The fullback is the most modernized position in soccer today in relation to fullbacks of years past; today’s fullbacks are technically adept, tactically flexible, and have a much higher propensity to enter the attack. While tactical wrinkles such as underlapping and inverted fullbacks have swept through Europe, the overlapping run of a fullback remains an important way to generate chances from lower quality positions. Using purely raw tracking data, I will attempt to generate an exhaustive list of all overlapping runs that fullbacks undertook throughout the course of a game, then quantify them against each other. NB: I got this data from a confidential source, so this writeup is anonymized.

--

After watching the provided tape and taking shorthand notes of individual overlap occurrences and their effectiveness, I strived to recreate the database purely from tracking data. First, I needed to be able to extract the fullback overlap. I started by looking at Fullback1 (coincidentally, one of my favorite MLS players of the past couple years) and his first half runs. To identify the runs, I took the relatively low threshold of 3.8 m/s as the basis for FB1 being on the move, then using a rowwise difference function, I identified consecutive frames where FB1 was running. After creating a new table with start points and endpoints of runs, I extracted the length of each run then filtered for forward runs of length greater than 10 meters.

The hard part lay in identifying an overlapping run specifically. I settled on iterating through all frames of each of FB1’s forward runs and checking to see if FB1 was ever in line with the ball while staying towards the touchline. After tediously tweaking this series of conditionals, I eventually matched my python output to what I saw in the tape.

--

Now that I had built a model, would it be able to generalize to the other fullbacks? I tried FB2 (the other sided fullback) next and, to my disappointment, the model did not work. I realized then that my current model was extremely overfitted to predict the specific kind of overlapping that FB1 does, so I immediately switched gears to try to come up with a generic function usable for all fullbacks.  This generic function now includes the vertical and horizontal differences in the ball as well as the closest player to the ball (essentially a form of an on-ball metric), and a control to make sure overlapping runs are away from the center channel of the field.

Finally, I needed to quantify the overlapping runs. First, I got all the timestamps of the overlapping runs, the id’s of the player overlapping, and the id’s of the teammate they overlapped. Next, I calculated the distance of the ball to the opposing goal at the start of the run, the minimum distance over the course of the next 20 seconds where the last team to touch the ball was the same as the fullback, and the difference between these as a “ball progression” statistic.

Returned is a timestamped list of each overlapping run, ranked in decreasing order of overall value. For the sake of simplicity, I set the overall value assigned to each overlapping run to be the average of the normalized ball progression statistic and the minimum distance to goal over the next 20 seconds.

Overlapping runs are valued for two things: their ability to allow for ball progression due to disorganizing the defense and creating high valued chances from low valued areas via both crossing the ball and opening passing lanes. With this model, I strove to be able to capture both on-ball and off-ball value of an overlapping run, which is why there are no on-ball measurements here, even though that would most likely translate to a more straightforward model.

Of course, you can’t attribute all the ball progression and chance creation to the fullbacks in each of these cases — if the attacking mid plays a slick pass through the middle while FB1 happens to be overlapping around him, FB1 is almost certainly not the one creating value here. However, the thought behind this is that the space through the defense might not have opened up if it weren’t for that FB1 run.

--

If this tracking data was paired with event data, we could have synced them together and come up with some pretty impactful insights through applying an expected threat or expected goals added model. However, these are still usually lacking at quantifying off-ball movement, which is a channel that overlapping fullback runs tend to influence the game through, as was discussed prior. Also, in addition to a last\_touch\_team\_id, a last\_touch\_player\_id would have been helpful; I had to use this in my code, and having to create it myself, my workaround introduced a bit more error.

Probably the most challenging part of this model was figuring out the parameters that needed fine-tuning, then realizing that I was overfitting the data. Generalizing the code from one fullback to all fullbacks was pretty challenging but also rewarding, as, bear with me, the boolean values to positive and negative one integer typecasting that I did in aiding the ball\_y\_player\_diff parameter was fairly cool to figure out.

--

Overall, this was a very rewarding experience getting, again, to get up close and intimate with tracking data. While it would have been a great exercise to pair it with event data and see what I could have come up with, it was fun only having tracking data and not being able to fall back on seeing what the event data says. Even though this model was fairly noisy, I feel I both learned a lot and reaffirmed my passion for soccer. It’s fun translating data into actionable insight, and even more fun when it’s with a sport you love!