

Sesión Proyectos

Actividad 1. Exploración de datasets

Jesús Lagos Milla

Telefonica

Instituto Tecnológico Telefónica_

Datasets seleccionados:

1. Fichero de eventos (F24) de un partido de fútbol (Valencia-Real Madrid ene-2018)
2. Ficheros de matrices de pase (2), uno por equipo para el partido citado

Acceso: <https://github.com/Jelagmil/DataF24>

Formato de los ficheros: xml

Proveedor de la información: Stats Perform (Opta)

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Copyright 2001-2018 Opta Sportsdata Ltd. All rights reserved. -->
<!-- PRODUCTION HEADER
     produced on:   valde-job@opta.net
     production time: 20180127T17:13:50Z
     production module: Opta::Feed::XML::Soccer::F24
-->
<!--
-->
<Game timestamp="2018-01-31T21:17:12">
  <Game id="943000" away_score="4" away_team_name="Real Madrid" competition_id="23" competition_name="Spanish La Liga" game_date="2018-01-27T15:15:00" home_score="1" home_team_id="191"
  <Event id="1845328763" event_id="1" type_id="34" period_id="16" min="0" sec="0" team_id="186" outcome="1" x="0.0" y="0.0" timestamp="2018-01-27T14:08:47" last_modified="2018-01-27T14:08:47" version="1"
    <Q id="1524420932" qualifier_id="194" value="39563" />
    <Q id="1969475087" qualifier_id="131" value="1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 0, 0, 0, 0, 0, 0" />
    <Q id="1770462179" qualifier_id="197" value="3894" />
    <Q id="1541035455" qualifier_id="30" value="28411, 88483, 39563, 61256, 90152, 88477, 37055, 44989, 19927, 36903, 14937, 39790, 222784, 107593, 192364, 174292, 199248, 91651" />
    <Q id="2129716943" qualifier_id="59" value="1, 2, 12, 14, 5, 6, 10, 8, 9, 11, 7, 13, 15, 17, 18, 20, 21, 23" />
    <Q id="109848974" qualifier_id="227" value="0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0" />
    <Q id="1856941733" qualifier_id="130" value="4" />
    <Q id="1366046785" qualifier_id="44" value="1, 2, 2, 3, 2, 2, 3, 3, 4, 4, 4, 5, 5, 5, 5, 5" />
  </Event>
  <Event id="1874742238" event_id="1" type_id="34" period_id="16" min="0" sec="0" team_id="191" outcome="1" x="0.0" y="0.0" timestamp="2018-01-27T14:13:57" last_modified="2018-01-27T15:44:38" version="1"
    <Q id="1539139093" qualifier_id="197" value="1597" />
    <Q id="1321330214" qualifier_id="44" value="1, 2, 2, 3, 2, 2, 3, 3, 4, 4, 3, 5, 5, 5, 5, 5" />
    <Q id="1985354522" qualifier_id="30" value="69752, 86153, 221306, 51952, 28525, 54864, 181284, 88754, 80954, 150360, 132105, 116732, 198499, 108093, 60270, 156689, 224860, 230598" />
    <Q id="1394223480" qualifier_id="194" value="51952" />
    <Q id="2034294525" qualifier_id="59" value="13, 21, 15, 10, 24, 17, 7, 16, 19, 22, 14, 1, 6, 8, 9, 11, 18, 30" />
    <Q id="1131089549" qualifier_id="227" value="0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0" />
    <Q id="1383308461" qualifier_id="131" value="1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 0, 0, 0, 0, 0" />
    <Q id="1281883221" qualifier_id="130" value="2" />
  </Event>
  <Event id="1251517710" event_id="2" type_id="32" period_id="1" min="0" sec="0" team_id="191" outcome="1" x="0.0" y="0.0" timestamp="2018-01-27T15:15:10.767" last_modified="2018-01-27T15:15:11" version="1"
    <Q id="1344952635" qualifier_id="127" value="Left to Right" />
  </Event>
  <Event id="1716312310" event_id="2" type_id="32" period_id="1" min="0" sec="0" team_id="186" outcome="1" x="0.0" y="0.0" timestamp="2018-01-27T15:15:10.767" last_modified="2018-01-27T15:15:11" version="1"
    <Q id="1551813265" qualifier_id="127" value="Right to Left" />
  </Event>
  <Event id="1688015133" event_id="3" type_id="1" period_id="1" min="0" sec="1" player_id="80954" team_id="191" outcome="1" x="49.8" y="51.2" timestamp="2018-01-27T15:15:11.767" last_modified="2018-01-27T15:15:11.767" version="1"
    <Q id="179918942" qualifier_id="213" value="3.5" />
    <Q id="1639188950" qualifier_id="279" value="8" />
    <Q id="1573955245" qualifier_id="140" value="33.0" />
    <Q id="1533869866" qualifier_id="56" value="Back" />
    <Q id="1835951361" qualifier_id="141" value="41.0" />
    <Q id="190986143" qualifier_id="212" value="19.0" />
  </Event>
  <Event id="1165699801" event_id="4" type_id="1" period_id="1" min="0" sec="2" player_id="51952" team_id="191" outcome="1" x="33.0" y="41.0" timestamp="2018-01-27T15:15:13.406" last_modified="2018-01-27T15:15:13.406" version="1"
    <Q id="1143843085" qualifier_id="213" value="1.8" />
    <Q id="1366225267" qualifier_id="141" value="68.0" />
    <Q id="1364845060" qualifier_id="140" value="39.0" />
    <Q id="1825594666" qualifier_id="56" value="Back" />
    <Q id="1369496450" qualifier_id="212" value="19.5" />
  </Event>
</Game>
```

Ejemplo de fichero F24

Fichero F24. Eventos de un partido



F24

- Fichero de eventing de cada partido
- 380 partidos por temporada
- Xml de ~11k líneas por partido
- Transformado a dataframe son cerca de 2k eventos y 250 columnas por partido

```
gol
<Event id="1811957136" event_id="316" type_id="16" period_id="1" min="26" sec="57" player_id="242831" team_id="174" outcome="1"
  <Q id="1267782932" qualifier_id="25" /> corner
  <Q id="1198279277" qualifier_id="103" value="3.2" /> altura
  <Q id="1838396608" qualifier_id="231" value="52.9" /> Posición z portero
  <Q id="1612651454" qualifier_id="29" /> asistido
  <Q id="1119577594" qualifier_id="61" /> Área pequeña izquierda
  <Q id="1107292632" qualifier_id="102" value="45.8" /> Posición y portero
  <Q id="1400046745" qualifier_id="154" />
  <Q id="2144578788" qualifier_id="230" value="99.0" />
  <Q id="1588250953" qualifier_id="55" value="314" /> Evento anterior
  <Q id="1756663226" qualifier_id="328" /> Golpeo primer toque
  <Q id="1541297950" qualifier_id="56" value="Center" />
  <Q id="1799251201" qualifier_id="80" />
  <Q id="1976485792" qualifier_id="15" /> Remate cabeza
</Event>
```

Explicación de un evento de un partido

Event_ID	Name	Event_ID	Name	Event_ID	Name	Event_ID	Name
1	Pass	17	Card	39	Temp_Attempt	59	Keeper Sweeper
2	Offside Pass	18	Player Off	40	Formation change	60	Chancemissd
3	Take On	19	Player on	41	Punch	61	Ball touch
4	Foul	20	Player retired	42	Good skill	63	Temp_Save
5	Out	21	Player returns	43	Deleted event	64	Resume
6	Corner Awarded	22	Player becomes goalkeeper	44	Aerial	65	Contentious referee
7	Tackle	23	Goalkeeper becomes player	45	Challenge	66	Possession Data
8	Interception	24	Condition change	47	Rescinded card	67	50/50
8	Interception	25	Official change	49	Ball recovery	68	Referee Drop Ball
8	Interception	27	Start delay	50	Dispossessed	69	Failed to Block
9	Turnover	28	End delay	51	Error	70	Injury Time Announcement
10	Save	30	End	52	Keeper pick-up	71	Coach Setup
11	Claim	32	Start	53	Cross not claimed	72	Caught Offside
12	Clearance	34	Team set up	54	Smother	73	Other BallContact
13	Miss	35	Player changed position	55	Offside provoked	74	Blocked Pass
14	Post	36	Player changed Jersey number	56	Shield ball opp an opponent as it rolls out of play	75	Delayed Start
15	Attempt Saved	37	Collection End	57	Foul throw-in	76	Early end
16	Goal	38	Temp_Goal	58	Penalty faced	77	Player Off Pitch
						79	Coverage interruption

Tipos de Evento en el fichero

Telefónica

Instituto Tecnológico Telefónica

Fichero Pass Matrix. Pases entre jugadores



- Fichero de pases entre jugadores de cada partido
- 2 ficheros por partido para los 380 partidos por temporada
- Xml de ~300 líneas por partido

Pass_matrix

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Copyright 2001-2018 Opta Sportsdata Ltd. All rights reserved. -->
<!-- PRODUCTION HEADER
      produced on:      valde-jobq-a02.nexus.opta.net
      production time:   20180131T211714,368Z
      production module: Opta::Feed::XML::Soccer::F27
-->
<SoccerFeed act_minute="92" act_period="2" act_second="59" away_team_id="186" away_team_name="Real Madrid" competition_id="27" competition_name="La Liga"
  <Player cross_lost="1" cross_success="1" jersey_num="18" pass_lost="2" pass_success="23" player_id="224860" player_name="Carlos Soler" position="Midfielder"
    <Player player_id="86153" player_name="Martin Montoya">9</Player>
    <Player player_id="28525" player_name="Ezequiel Garay">3</Player>
    <Player player_id="150360" player_name="Santi Mina">2</Player>
    <Player player_id="88754" player_name="Geoffrey Kondogbia">2</Player>
    <Player cross="1" player_id="132105" player_name="José Gayá">1</Player>
    <Player player_id="51952" player_name="Daniel Parejo">5</Player>
    <Player player_id="221306" player_name="Toni Lato">1</Player>
  </Player>
  <Player cross_lost="0" cross_success="0" jersey_num="13" pass_lost="5" pass_success="14" player_id="69752" player_name="Neto" position="Goalkeeper"
    <Player player_id="224860" player_name="Carlos Soler">1</Player>
    <Player player_id="56864" player_name="Francis Coquelin">1</Player>
    <Player player_id="80954" player_name="Rodrigo Moreno">1</Player>
    <Player player_id="28525" player_name="Ezequiel Garay">4</Player>
    <Player player_id="88754" player_name="Geoffrey Kondogbia">1</Player>
    <Player player_id="51952" player_name="Daniel Parejo">3</Player>
    <Player player_id="221306" player_name="Toni Lato">3</Player>
  </Player>
```

Jugador que realiza el pase

Jugadores que reciben el pase y número de pases

Caso de uso

Objetivo: Cuadro de mando de visualización de fases y métricas de juegos a partir del fichero F24 y Pass Matrix

Descripción: El mundo de los datos ha llegado al deporte para complementar los análisis tradicionales (tácticos y video) y poder aplicarlos a áreas del fútbol como:

- Rendimiento físico
- Rendimiento y métricas de juego
- Scouting
- Marketing

Gracias al fichero F24 y pass matrix podemos crear una herramienta de visualización para el cuerpo técnico que ayude a la toma de decisiones, así como la construcción de nuevas métricas, como xG (goles esperados a partir de modelos ML), xA (Asistencias Esperadas), xT, xBuildUp, xChain y entender los datos de un partido a partir por ejemplo de análisis de redes (SNA).

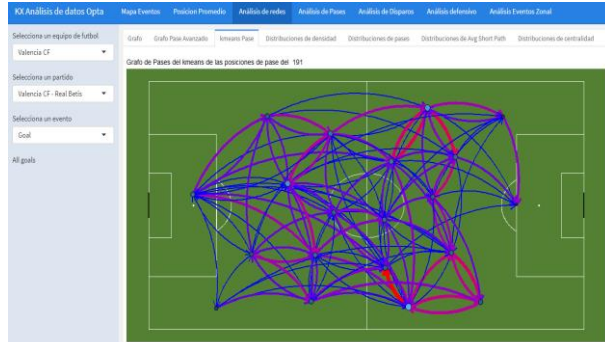
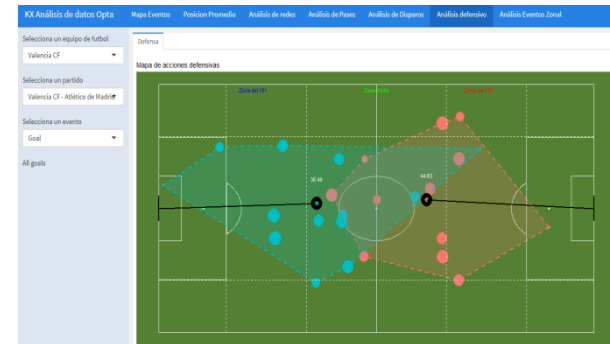
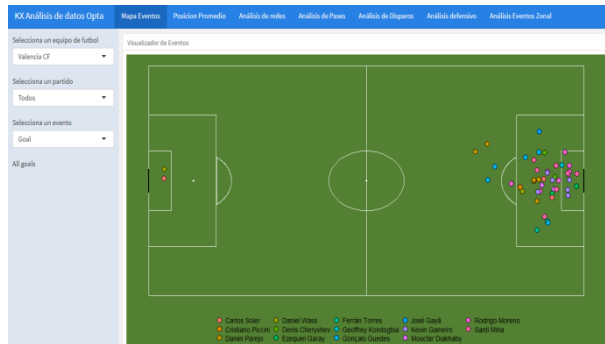
Mockup de la herramienta

Esta imagen son las funciones esperadas a obtener con el fichero F24 y Pass Matrix. Corresponden a la librería de R “soccergraphR” de la que soy creador y se puede descargar en <https://github.com/Jelagmil/soccergraphR>

A partir de dichas funciones construir un CdM que muestre:

- Localización de eventos en un campo de fútbol
- Posición promedio de los jugadores a partir de sus eventos
- Análisis de redes:
 - Grafo con posición promedio como nodo, conexiones, intermediación y densidad para todos los pases
 - Similar pero para pases avanzados
 - Grafo donde la posición no es del jugador si no de zona de pase con un modelo de segmentación como kmeans parametrizando “k”.
 - Análisis de pases con mapas polar, de calor y flechas
 - Análisis de disparos con visión frontal, aérea, lanzamientos de corners y análisis de contras (robo-disparo en segundos)
 - Análisis defensivo a partir de los eventos defensivos realizando un Convex-hull, marcando centroide y distancia a portería.
 - Análisis de eventos por zonas por partido y cada 15’
 - Modelo Expected Goal a partir de base de datos de disparos. GBM o xgBoost.

Mockup de la herramienta



Telefónica

Instituto Tecnológico Telefónica