Automating insight extraction from football data visualizations

Football data visualizations convey an incredible amount of information very quickly. Moves, passing directions and, virtually, every type of event that can be recorded in a football match is able to be represented on a graph.

A simple method to extract insight from a certain type of visualization is to execute it over two different datasets (events by two different players, or just by a player when playing in two different positions) and compare the outputs.

We define metrics that compare visualizations (heatmaps & passing sonars) to automate this process of extracting insight.

Heatmap Comparisons
A heatmap is a measure on the pitch. We compare two heatmaps by measuring the cost of optimally transporting one to the other (this is the earth mover distance).

David Silva’s, James Maddison’s and Christian Eriksen’s heatmaps in EPL 2018/19

Sterling’s, Hazard’s and Mané’s passing sonars of all their passes in EPL 2018/19

Passing Sonar Comparisons
A passing sonar is a 16-dimensional vector with positive components that add up to 1. We compare two passing sonars via a modified \( l_1 \) distance that penalises differences in distant components more than close-by ones.

Bivariate Comparisons

Expected Passing Sonars
The expected passing sonar of a player given his heatmap is the passing sonar an average player would have if his heatmap coincided with the heatmap of the player under study.

Firmino’s (top-left), Kane’s (top-right), Rondón’s (bottom-left) and Wilson’s (bottom-right) heatmaps and passing sonars, respectively.

Compare Firmino to other EPL strikers with, at least, 2000 minutes on field last season.

<table>
<thead>
<tr>
<th></th>
<th>Rank based on distance to Firmino’s heatmap</th>
<th>Rank based on distance to Firmino’s passing sonar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kane</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>Rondón</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Wilson</td>
<td>16</td>
<td>1</td>
</tr>
</tbody>
</table>

High predictability: David Brooks
Average predictability: Richarlison
Low predictability: Felipe Anderson

(Top: Real passing sonar; bottom: expected passing sonar given the player’s heatmap)