Team Members

Ashwin Nair **Archit Dhanani** **Mike Hoang** **Andy Peng**

Abstract

A soccer manager's main job is to select a squad that has the best chance of winning games given a limited budget. It is reasonable to assume that each player's quality and rating determines the chances of their team to win games. An optimization model of a soccer team is developed to assist the manager in selecting the best team using six attributes for each player position. The result is an optimization program that takes inputs about the desired play style and available budget, and returns the optimal player lineup.

Introduction

The objective of this project is to generate an optimal player lineup given a maximum budget and desired playstyle. The significant decisions to be made are how to quantify the quality of each player and how to quantify the playstyle. To accomplish this goal, it was decided that a soccer player's overall skill rating is defined by six individual attributes (more specific details are described in the problem statement). The team's playstyle is determined by two dimensions: aggressive vs. defensive, and wide vs. center. These playstyle variables were quantified and given reasonable constraints and bounds. Finally, the optimization problem is formulated in terms of these playstyle variables and the available budget.

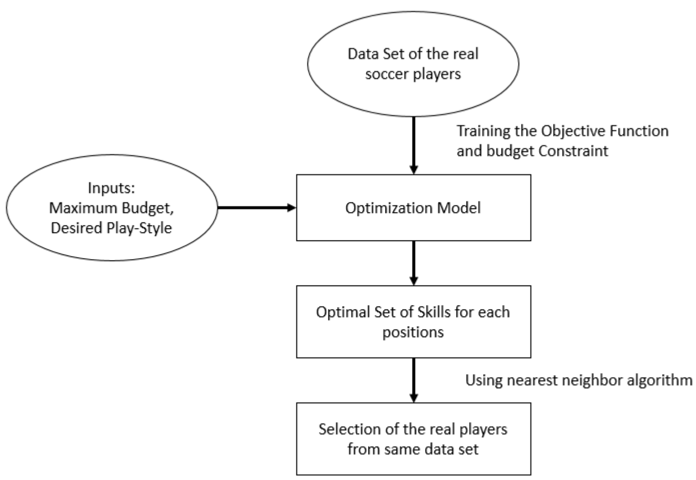

Figure 1: Block Diagram for the model

Problem Statement

As soccer team consists of eleven players, to reduce the dimensionality of the problem we have categorized these eleven players into seven categories of positions: the striker, wingers, center midfielder, wide midfielders, center backs, full backs, and goalkeeper. For each player in our model, we have considered six attributes as the decision variables. Each outfield player is described by the following attributes: pace, dribbling, passing, shooting, defensive skills and physical strength. The goalkeeper, as the only exception, is described by the following attributes: positioning, diving, handling, kicking, reflexes and physical strength. In total, our model has forty-two decision variables (i.e. $x_1, x_2, x_3, ..., x_{42}$).

Objective Function

The proposed optimization problem is to maximize the total skill level of a soccer team with respect to individual attributes of each player.

Each position is a function of the respective six decision variables that describe the player's overall skill. The formation of the players that we have considered is 4-3-3, i.e. one goal-keeper, four defenders, three midfielders and three attacking players. Hence, the objective function is the weight sum of all the position function which can be expressed as:

$$f(x) = w_1 f_1 + w_2 f_2 + w_3 f_3 + w_4 f_4$$

where:

$$f_1 = w_{ST} r_1 + w_W r_2 + w_W r_3$$

$$f_2 = w_{CM} r_4 + w_{WM} r_5 + w_{WM} r_6$$

$$f_3 = w_{CB} r_7 + w_{CB} r_8 + w_{FB} r_9 + w_{FB} r_{10}$$

$$f_4 = r_{11}$$

$$w_1 + w_2 + w_3 + w_4 = 1;\ w_{ST} + 2w_w = 1;\ w_{CM} + 2w_{WM} = 1;\ 2w_{CB} + 2w_{WB} = 1$$

The weights w_1, w_2, w_3, w_4, w_{ST}, w_W, w_{CM}, w_{WM}, w_{CB} and w_{FB} are discrete values assigned based on the desired style of play where:

w_1, w_2, w_3, w_4 define how offensive or defensive needs to be. w_{ST}, w_W, w_{CM}, w_{WM}, w_{CB} and w_{FB} define style of the formation the manager wants i.e., center oriented or more on wings. The style of play is decided by the user.

Definition of the overall rating for each position and the constraints on them.

1. **One Striker (ST)**: $r_1 = \sum(\beta_i x_i)$ for i = 1...6
2. **Two Wingers (W)**: $r_2 = r_3 = \sum(\beta_i x_i)$ for i = 7...12
3. **One Center Midfielder (CM)**: $r_4 = \sum(\beta_i x_i)$ for i = 13...18
4. **Two Wide Midfielders (WM)**: $r_5 = r_6 = \sum(\beta_i x_i)$ for i = 19...24
5. **Two Center Backs (CB)**: $r_7 = r_8 = \sum(\beta_i x_i)$ for i = 25...30
6. **Two Full Backs (FB)**: $r_9 = r_{10} = \sum(\beta_i x_i)$ for i = 31...36
7. **One Goalkeeper (GK)**: $r_{11} = \sum(\beta_i x_i)$ for i = 37...42

Budget Constraint

An important constraint in our model is the budget constraint. To satisfy this constraint in our problem, we needed a model to relate the attributes of a desired player to a market value so that the cost of a player can be taken into consideration in our optimization problem. To obtain this relationship, we looked at real market value data from about 5000 players.

To simplify the model, we created an intermediate value representing the overall performance of a player. This value is a function of the six attributes and the position of a player. We obtain this function through linear regression models for each position that relate a player's six attributes to their "overall rating" on the FIFA Index website.

Then, we used the market value data to determine a relationship between a player's overall rating and their market value in each position. The models that we developed from our data are of the form:

$$c = ae^{dr^5}$$

where:

r = Overall rating of the player

c = Market Value of the player

a and d are the regression coefficients that change with respect to the different positions

Every player in the team is assigned a market value(MV) with function of overall rating defined earlier. Hence, the total expense to build the team cannot be greater than the maximum budget given by the user.

Total expense on the players can be defined as:

$$c = \sum_{i=1}^{11} c_i = \sum_{i=1}^{11} a_i e^{d_i r_i^5}$$

where c_1 = Market Value of Striker (ST)

$c_2 = c_3$ = Market Value of Left and Right Winger (W)

c_4 = Market Value of Center Midfielder (CM)

$c_5 = c_6$ = Market Value of Wide Midfielders (WM)

$c_7 = c_8$ = Market Value of Center Backs (CB)

$c_9 = c_{10}$ = Market Value of Full Backs (FB)

c_{11} = Market Value of Goalkeeper (GK)

a_i and d_i are weights that change according to position

The total expense is constrained by the maximum budget(in dollars) as follows:

$c \leq b$ such that b $\in [9 \times 10^5, 10^9]$

Negative Null Form

$$\min_{\mathbf{x}\in\mathbb{R}^{42}} \quad -f(x) = -(w_1 f_1 + w_2 f_2 + w_3 f_3 + w_4 f_4)$$

where:

$$f_1 = w_{ST} r_1 + w_W r_2 + w_W r_3$$
$$f_2 = w_{CM} r_4 + w_{WM} r_5 + w_{WM} r_6$$
$$f_3 = w_{CB} r_7 + w_{CB} r_8 + w_{FB} r_9 + w_{FB} r_{10}$$
$$f_4 = r_{11}$$

w.r.t.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{42} \end{bmatrix}$$

subject to

$$g_1 = \sum_{i=1}^{11} c_i - b \leq 0,$$
$$g_2 = \mathbf{r}_{\text{LB}} - \mathbf{r}_{\text{i}} \leq 0; \qquad i \in [1, 11]$$
$$g_3 = \mathbf{r}_{\text{i}} - \mathbf{r}_{\text{UB}} \leq 0; \qquad i \in [1, 11]$$
$$g_4 = \mathbf{x}_{\text{LB}} - \mathbf{x} \leq 0; \qquad \mathbf{x} \in \mathbb{R}^{42}$$
$$g_5 = \mathbf{x} - \mathbf{x}_{\text{UB}} \leq 0; \qquad \mathbf{x} \in \mathbb{R}^{42}$$

Symbol	Description	Value
$\mathbf{x}_{\text{LB}}$	Lower bound of Player Attributes	$0 \ \forall \ \text{x}_{\text{LB}_i}$
$\mathbf{x}_{\text{UB}}$	Upper bound of Player Attributes	$100 \ \forall \ \text{x}_{\text{UB}_i}$
b	Maximum allowed budget	User Defined
$\mathbf{r}_{\text{LB}}$	Lower bound of Overall Rating	$45 \ \forall \ \text{r}_{\text{LB}_i}$
$\mathbf{r}_{\text{UB}}$	Upper bound of Overall Rating	$100 \ \forall \ \text{r}_{\text{UB}_i}$

Analysis of Problem Statement

Monotonicity Analysis

All of the decision variables share the same behavior, so for simplicity, they have been combined in one single vector in the monotonicity analysis below:

Function	**x**
f	-
g_1	+
g_2	-
g_3	+
g_4	-
g_5	+

As seen above, at least one of budget constraint or either upper bound limits must be active, depending on the value of the maximum budget. This problem is well-bounded since the objective function is monotonically decreasing with respect to all the decision variables and there exist three monotonically increasing constraints for each decision variable.

In this problem, there are as many active constraints as decision variables, which means that there is zero degrees of freedom. g_3, g_4 and g_5 are natural constraints since all ratings have to be between 0 and 100. g_2 is a practical constraint since it was decided that a player's overall rating should not fall below 45, which was the rating of the worst player gathered from the data of real players. g_1 is a practical constraint as it is user defined according to the given weights for the playstyle.

It was assumed that a player's overall rating depends on his six individual attributes. It was also assumed that the market value only depends on a player's overall rating and position. These assumptions are valid within the range of play rating data obtained from our sources. We define our constraints so that the feasible domain is restricted to only the region where our assumptions are valid.

This problem is a NLP since the budget constraint is non-linear and the domain is continuous. The objective function is linear, which means it is also convex and it does not have multiple local minima (it only has one global minimum).

Optimization Study

To solve the optimization problem we used an interior point algorithm to obtain a solution. This algorithm is implemented in the default settings for the fmincon function in MATLAB. Since we have monotonous functions for our objective and constraints, we felt that a simple algorithm that uses basic KKT conditions to arrive a solution was best. The interior point algorithm for the fmincon function satisfies these requirements.

In formulating our problem, we considered the scaling for the market value because the range for our market values were from 25,000 to 120,000,000 in dollars and our overall ratings varies from 45 to 100. We scaled these values by redefining our function in terms of millions of dollars so that our market values ranged from 0.025 to 120.

As an example we optimized for a team with a budget of 104 million and a play style evenly split between offensive and defensive and between wide and center. The global optimum for the inputs given in the image below is -81.55, so the optimum team rating is 81.55.

The results show the desired attributes for the best team with the given play style and budget. We also have the option of choosing the players with the closest match for those attributes from our database of players. These players are matched by reducing the root mean squared deviation from the optimal attributes.

Figure 2: GUI

Our criteria for convergence were 1e-6 for the first-order optimality measure which is the absolute value of the gradient of the LaGrangian, which should be zero at the minimum, and 1e-6 for the constraint violation tolerance which should be zero to satisfy all constraints. We also set the maximum iteration for our optimization program to be 1000000.

Sensitivity Analysis

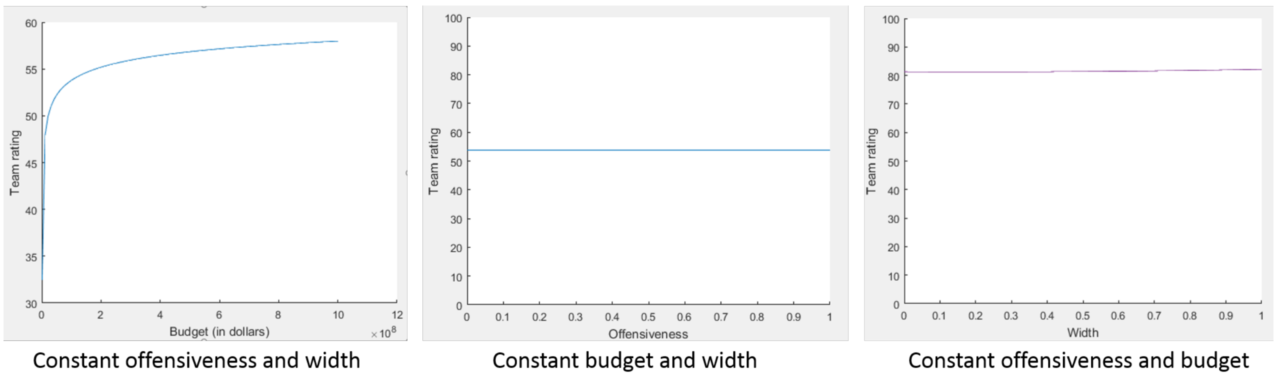

Figure 3: Sensitivity Analysis Plots

We conducted a parametric study on the effect that budget and play style have on the overall rating of the team. We found that an increase in budget with constant play style leads to increases in team rating. There are diminishing while changes to play style with a constant budget do not.

References

- FIFA Index - For player attributes and ratings
- Transfer Market - For current market value of players
- Lagrange Multipliers - Scholarly Article
- Constrained Optimization Using Lagrange Multipliers - Scholarly Article
- Futhead - For player attributes and ratings