Optimal Football strategies: An MDP approach

Elliot Orenstein, Felipe Simon

1 Introduction

Football is a team sport mainly played in the U.S. The biggest league is the National Football League, which is made up of 32 teams from all parts of the country. The main objective of the game is to score more points than the adversary. There are 4 ways of scoring points: Touchdown (6 points), Field goal (3 points), Safety (2 points) and Try after touchdown (1 point for an extra point kick, 2 points for a completed two-point conversion). A touchdown is scored when the ball is carried or passed to the opposition end zone. For a detailed description of the rules of scoring we refer to [4].

A team has 4 opportunities (called downs) to either score a touchdown, try a field goal, punt or gain at least 10 yards. If a team manages to gain 10 yards then they get a first down (4 more opportunities). Otherwise, the ball is turned over to the other team (the way in which it is turned over depends on the previous action).

Recently there has been an increasing interest in the use of statistical methods to support decision making in competitive sports. In particular, Machine Learning and Artificial Intelligence has seen many applications in these types of settings. For a survey of some of these applications we refer to [3].

In this project we study how to model a game of football using an MDP in order to calculate the optimal strategy that should be taken by the offensive team to obtain a first down. The closest previous work to the one presented here is [1]. There, the authors also use a state-based model with stochastic transitions to capture the random nature of football. The main differences from their work and this one are: i) their objective function differs from ours in that a success for their model is a touchdown whereas in our case obtaining a first down is sufficient, and ii) they include a game-theoretic dimension to their model where the transition probabilities depend not only on the action of the team but also on the defensive arrangement of the opposition.

[1] is not the only work where game theory is used to model football strategies. [2] presents a framework to incorporate information from the match to improve decision making in real time. Their work is more focused on data created while the game is being played. In our model, historic data is more important and information created during the game is only informative to find the current state of the world.

1.1 Football plays as an MDP

Football lends itself to be modeled as a Markovian decision process better than other sports because yards and downs left are a natural way of modeling the states of the game. It is also not far-fetched to assume that based on these states the Markovian property will hold. That is to say, the next yardage and downs left will only depend on the current play and the history of how a team got to its current state has no effect on the future of the game.

2 Model

An MDP is a Markov chain in which an agent takes actions and thereby receives rewards. In this context, yards remaining until the first down is reached $S_y = \{0, 1, \ldots, 20\}$ and the number of downs remaining $S_d = \{0, 1, 2, 3, 4\}$, where $s_d = 0$ is an auxiliary state, comprise our state space $S = S_y \times S_d$. From the current state, the agent takes an action $x \in X$, which for our problem is the play type: run, pass, or punt. The action chosen from the current state will provide an immediate reward to the agent $C(s, x)$ and will transition the agent to the next state according to the transition matrix P.

The goal of our agent is to reach the first down (i.e., to have zero yards remaining) before using all of his downs. There are three mutually exclusive outcomes for the agent in pursuit of this goal: i) the agent reaches the first down within the four downs allotted, ii) the agent retains possession but does not reach the first down within the four downs, or iii) the agent forfeits possession (i.e., punts) within the four possessions. The final outcomes are expressed by adding auxiliary states of zero downs remaining $(s_y, 0), \forall s_y \in S_y$ and punting $(-1, s_d), \forall s_d \in S_d$ to our state space. The rewards associated with each state and action are

$$\forall x \in X : C((s_y, s_d), x) = \begin{cases} 0, & s_d > 0 \\ 0, & s_y = -1 \\ -1, & s_d = 0,\ s_y > 0 \\ 1, & s_d = 0,\ s_y = 0 \end{cases}$$

The context of our problem dictates that only the final outcome determines rewards (1, -1, and 0, respectively). All states with positive downs remaining do not generate rewards. Therefore, actions are only related to rewards through their effect on the agent's transitions.

With the state space, action space, rewards, and transition probabilities known, the agent's objective is then to choose the actions that yield the highest reward. Each state then has value $V_t(S_t)$ corresponding to the rewards that can be generated by taking the optimal action at that state. Finding the values of all states is then the reward of taking an action at that state plus the discounted expected values that come from the next state the agent transitions to as a result of his action.

One way of expressing the agent's objective in an MDP is for the agent to optimize the value of each state through his choice of actions. This is known as the Bellman equation [5]:

$$V(S_t) = \max_{x_t \in X}(C(S_t, x_t) + \gamma \sum_{s \in S} \text{Prob}(S_{t+1} = s | S_t, x_t) V(s))$$

where γ ($\gamma = 1$ for our purposes[1]) is the discount factor. From the Bellman equation, we can not only find the values of each state, but also the optimal policy of actions.

[1]This is because of the form of our reward function. Since the agent only receives a reward at the last step, having a $\gamma \neq 1$ would be equivalent to rescaling the rewards in the final down.

2.1 Solving MDP using LP

We can alternatively solve the MDP through the following linear program [5]:

$$\begin{aligned} \min \quad & \sum_{s \in S} \beta_s y_s \\ \text{s.t.} \quad & y_s \geq C(s,x) + \gamma \textstyle\sum_{s' \in S} \text{Prob}(s'|s,x) y_{s'}, \forall s \in S, x \in X \end{aligned}$$

The optimal solution of this LP provides us with the values of each state. We can relate the optimal solution of this LP to the solution of the Bellman equation by looking at the binding constraints. There are three constraints associated with each y_s, one for each action. The value of y_s must be greater than or equal to each of its right-hand sides, which are the expected values of the state given the action taken. The binding constraint is the one that therefore maximizes the value of the state and the action associated with it is the optimal action for that state. In other words, the optimal solution to the LP indicates what the expected values of each state will be when taking the optimal policy and what that optimal policy is. The β_s are positive parameters that can be used to tune the model. In our case, in favour of having a simple model, we chose $\beta_s = 1$. This way no state will be unjustifiably favored.

2.2 Recovering optimal policy

To find the optimal policy, we need only to find the binding constraints, which can be seen through the optimal solution to the dual problem. From complementary slackness, we know that non-zero values of the dual optimal solution correspond to the binding constraints of the primal problem. Zero-valued dual variables at optimality indicate there is slack in the associated primal constraint. For the purposes of our analysis, we are not concerned with actions taken in the auxiliary states.

3 Data

In order to estimate the transition probabilities we use data from NFL games from the calendar year 2021. We have very granular data, that includes for every play made, whether it was a rush or a pass, the number of yards gained (or lost), among other variables. The dataset consists of $34,797$ rows.

3.1 Distribution of passes

Passes are seen as more risky than rushes but with the upside that more yards can be made from them. Figure 1 shows the distribution of yards gained from successful passes. It is interesting that the mode of the distribution is 10, this might be seen as a rational response to the rules of the game. Another important aspect is that 34% of passes are incomplete, that is no yard difference is accomplished from the play.

3.2 Distribution of running plays

Running plays (or rushes) are on distribution less risky but they usually entail less reward. Though in occasion a great play might come out from a running play (as seen in Figure 2) the distribution is highly concentrated between $[-2, 15]$.

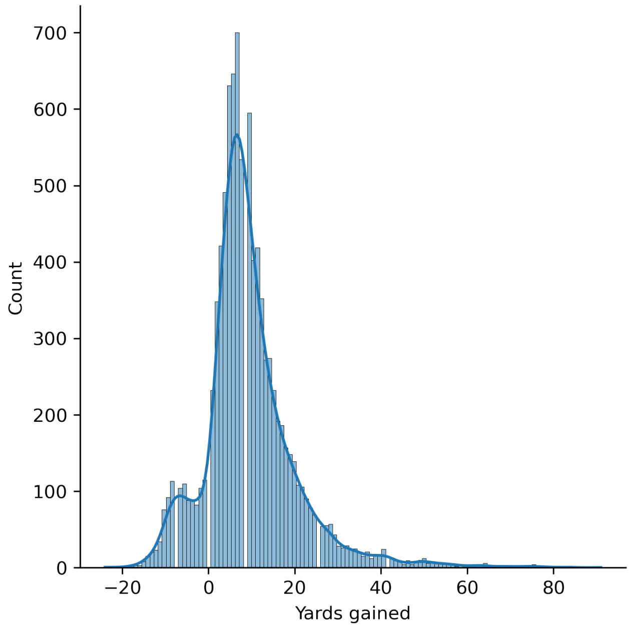

Figure 1: Distribution of yards gained from successful passing attempts

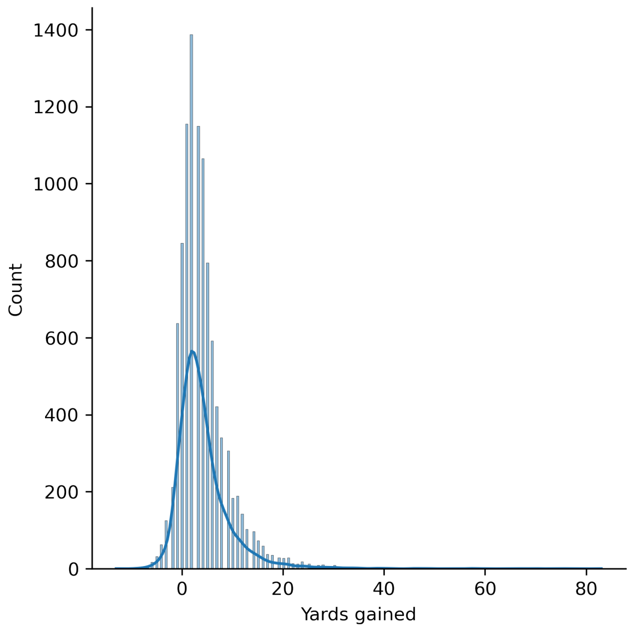

Figure 2: Distribution of yards gained from rushes

3.3 Transition matrices

Based on these empirical distributions we estimate the transition probabilities of each action in the following manner: Let $\hat{F}_x$ be the empirical distribution of action $x \in \{pass, run\}$. Then the probability of gaining k yards given action x is:

$$\begin{aligned} P(k \leq z < k+1|x) &= 1 - \hat{F}_x(k) - (1 - \hat{F}_x(k+1)) \\ &= \hat{F}_x(k+1) - \hat{F}_x(k) \end{aligned}$$

We can use this expression to calculate the transitions from yard i to yard j:

$$P_{ij|x} = \begin{cases} \hat{F}_x(i-j) - \hat{F}_x(i-j-1) & 0 < j < 20 \\ \hat{F}_x(i-20) & j = 20 \\ 1 - \hat{F}_x(i) & j = 0 \end{cases}$$

4 Results

Solving our MDP provides us with the optimal policy at each state and the value of each state under this optimal policy. The values of each state are seen in Figure 3:

	1	2	3	4	5
1	1.0000	1.0000	1.0000	1.0000	1.0000
2	-1.0000	0.6242	0.8906	0.9623	0.9856
3	-1.0000	0.3983	0.8108	0.9333	0.9742
4	-1.0000	0.1319	0.7067	0.8918	0.9571
5	-1.0000	0.0318	0.5948	0.8417	0.9348
6	-1.0000	1.9676e-10	0.5178	0.7847	0.9069
7	-1.0000	1.9737e-10	0.4830	0.7315	0.8752
8	-1.0000	2.0658e-10	0.4438	0.6983	0.8404
9	-1.0000	2.0919e-10	0.3998	0.6603	0.8087
10	-1.0000	2.1665e-10	0.3576	0.6213	0.7808
11	-1.0000	2.2257e-10	0.3152	0.5800	0.7496
12	-1.0000	2.2882e-10	0.2794	0.5412	0.7180
13	-1.0000	2.3629e-10	0.2460	0.5029	0.6852
14	-1.0000	2.4425e-10	0.2178	0.4671	0.6529
15	-1.0000	2.5394e-10	0.1934	0.4334	0.6209
16	-1.0000	2.6398e-10	0.1720	0.4017	0.5898
17	-1.0000	2.7453e-10	0.1533	0.3715	0.5590
18	-1.0000	2.8436e-10	0.1371	0.3435	0.5291
19	-1.0000	2.9326e-10	0.1225	0.3169	0.4994
20	-1.0000	3.0090e-10	0.1099	0.2920	0.4703
21	-1.0000	4.9014e-10	0.0985	0.2359	0.3525
22	3.4218e-11	3.6745e-10	8.1598e-10	1.4921e-09	2.8555e-09

Figure 3: Values of each state under optimal policy

The (i, j) element of the table corresponds to the $(y_s + 1, y_d + 1)$ state. We see that the expected values of the zero-down state are exactly the rewards achieved at each of these states.

Two key insights arise from these values. Firstly, with one down remaining, we see that the value is non-zero from one to four yards remaining and zero for all yards remaining greater. As we will see from the optimal policy, this indicates that the agent can gain more rewards in expectation by using all four downs rather than forfeiting possession. While our model does not consider time, score, field position, in-game history, the defensive scheme, and other important factors that should be considered when making a play call decision, this finding is in direct contrast to conventional strategies and suggests that teams should be less conservative on fourth down.

Clearly, for a fixed number of yards remaining, states with more downs have higher values. Likewise, for a fixed number of downs remaining, states with fewer yards to go have higher values. Having all values allows us to compare states across all dimensions to one another. For example, three yards and two downs remaining and seven yards with three downs remaining each have a value of approximately 0.7. These comparisons across states could be especially important in decision-making when the defense commits a penalty. Often in these situations, coaches are offered the chance to accept or decline penalties that occurred during the previous play. Accepting would allow the previous play's results to hold and the state to transition as it happened on the field. Declining would instead provide the team to take another action from the previous state.

We derive the optimal policy through the dual optimal solution. In Figures 4, 5, and 6, we've removed dual variables associated with the auxiliary states (i.e., punting and zero-downed states) as well as the zero-yard states. The latter was removed because the agent's goal has been achieved and no further actions are needed.

Each state has three dual variables, one for each action. The action corresponding to the non-zero value of the state's dual variable is the optimal action for that state. For example, the first row, first column state (one yard and one down remaining) has a non-zero value for the dual variable associated with the run constraint, so running is the optimal action in order to maximize the agent's value. Following this first column (one down remaining), we see that the optimal policy is to run from one to three yards remaining, pass with four yards remaining, and punt for all distances further out. We saw previously that the agent should "go for it" on fourth down with four yards or fewer, but now we have the optimal actions. We also see that the agent never punts before fourth down, which is logical given that there is no loss of rewards until the zero-down state. If we were to include the possibility of turnovers, or plays in which the offense forfeits possession through error (e.g., interception, fumble), this could change the dynamics of punting strategies. With the optimal policy, the agent now has a play-calling strategy for each state.

{[3.0034]}	{[2.7121]}	{[1.9851]}	{[1.0000]}
{[2.8995]}	{[2.6626]}	{[2.0239]}	{[1.0000]}
{[2.6266]}	{[2.5080]}	{[1.9999]}	{[1.0000]}
{[3.2297e-09]}	{[2.2843]}	{[1.9296]}	{[1.0000]}
{[1.3206e-09]}	{[1.0693e-08]}	{[1.8404]}	{[1.0000]}
{[8.7120e-10]}	{[3.8919e-09]}	{[6.9876e-08]}	{[1.0000]}
{[6.9506e-10]}	{[2.7318e-09]}	{[1.2992e-08]}	{[1.0000]}
{[6.0923e-10]}	{[2.4186e-09]}	{[8.4357e-09]}	{[9.2473e-08]}
{[5.5319e-10]}	{[2.3355e-09]}	{[6.6969e-09]}	{[3.2621e-08]}
{[5.1150e-10]}	{[2.4078e-09]}	{[6.0123e-09]}	{[2.2525e-08]}
{[4.8982e-10]}	{[2.5322e-09]}	{[5.5119e-09]}	{[1.7182e-08]}
{[4.6850e-10]}	{[2.7073e-09]}	{[5.1884e-09]}	{[1.4380e-08]}
{[4.5342e-10]}	{[2.9061e-09]}	{[4.9095e-09]}	{[1.2238e-08]}
{[4.4396e-10]}	{[3.1306e-09]}	{[4.7092e-09]}	{[1.0657e-08]}
{[4.3598e-10]}	{[3.3790e-09]}	{[4.5681e-09]}	{[9.3629e-09]}
{[4.3180e-10]}	{[3.6666e-09]}	{[4.4899e-09]}	{[8.3967e-09]}
{[4.3241e-10]}	{[3.9802e-09]}	{[4.4273e-09]}	{[7.3954e-09]}
{[4.3931e-10]}	{[4.3737e-09]}	{[4.4030e-09]}	{[6.5197e-09]}
{[4.6789e-10]}	{[4.8081e-09]}	{[4.2342e-09]}	{[5.0830e-09]}
{[5.1244e-10]}	{[5.3032e-09]}	{[6.1470e-09]}	{[4.5223e-08]}

Figure 4: Dual optimal solution associated with run constraints

```
{[8.6740e-10]}   {[4.4234e-09]}   {[9.9980e-09]}   {[2.0447e-08]}
{[1.5100e-09]}   {[4.1301e-09]}   {[9.0757e-09]}   {[1.8321e-08]}
{[9.3286e-09]}   {[4.1116e-09]}   {[8.7855e-09]}   {[1.7044e-08]}
{[    2.3682]}   {[1.0141e-08]}   {[9.2575e-09]}   {[1.6870e-08]}
{[1.0881e-08]}   {[    2.0730]}   {[1.6298e-08]}   {[1.8140e-08]}
{[3.2390e-09]}   {[    2.4743]}   {[    1.7176]}   {[2.3297e-08]}
{[1.8926e-09]}   {[    2.2972]}   {[    1.6037]}   {[6.7294e-08]}
{[1.3028e-09]}   {[    2.3094]}   {[    1.8458]}   {[    1.0000]}
{[1.0524e-09]}   {[    2.2225]}   {[    1.7679]}   {[    1.0000]}
{[8.7415e-10]}   {[    2.1583]}   {[    1.7250]}   {[    1.0000]}
{[7.8835e-10]}   {[    2.0965]}   {[    1.6905]}   {[    1.0000]}
{[7.1623e-10]}   {[    2.0288]}   {[    1.6505]}   {[    1.0000]}
{[6.6697e-10]}   {[    1.9635]}   {[    1.6159]}   {[    1.0000]}
{[6.3408e-10]}   {[    1.8904]}   {[    1.5739]}   {[    1.0000]}
{[6.0373e-10]}   {[    1.8210]}   {[    1.5361]}   {[    1.0000]}
{[5.8048e-10]}   {[    1.7573]}   {[    1.5001]}   {[    1.0000]}
{[5.6337e-10]}   {[    1.7022]}   {[    1.4713]}   {[    1.0000]}
{[5.4893e-10]}   {[    1.6557]}   {[    1.4475]}   {[    1.0000]}
{[5.3854e-10]}   {[    1.6178]}   {[    1.4261]}   {[    1.0000]}
{[9.5373e-10]}   {[    1.0000]}   {[    1.0000]}   {[    1.0000]}
```

Figure 5: Dual optimal solution associated with pass constraints

```
{[6.3162e-10]}   {[4.4248e-10]}   {[4.0897e-10]}   {[3.9830e-10]}
{[9.8886e-10]}   {[4.8592e-10]}   {[4.2167e-10]}   {[4.0289e-10]}
{[2.9793e-09]}   {[5.5720e-10]}   {[4.4119e-10]}   {[4.0998e-10]}
{[1.2710e-08]}   {[6.6137e-10]}   {[4.6726e-10]}   {[4.1967e-10]}
{[    2.8436]}   {[7.5961e-10]}   {[5.0085e-10]}   {[4.3248e-10]}
{[    2.8010]}   {[8.1440e-10]}   {[5.3742e-10]}   {[4.4816e-10]}
{[    2.6560]}   {[8.8628e-10]}   {[5.6256e-10]}   {[4.6702e-10]}
{[    2.5874]}   {[9.8355e-10]}   {[5.9501e-10]}   {[4.8521e-10]}
{[    2.4971]}   {[1.0989e-09]}   {[6.3228e-10]}   {[5.0219e-10]}
{[    2.4163]}   {[1.2462e-09]}   {[6.7710e-10]}   {[5.2257e-10]}
{[    2.3385]}   {[1.4051e-09]}   {[7.2545e-10]}   {[5.4514e-10]}
{[    2.2516]}   {[1.5944e-09]}   {[7.8031e-10]}   {[5.7093e-10]}
{[    2.1666]}   {[1.7994e-09]}   {[8.3970e-10]}   {[5.9893e-10]}
{[    2.0707]}   {[2.0249e-09]}   {[9.0453e-10]}   {[6.2956e-10]}
{[    1.9805]}   {[2.2752e-09]}   {[9.7542e-10]}   {[6.6260e-10]}
{[    1.8937]}   {[2.5510e-09]}   {[1.0538e-09]}   {[6.9895e-10]}
{[    1.8204]}   {[2.8502e-09]}   {[1.1388e-09]}   {[7.3830e-10]}
{[    1.7588]}   {[3.1864e-09]}   {[1.2333e-09]}   {[7.8197e-10]}
{[    1.7105]}   {[3.5514e-09]}   {[1.3375e-09]}   {[8.3037e-10]}
{[    1.0000]}   {[3.9445e-09]}   {[1.6496e-09]}   {[1.1076e-09]}
```

Figure 6: Dual optimal solution associated with punt constraints

5 Conclusion and next steps

A drawback of this decision-making process is that these strategies are pure (or deterministic). If the defense were to solve the same MDP, it would know exactly which play were to occur and scheme accordingly. Next steps of our model would be to incorporate a game-theoretic approach. This would not only change our LP, but also the data as we would need to further divide the distributions of yards gained by play against each defensive strategy.

References

[1] Jess S Boronico and Scott L Newbert. “Play calling strategy in American football: a game-theoretic stochastic dynamic programming approach”. In: *Journal of Sport Management* 13.2 (1999), pp. 103–113.

[2] Jeremy D Jordan, Sharif H Melouk, and Marcus B Perry. “Optimizing football game play calling”. In: *Journal of Quantitative Analysis in Sports* 5.2 (2009).

[3] Ryan Beal, Timothy J Norman, and Sarvapali D Ramchurn. “Artificial intelligence for team sports: a survey”. In: *The Knowledge Engineering Review* 34 (2019).

[4] NFL. *NFL rulebook*. 2021. URL: `https://operations.nfl.com/the-rules/nfl-video-rulebook/scoring-plays/` (visited on 12/10/2021).

[5] Shuzhong Zhang. *Dynamic Optimization and Applications*.