Sports Analytics Conference Videos Here records many of the speeches in MIT Slogan Sports Analytics Conference (SSAC) sicne 2012.

2.2 Selected Papers

Tracking data driven

<Data-Driven Ghosting using Deep Imitation Learning>, 2017. Published on MIT-Slogan Sports Analytics Conference (SSAC). TBA.

<BasketballGAN: Generating Basketball Play Simulation Through Sketching>, 2019. Published on ACM MultiMedia (ACM MM). TBA.

Statistical data driven

<Actions Speak Louder than Goals: Valuing Player Actions in Soccer>, 2019. KDD best paper of application track. TBA.

Visual data driven (Mostly Tracking)

<Multiple Object Tracking in Soccer Videos using Topographic Surface Analysis>, 2019. TBA.

<Tracking Multiple People in a Multi-Camera Environment>, 2018. TBA.

<Soccer: Who Has The Ball? Generating Visual Analytics and Player Statistics>, 2018. Published on CVPR workshop. TBA.

<A Survey on Player Tracking in Soccer Videos>, 2017. TBA.

<A Survey on Content-aware Video Analysis for Sports>, 2017. TBA.

<Event Recognition in Broadcast Soccer Videos>, 2016. Published in ICVGIP16, Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing.

<Detecting events and key actors in multi-person videos> By Google and Stanford researchers (Feifei Li).

<Learning to Track and Identify Players from Broadcast Sports Videos>, 2012. Published on IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE.

<Ball Tracking and Action Recognition of Soccer Players in TV Broadcast Videos>, 2007. This is a theis which detailed describe various methods of Video processing in football analysis. TBA.

<Visually Tracking Football Games Based on TV Broadcasts>, 2007. Published in IJCAI07. This paper is an extension of <ASPOGAMO: Automated Sports Game Analysis Models>. They propose ASPOGAMO, a visual tracking system that determines the coordinates and trajectories of football players in camera view based on TV broadcasts.

<Players and Ball Detection in Soccer Videos Based on Color Segmentation and Shape Analysis>, 2007. Published in MCAM 2007. This paper proposes a scheme to detect and locate the players and the ball on the grass playfield in soccer videos.

<Soccer video analysis by ball, player and referee tracking>, 2006. TBA.

2.3 People

Patrick Lucey is currently the VP of Artificial Intelligence at STATS, whose research interests cover artificial intelligence and interactive machine learning in sporting domains. Many of his publications which can be found on his webpage coincide with my interests.

CHAPTER 3

Data Sources & Competitions

Useful public data source, some are free while others must be bought from data companies. In addition, sports analytics challenge and competitions are collected, some of which offers off-line data for testing insights on. There are also some training environments that can be used for reinforcement learning included. These collections mainly focus on football while some may cover other sports.

3.1 Data Sources

KaggleData KaggleData is kept by Kaggle, the world's famous data mining competition platform. There are kinds of football data while they also keeping basketball data and cricket data.

FootballData Jokecamp created this repository of Football/Soccer data for anyone to use. He save them here as he find them or build the files. Data is in mostly csv and json formats.

Free football data from StatsBomb This repository is committed by StatsBomb to sharing new data and research publicly to enhance understanding of the game of Football. They want to actively encourage new research and analysis at all levels. Therefore they have made certain leagues of StatsBomb Data freely available for public use for research projects and genuine interest in football analytics.

3.2 Training Environments

Google Research Football: A Novel Reinforcement Learning Environment Google Research Football is an RL environment based on open-source game Gameplay Football. Read the related paper here.

DeepMind MuJoCo Multi-Agent Soccer Environment DeepMind MuJoCo Multi-Agent Soccer Environment, a simulator which is a 2v2 soccer game using the MuJoCo physics engine. Read the related paper here.

3.3 Competitions

PSG Sports Analytics Challenge PSG Sports Analytics Challenge (finished) is a football analytics challenge held by Paris Saint-Germain and École Polytechnique. This challenge provides Opta event data and InStat tracking data, which can be only download by those participants, and ask participants to predict something like possession and design useful tools for football analytics.

Some fans have open their solution to this competition on github with more details about this competition:

1. https://github.com/Ericonaldo/Sports-Analytics-Challenge
2. https://github.com/Logosxxw/challenge_PSG
3. https://github.com/dam-grassman/psg-challenge

Football Player's Worth Estimation Football Player's Worth Estimation (open) is an exercise, which asks people to predict the player's market value based on the player's information and ability values. This competition has public dataset that is available to download until today.

March Machine Learning Mania (Google Cloud & NCAA® ML Competition) These are challenges on Kaggle, which ask data scientists to use machine learning methods to predict winners and losers of the men's 2016 NCAA basketball tournament since 2014. There are many open source solutions and discussions. The dataset can be download until now and they also keep updating the NCAA basketball dataset as far back as 1894. Competitions are listed bellow:

1. March Machine Learning Mania
2. March Machine Learning Mania 2015
3. March Machine Learning Mania 2016
4. March Machine Learning Mania 2017
5. Google Cloud & NCAA® ML Competition 2018-Men's
6. Google Cloud & NCAA® ML Competition 2018-Women's
7. Google Cloud & NCAA® ML Competition 2019-Men's
8. Google Cloud & NCAA® ML Competition 2019-Women's

CHAPTER 4

Football Analytics Companies & Products

Football analytics companies, including general analytics companies that cover other sports. These companies are mainly data collectors, statistical analyzers and special solution providers. Companies using AI (Artificial Intelligence) are valued much more. Related product demo vediosa are also contained.

SAP SAP is a German multinational software corporation founded in 1972 that makes enterprise software to manage business operations and customer relations. Due to its powerful ability of cloud computation, SAP provides numerous kinds of solutions, including data analytics and database. Specifically, SAP offers SAP-Sports-One as a sports team management platform based on SAP-HANA, which helps Germany win the 2014 FIFA world cup.

There are some demos about how the platform works, please click here and here.

InStat InStat is a sports performance analysis company founded in Moscow, Russia in 2007, providing professional tools and commercial data for performance evaluation, scouting, fitness analysis and a panoramic filming technology. InStat aims to support football professionals with statistical data. It offers statistical reports and online video platform called InStat Scout which allows coaches to access stats linked to the supporting videos of actions. InStat recently provides tracking data support with 2D model of the match, which is also introduced to ice hockey and basketball.

Demos for InStat Scout can be found here.

Opta Sports Opta, founded in 1996 and has been operated as a wholly owned subsidiary of Perform Group since 2013, is a British sports analytics company based in the United Kingdom. Opta provides data for 30 sports in 70 countries, with clients ranging from leagues to broadcasters and betting websites. Opta offers detailed data for not only footballs but also other sports. They record event and tracking data and sell it to companies and individuals. They hold Sports Analytics Forum every year, in which many football analysist show their exciting researches. Some vedios from OptaPro Analytics Forum 2019 can be found here, from which one can draw interesting ideas.

Camvision Camvision establised in 2007, Czech, is a technology company focused on providing intelligent video-capture systems for sporting events. They claim to have extensive research experience in the sports video field and they are able to deliver automated solutions for sports-oriented video analysis through SoTA technology. Their product called PANORIS will record the match autonomously with analyzing every movement.

Go to their homepage for demos or visit here.

StatsBomb StatsBomb are a team with headquarter in Bath, UK and offices in Boston and Cairo, founded in 2016. It is a team of analysts, data scientists, computer engineers, with leadership experience in Football and Technology. They have recently acquired ArqamFC, an experienced data collection team based in Cairo. Arqam bring technology, dedication and are committed to a shared vision with StatsBomb for football data. They provide training courses in Professional Football Analytics. StatsBomb collects data and use it for building practical models to solve real-world problems.

MT-Sports MT-Sports is a company founded in 2015, China, which aims to develop smart wearable devices to collect sports data, and use it for data visualization and sports data mining. Their products has been applied to Chinese national football team, and has became the first choice of sports education information construction in China.

Champdas Champdas is founded in 2016, Shanghai, China. It is a sports internet company with independent property rights, integrating data collection, data mining and data productization. Each game can collect 15000+ related data and present it in real time on Champdas DATA. They focus on the domestic events to provide live broadcasts of the Super League and the Chinese League, and will also pay attention to the performance of the national team.

Tongdaoweiye Tongdaoweiye is founded in 2015, Beijing, China. It is a football big data company that covers data collection, modeling analysis, data application and information system. They offers data service and transfering service and they claim to have management platform, data platform, wearable device and other technologies. They cooperate with Chinese Football Association and Chinese professional teams.

WYSCOUT WYSCOUT is a data-driven company founded in 2007, Italy. They serve football analysis services and tools and they have built their own dataset since 12 years ago. They have products for providing game vedio data, a platform of game anylysis with statistics and a platform for scout with statistics.

SkillCorner SkillCorner is a computer vision startup founded in 2016, Paris, France, who has built an AI-powered video tracking technology based on deep learning capable of recognizing, positioning and following in real time, the football players, the referee and the ball from any broadcasted game. SkillCorner's tracking algorithm for live broadcasted games is the best available with up to 95% accuracy. From previously unavailable raw tracking data to live match visualization, SkillCorner's products are at the forefront of technological innovation.

CHAPTER 5

Sports Analytics Reference

Some useful reference webpages that provide interesting analytics information on other sports, like basketball, American football, lacrosse, baseball and others.

5.1 Basketball (Mainly on NBA)

NBAstuffer NBAstuffer is a website started out as a hobby-site by Serhat Ugur in 2007, has grown into a reputable stats-reference that delivers unique metrics and NBA analytics content some of which can't be found anywhere else. Supported by dynamic charts and visualizations, it not only provides daily updated advanced NBA stats tables, but also have built popular research tools such as schedule analysis, rest days stats, analytics primer, DraftKings & FanDuel NBA DFS cheatsheet and DFS lineup optimizer. Their recent product - BigDataBall datasets offers cleaned-up, aggregated and enriched box score stats, odds, play-by-play logs, and DFS data in easy to manipulate spreadsheets.

82games 82games is geared towards providing innovative statistical coverage and analysis of the NBA for team executives, coaches, fans, and the media. They have detailed team stats, player stats, statistical reports and other interesting information.

Basketball-reference Basketball-reference constains basketball stats and history statistics, scores, and history for the NBA, ABA, WNBA, and top European competition, which can be used as a stable data source.

Thehothand Thehothand is a blog that posts on analyzing sports streakiness by Texas Tech Professor Alan Reifman. He begins with the Hot Hand theory and even publish a related book. In there, more referece websites that focus on sports statistical analytics can be found.

5.2 Amerinan Football (Mainly on NFL)

Advanced Football Analytics Advanced Football Analytics is a website created by Brian Burke to share research and analysis of NFL football. Instead of opinion and intuition, AFA uses facts and data to make useful analysis. AFA made modern football analytics a reality by pioneering the tools that have become the standard methods for

advanced football analysis. AFA features original research, advanced analysis, novel statistics, practical tools, innovative models, and captivating visualizations.

Football-reference Football-reference constains football stats and history statistics. There are complete source for current and historical NFL, AFL, and AAFC players, teams, scores and leaders, which can be used as a stable data source.

PhD Football PhD Football is matained by Andrew S.-Rook, who is a data scientist at Capital One, working to help banking better. He spend a lot of my (increasingly rare) free time doing NFL analytics.

Fantasy Football Analytics Fantasy Football Analytics is a website for statistical analysis in fantasy football. It is a community for people who want to: improve their performance in fantasy football using stats; learn advanced statistical analysis approaches; or learn the statistical software R. They try to make all of our scripts available for free. One of their contributions is that they propose to calculate "wisdom of the crowd" projections by aggegating multiple sources of projections, which allows users to aggregate the sources by calculating a mean, weighted average, or robust average.

Armchair Analysis Armchair Analysis provides detailed and affordable NFL stats and related analysis for every play of every game for every player since the 2000 season. They've done the heavy lifting: instantly connect 700+ data points across 30 different tables thanks to specific ID's for plays, players and games. and they claim that they don't rely on scraped and regurgitated NFL stats, in addition to charting basic play-data, they chart 38 different custom variables that are available 7 days after each game. QB Pressure, Hits, Hurries; Defenders in the Box; Contested Balls. They offer data with .csv and APIs.

5.3 Baseball

Sean Lahman's Homepage Sean Lahman's Homepage is maintained by Sean Lahman, who has broken new ground in the field of sports statistics, creating historical databases for use in both print and digital projects. His work spans the spectrum of sports: baseball, pro and college football, pro and college basketball, auto racing, tennis, boxing and the Olympic games. His Baseball Archive web site was one of the earliest sources for baseball information on the Internet, and he headed the first significant effort to make a database of baseball statistics freely available to the general public. Lahman also contributed to pioneering efforts at websites like Baseball-Reference.com, Pro-Football-Reference.com, and BasketballReference.com.

Baseball-reference Baseball-reference constains baseball stats and history statistics. There are complete source for current and historical baseball players, teams, scores and leaders, which can be used as a stable data source.

5.4 Lacrosse

Lacrosse Analytics Lacrosse Analytics is a webpage that provides lacrosse analytics.