Optimising Football Lineups

Problem Statement

We will select a team, and optimise its line up for each game in a premier league season, and maximise their points total. To determine who plays which game, we will take the 25 man squad for a game, and select 11 players. We will select these 11 players by first requiring that we need 1 goalkeeper, 3-5 defenders, 3 - 5 midfielders, and 1 - 3 attackers. We will then assign each player a rating based on acquired data, and try to select our line up such that we maximise the total rating in each section of the field while making sure the total team average rating is as close as possible to the opposing teams. This way, we will not select our best players for games they are not needed, and they can be saved for games for more difficult opposition. When considering the opposition, for now, we will only use their average ratings from our data. We will also introduce a cost function on the players ratings. If a player plays games consecutively, their rating will decrease by some amount determined by our cost function, to emulate risk of injury (if a players’ rating decreases by too much for playing, for example, 5 games in a row, then they will not be selected for the next game, which could represent how an overplayed player could get injured). Similarly, a players’ rating will also decrease if they play a game too close to a previous one, in terms of time.

Some players may not have a cost. For example, Virgil Van Dijk plays nearly every game for Liverpool without injury, so we will give him a 0 cost function. In a similar fashion, Marco Reus is highly injury prone, so he could have a much more severe cost function.

Now that we have determined the line up, we will determine who will win the game based on the average rating of both teams, and for now, we can select the winner if they have a higher rating, outside of some epsilon range. If the average rating of our teams’ players and the opposing team are within some epsilon, we can consider that a draw and give our team a point. If our rating is greater by epsilon, we can give us 3 points, else give 0.

This will be the model we initially create, and improve upon.

Assumptions

For now, we will not consider champions league games, and internationals games, and freak injuries. We will not take into account transfer windows, and players/managers getting sacked. We will also not consider in and out of form players, and finally we will assume that a team’s average rating does not change throughout a season.

Data

For our team, we will scrape data from FIFA for all the individual players. Initially for a simpler model, we will just scrape a player’s position and the rating FIFA gives them. So, for instance, for Lionel Messi, we would want to extract that he is a 91-rated Right Winger. If we want to add complexity to our model and make it a bit more realistic, we can also figure out that Messi can play as an Attacking Midfielder, with a slight penalty to his rating, since that is not his original position.

In addition, for all opponents that we face, we will just scrape their average overall rating. FIFA, by default, tries to play the optimal lineup for every single team, so our model will be trying to optimise against the best possible opposing team that it can face.

Evaluation:

We can evaluate our model by comparing the individual game lineups, and the outcome predicted to what actually happened in the game. So, for instance, last season, Liverpool drew Tottenham in a game which ended up being very important for the title race, as Liverpool lost the league by a single point. We can see if our model suggests an improved starting lineup which would give a higher chance of victory.

In addition, we can run sanity checks on our results as well. So, we want to make sure that our model is not, for instance, playing a Liverpool youth team player against Manchester City for example, in an attempt to rest some better players for easier fixtures. That game is important, so we want the best possible odds of winning in this case.

Next steps

We will look into adding more constraints for selecting the line ups, perhaps implementing some sort of in and out of form metric, have something more concrete for injuries. We will also try to determine a better way for who wins a match; we are looking into some ML models that could take the data we determine and predict a winner, that we can feed back into our optimization model. We could also try to model an optimised transfer window for our team, or adjust the squad based on the real life transfer window when we get to that stage in the model. Finally, we could also consider the games outside of the league, such as the champions league, international games etc.