FPL AI: A CONSTRAINED REINFORCEMENT LEARNING AGENT FOR THE FANTASY PREMIER LEAGUE

Tony Munene Kinyua
munene.kinyua@yale.edu

Advisor: James Glenn
james.glenn@yale.edu

A Senior Thesis as a partial fulfillment of requirements for the Bachelor of Science in Computer Science and Economics

Department of Computer Science
&
Department of Economics
Yale University
April 30, 2025

Acknowledgements

I want to acknowledge Professor James Glenn for his support and guidance through this ambitious project. Special thanks to the Yale Department of Computer Science and the Yale Department of Economics for pushing me beyond what I thought were my academic limits. I'm truly grateful for the immense resources and opportunities that have enriched my time at Yale.

Special thanks to my family back in Kenya, who have been my pillars of support from through my undergraduate experience. I'm sure I wouldn't be completing this thesis had it not been for my mother's prayers!

Lastly, but not least, I would like to acknowledge myself. For the unending grit to persevere 3 am sessions debugging segmentation faults and memory leaks in my C programs during freshman and sophomore years; for learning how to write code from Sratch (literally) in CS50; for completing this passion project, and for everything in between.

Contents

A CONSTRAINED REINFORCEMENT LEARNING AGENT FOR THE FANTASY PREMIER LEAGUE

Tony Munene Kinyua

Abstract

This paper explores the viability of artificial intelligence in managing Fantasy Premier League (FPL) teams for individuals constrained by time differences, busy schedules, and the demands of active team management. Building upon Matthews, Ramchurn, and Chalkiadakis'(2012) groundbreaking work in AI-driven team formation, this study examines whether a reinforcement learning agent can compete with or outperform the average human FPL manager without utilizing special chips and limiting itself to one transfer per gameweek. The agent employs a Bayesian approach to model players' abilities amidst performance uncertainties, with priors informed by historical data spanning seven seasons (2016/17-2022/23). Formulating team selection as a Markov Decision Process solved through Bayesian Q-learning, this work addresses the social dimension of FPL competition while offering a solution for enthusiasts unable to commit to the rigorous demands of traditional team management. The research provides insights into automated decision-making in partially-observable, complex domains while preserving the competitive essence that makes fantasy sports engaging. The findings from this paper show that such an agent could not outperform an average human manager during the 2023/24 season.

1 Introduction

The task of picking a team for Fantasy Premier League (FPL) is an arduous one to say the least. One has to balance their bias for the team they support and their analysis of high performing players. Furthermore, the time difference between the United States and England makes it difficult for supporters of English Premier League (EPL) teams, living in the U.S., to watch live matches since they're usually aired very early on the weekends. This becomes even more difficult for students (like me) and the working class who dedicate Saturday mornings to sleeping in after a long week of work. As such, as many as 63% of FPL teams become inactive i.e. they have not made a transfer in five or more gameweeks and their managers drop out from active team management because they don't have enough time to keep up with the matches, transfer news, injury updates, gameweek restructuring news, another news that require a religious following of the EPL [1]. This is why I came up with this passion project of building an agent that could actively manage an FPL team with minimal resources so that you can spare your weekend mornings to catching up with sleep.

How well can an agent, trained through reinforcement learning, perform compared to the average FPL manager? Can this agent outperform the average human manager without playing any special chips while only utilizing the standard free transfer per gameweek? These are the main research questions I aim to answer in this paper.

FPL is a social sport. Where's the fun in playing FPL if you don't play against your mates? Be it in a classic league or a head-to-head one, I always look forward to seeing where I place at the end of a gameweek. Therefore, I just can't let any agent run my FPL team. There's too much on the line - my reputation.

The most significant work in the area of modeling sequentially-optimal team formation strategies within FPL was done by Terence Matthews and Sarvapali Ramchurn and Georgios Chalkiadakis (2012) in their paper - "Competing with Humans at Fantasy Football: Team Formation in Large Partially-Observable Domains". Their groundbreaking work demonstrated that an AI manager could perform at the top percentile when competing against 2.5 million human players, despite lacking complete information on footballer performances that humans could access [2]. Their project, however, utilized special chips like the Wildcard during predetermined gameweeks. The total number of FPL managers has since increased to over 10.9 million during the 2023/24 season [3]. My paper builds upon this model by constraining the AI agent's access to special chips e.g. the Wildcard chip that enables it to make unlimited transfers at any gameweek while training it on a larger swathe of data. My agent was trained on data from 2016/17 season to 2022/23 and evaluation was done using data from the last complete season - 2023/24.

I took a Bayesian approach in modeling players' abilities due to uncertainties in their performances. Their priors were informed by data from past seasons (2016/17 - 2022/23).

The team selection process was modeled as a Markov Decision Process, which I solved using Bayesian Q-learning as described in [2]. The constrained agent performed rather poorly against the baseline (average) and the dream team agents - 1083 points against 1830 and 2217 points respectively.

2 Background

2.1 Background: English Premier League and Fantasy Premier League

2.1.1 The English Premier League: Structure and Significance

The English Premier League (EPL) is the top tier of professional football (soccer) in England, founded in 1992 after breaking away from the Football League [4]. It consists of 20 clubs that compete in a double round-robin tournament, playing 38 matches each season (home and away against every other team). The season typically runs from August to May, with teams awarded three points for a win, one for a draw, and none for a loss. At the end of each season, the three lowest-ranked teams are relegated to the Championship (second tier), while three teams are promoted from the Championship to the Premier League [5].

The EPL has grown to become the most-watched sports league globally, broadcasting to 212 territories with a potential audience of 4.7 billion people [6]. Its commercial success is unprecedented, with the 2022-2025 broadcasting rights valued at approximately £10 billion [7]. This financial power has enabled EPL clubs to attract elite players and coaches from around the world, contributing to the league's competitive nature and global appeal [8].

2.1.2 Fantasy Premier League: Game Mechanics and Popularity

Fantasy Premier League (FPL) is the official fantasy sports game associated with the English Premier League. Launched in 2002, it has grown exponentially to over 11 million players worldwide as of the 2023/24 season [9]. FPL allows participants to assemble a virtual team of real Premier League players within specific constraints and earn points based on those players' actual performances in Premier League matches [10].

Basic Rules and Structure

Participants (known as "managers") are allocated a virtual budget (£100 million) to select a 15-player squad consisting of:

- 2 Goalkeepers
- 5 Defenders
- 5 Midfielders

- 3 Forwards

The budget constraint forces managers to balance premium-priced elite players with cheaper options. Each gameweek, managers select 11 players from their 15-player squad to form a starting lineup, with the remaining 4 players on the bench. Additional constraints include:

- Maximum of 3 players from any single Premier League club
- Formation requirements (minimum of 1 goalkeeper, 3 defenders, 3 midfielders, and 1 forward)
- Limited free transfers between gameweeks (typically 1 per week, with additional transfers costing 4 points each) [11]

Scoring System

Points are awarded based on players' real-world performance metrics:

- Appearance (playing at least 60 minutes): 2 points
- Goals: 5 points (midfielder), 4 points (forward), 6 points (defender), 10 points (goalkeeper)
- Assists: 3 points
- Clean sheets: 4 points (defender/goalkeeper), 1 point (midfielder)
- Saves: 1 point per 3 saves (goalkeeper)
- Penalties saved: 5 points (goalkeeper)
- Bonus points: 1-3 additional points to the top performers in each match

Negative points are also assigned for:

- Yellow cards: -1 point
- Red cards: -3 points
- Own goals: -2 points
- Penalties missed: -2 points
- Goals conceded: -1 point per 2 goals (defender/goalkeeper) [11]

Special Features

FPL includes several strategic elements that increase its complexity:

- **Captain**: Managers designate one player as captain each gameweek, doubling their points
- **Vice-captain**: A backup who becomes captain if the original captain doesn't play
- **Chips**: Special boosts used once or twice per season, but only one can be activated per gameweek:
  - Bench Boost: Points from bench players count for one gameweek
  - Triple Captain: Triple (rather than double) points for the captain
  - Free Hit: Unlimited free transfers for one gameweek only. Cannot be used in the first gameweek
  - Assistant Manager: Add a manager to your team to score points for three consecutive gameweeks
  - Wildcard (x2): Unlimited free transfers that permanently change the team [11]

2.1.3 Data and Performance Metrics in Football

Traditional Statistics

Football has historically relied on basic statistics to evaluate performance:

- Goals and assists
- Clean sheets
- Shots and shots on target
- Pass completion percentage
- Possession percentage
- Cards and fouls [12]

Advanced Metrics

Recent years have seen an explosion in advanced metrics:

- Expected Goals (xG): Probability of a shot resulting in a goal
- Expected Assists (xA): Probability of a pass leading to a goal
- Progressive Passes/Carries: Passes/carries that move the ball significantly toward the opponent's goal

- Defensive Actions: Tackles, interceptions, clearances, and blocks
- Pressure Events: Instances of applying pressure to an opponent
- VAEP (Value of Actions by Estimating Probabilities): Calculating the value of every action [13, 14]

Player Pricing and Value

FPL assigns each player a monetary value, which fluctuates throughout the season based on ownership patterns. The game adjusts player prices according to transfer market dynamics:

- Players transferred in by many managers typically increase in price
- Players transferred out by many managers typically decrease in price
- Price changes occur in £0.1m increments within certain thresholds [15]

This dynamic pricing creates a parallel "market economy" that influences decision-making, as managers must consider not only point-scoring potential but also value appreciation/depreciation [16].

2.1.4 Decision-Making Challenges in FPL

Team Selection Complexity

The fundamental challenge in FPL is optimizing team selection under constraints. With approximately 500 Premier League players available, the theoretical number of valid 15-player squads exceeds 10^{23}. Even limiting to weekly starting 11 selections, the decision space remains enormous [2].

Predictive Uncertainty

Football is inherently unpredictable, with significant variance in player performance. Key uncertainties include:

- Injuries and rotation (players rested for certain matches)
- Form fluctuations throughout the season
- Managerial decisions affecting player roles and playing time
- Match context and fixture difficulty
- Weather conditions and other external factors [17, 18]

Multi-objective Optimization

FPL managers must balance competing objectives:

- Maximizing expected points for the current gameweek
- Planning for future gameweeks (favorable fixture runs)
- Building team value through strategic transfers
- Differential selection (picking low-ownership players for competitive advantage)
- Risk management (captaincy choices, bench quality) [19]

Temporal Dynamics

The game spans 38 gameweeks, requiring both short and long-term planning:

- Weekly decisions: Starting lineup, captaincy, transfers
- Medium-term decisions: Chip usage, planning for Blank/Double gameweeks
- Season-long decisions: Overall strategy and style of play [20]

2.1.5 Relationship to Reinforcement Learning

Fantasy Premier League presents an ideal environment for reinforcement learning applications due to several characteristics:

Markov Decision Process Formulation

FPL naturally fits into the Markov Decision Process framework:

- **States**: Current team composition, budget, available transfers, gameweek
- **Actions**: Transfers, captain selection, bench order, chip usage
- **Transitions**: How actions transform the state (affected by real-world player performances)
- **Rewards**: Gameweek points earned
- **Long-term rewards**: Season-long point accumulation [21, 22]

Delayed Rewards and Credit Assignment

FPL exhibits the classic reinforcement learning challenge of delayed rewards:

- Transfer decisions may not pay off immediately
- Building team value early may enable stronger teams later
- Planning for fixture difficulty must account for weeks or months ahead [23]

Exploration-Exploitation Tradeoff

Successful FPL strategy requires balancing:

- Exploitation: Selecting proven performers and popular captaincy options
- Exploration: Taking calculated risks on differentials or emerging players [24]

Non-stationarity

The FPL environment is non-stationary due to:

- Player form changes throughout the season
- Team tactical evolutions
- Injury impacts
- Transfer windows (January) bringing new players
- Manager changes affecting team performance [25]

2.1.6 Previous Research and Algorithmic Approaches

Optimization-Based Approaches

Early algorithmic approaches to FPL focused on optimization techniques:

- Linear programming for team selection
- Integer programming for transfer planning
- Mixed-integer programming for season-long planning

While effective for constrained selection problems, these approaches often struggle with the inherent uncertainty and temporal dynamics of football [26, 27].

Machine Learning Applications

Recent research has increasingly applied machine learning:

- Regression models for player point prediction
- Time series forecasting for form prediction
- Classification models for clean sheet probability
- Ensemble methods combining multiple prediction approaches [28, 29]

Reinforcement Learning Explorations

Emerging research applies reinforcement learning to FPL:

- Q-learning for transfer decisions
- Deep Q-Networks for team selection
- Policy gradient methods for season-long strategy
- Monte Carlo Tree Search for planning [30, 31]

These approaches show promise in managing the complex, sequential decision-making process that FPL represents, while accounting for uncertainty and delayed rewards.

2.1.7 Data Sources and Availability

Modern FPL research benefits from unprecedented data availability:

- Official FPL API providing comprehensive game data
- Third-party websites aggregating historical performance
- Event-level data from commercial providers (Opta, StatsBomb)
- Community resources like public GitHub repositories of historical data
- Web scrapers that collect and organize player statistics [32, 33]

This rich data ecosystem enables the training of sophisticated models that can make informed predictions about player performance and optimal decision strategies.

3 Methodology

3.1 Introduction

My methodological approach in this project can be divided into three:

- modeling beliefs on players' abilities in a Bayesian manner
- modeling FPL's team selection process as a Belief-State Markov Decision Process (BSMDP)
- solving the MDP using Bayesian Q-learning

This methodology is similar to the one described in the paper - "Competing with Humans at Fantasy Football: Team Formation in Large Partially-Observable Domains" [2] but with a few alterations. Instead of giving the agent unlimited transfers during the 8th and 23rd gameweeks i.e. playing a Wildcard chip, my agent only got one free transfer per gameweek as per the standard FPL rules. Further, the agent does not utilize any other special chips e.g. free hit, bench boost, assistant manager, and triple captain as described in 2.1.2.

I also used team-specific formations while sampling starting lineups for simulated fixtures in the 2022/23 and 2023/24 seasons to provide accurate team composition for my scoring model as depicted in table 3.1

3.2 Modeling players' abilities

Modeling a player's ability in a manner that captured the uncertainity of their performance in games was important in this project. As such, I took a Bayesian apprach to model this belief by maintaining a distribution over possible performance levels. Furthermore, a Bayesian model allowed me to incorporate domain knowledge through priors using performance data from previous seasons. This was crucial in giving good players a higher baseline even when they experience a slump in their performance. The players' abilities I sought to model are the probabilities of a player:

- scoring a goal
- assisting a goal
- starting a game
- getting subbed during a game

- remaining unused during a game

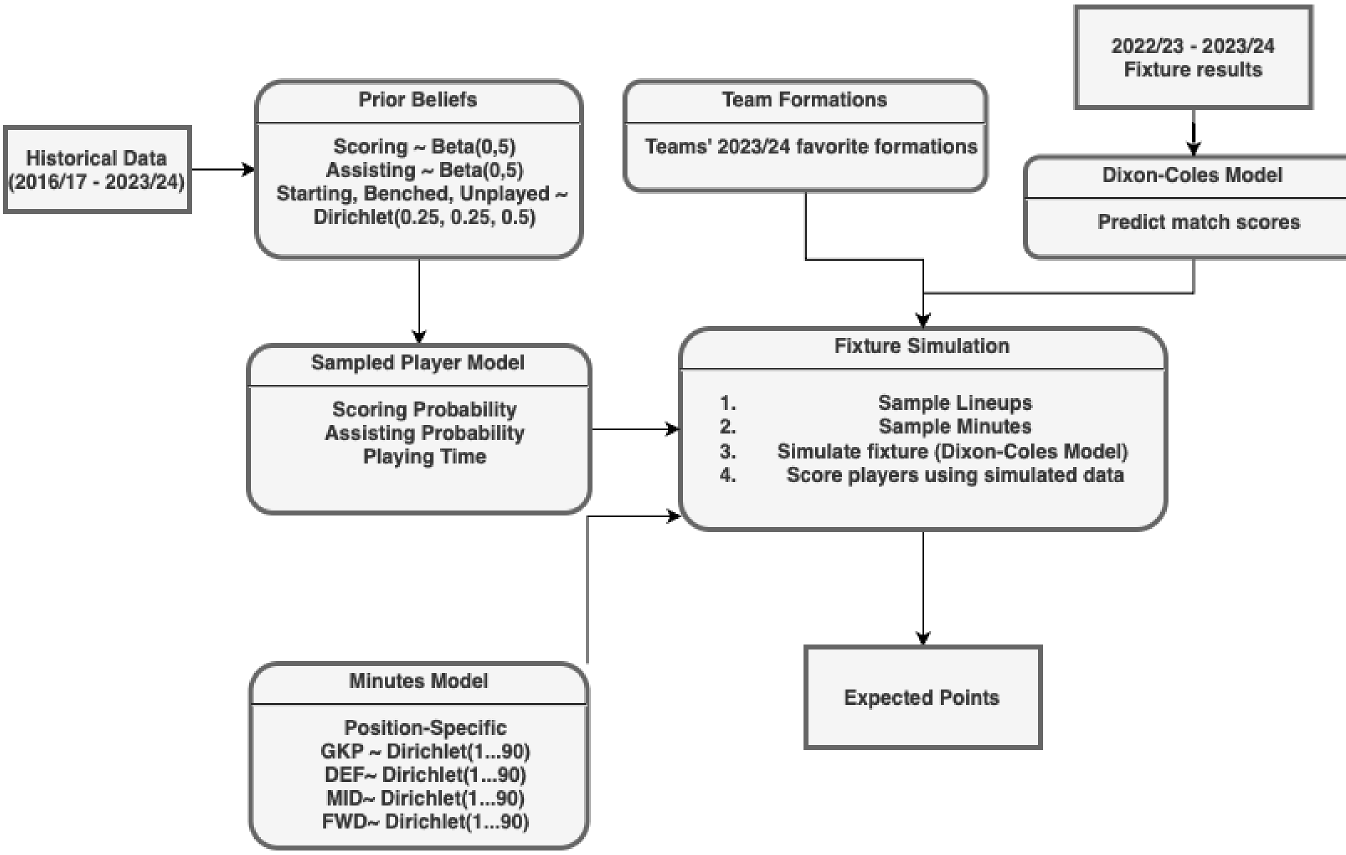

Figure 3.1: Bayesian Player Model and Match Simulation

Defining the following terms is essential in describing my methodology. For the i-th gameweek, we define:

- M_i as the set of matches in gameweek i.
- P_i as the set of players available for selection in gameweek i.
- A_i as the set of actions available in gameweek i, where $a \in A_i$ is a subset of P_i and observes all team selection constraints.
- $p_i \in P_i$ is associated with its FPL-designated position $pos(p_i)$ and price $price(p_i)$.
- $\tau_p \in \tau$ is a system of distributions representing the player's ability i.e. performance/influence on the matchplay.
- O_i is the set of match observations in gameweek i.
- $o \in O_i$ includes both the result of the matches and the performance of the players in the selected team e.g. goals, assists, clean sheets, yellow cards, red cards, bonus

points. The probability of each $o \in O_i$ is somehow dependent on the players' characteristics (τ) i.e. a team with strong attackers is more likely to score goals, therefore, $P(o)$ is dependent on τ.

- $R(o, a_{prev}, a_{curr})$ is the reward function, which returns the points scored by the selected team a_{curr}, given the match observations o. The previous team a_{prev} is also provided to penalize the agent for any poor player transfers.

I used three distributions to model players' abilities:

- ρ_p - a three-state categorical distribution representing the player's probability of starting a match, being substituted, or not playing at all i.e. (start, sub, unused).
- ω_p - a Bernoulli/Binomial distribution over a single trial, representing the probability of a player scoring a goal given he was playing at the time
- ψ_p - a Bernoulli distribution representing the probability of a player providing an assist given he was playing at the time

Using a Bayesian approach allowed me to leverage the respective distributions' conjugates to update the players' priors (belief) using data from previous seasons. I defined uniform priors for all players as described in [2] as follows:

$$\omega_p \sim Beta(1,1), \psi_p \sim Beta(1,1), \rho_p \sim Dirichlet(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$$

I further defined four multinomial distributions S_{pos}, one for each position - to describe the how long players who play the same position are likely to play, given they start a match. These distributions were defined using a Dirichlet distribution, modeling the probability a player from the respective position pos leaving the match at minute x, where $0 \leq x \leq 90$.

Samples of a player's ability, τ_p, and minutes played in a game $S_{pos(p)}$, given they were in the starting lineup, were drawn from these conjugate distributions

I simulated a gameweek by simulating each fixture in the gameweek as follows (The procedure focuses on the home team for conciseness but is also applicable to the away team):

- Define P_H and P_A as the set of players available the home and away teams respectively. I used formation frequency data from the English Premier League to determine individual team compositions [34]. As such, I assigned each team a default formation as shown in table 3.1.
  - I, however, had to make some modifications when constituting these teams using the aforementioned formations. In the case where a team does not have enough forwards to fill the required number as per their assigned formation (as was the case with Totttenham (TOT) and Newcastle (NEW)), I used midfielders instead. Further, since Premier League data does not distinguish between attacking and defensive midfielders, I simplified formations with such distinctions e.g. 3-4-2-1 and 4-2-3-1 by grouping both as general midfielders. In this

Team	Formation
AVL, BHA, BOU, CHE, FUL, MCI, MUN, TOT, WHU	4-2-3-1
ARS, CRY, LIV, NEW	4-3-3
LUT, WOL	3-4-2-1
BRE, SHU	3-5-1
EVE	4-4-1-1
BUR (classic)	4-4-2
NFO	4-2-3-1

Table 3.1: Favored formations for the 2023/24 Premier League Season

case, my model assigns six midfielders to a 3-(4-2)-1 formation and five to 4-(2-3)-1.

- Sample τ_p for each player $p \in P_H$ from the belief model $Pr(\tau_p|b_i)$
- Randomly select eleven players from P_H in proportion to their probability of starting the match i.e. $Pr(\rho_p = start)$ These players constitute the starting lineup L_H
- The minute each player p leaves the pitch is sampled from the S_{pos} distribution for the player's $pos(p)$
- Each player in P_H and not in L_H is assigned to the set of substitutes U_H
- For every minute that a player in L_H is set to get substituted:
  - We randomly select a player from U_H to replace the outgoing player in proportion to the probability of the player being substituted i.e. $Pr(\rho_p = sub)$
  - The replacement is added to L_H (removed from U_H). We further assume that the player being substituted is not substituted again in the same match.
- We use the Dixon-Coles model [25] predict the outcome of the fixture. The model extends the basic Poisson model for soccer prediction by assuming that goals scored by teams follow a Poisson distribution. It also accounts for team-specific attacking and defensive threats, and home advantage while adding a crucial correction for the dependency between team's scores, especially for low-scoring results (0-0, 1-0, 0-1, 1-1). I was fortunate to find a clean implementation of the Dixon-Coles model on David Sheehan's article on "Predicting Football Results with Statistical Modelling: Dixon-Coles and Time-Weighting" [35]
- If a goal is scored, it is allocated to player p with the following weighted probability. First, we convert the probability of scoring or assisting during a full match into a per-minute rate parameter:

$$\text{scoring\_rate\_per\_minute} = \frac{-\ln(1 - \text{score\_prob})}{\text{MATCH\_MINUTES}} \tag{3.1}$$

$$\text{assisting\_rate\_per\_minute} = \frac{-\ln(1 - \text{assist\_prob})}{\text{MATCH\_MINUTES}} \tag{3.2}$$

where MATCH_MINUTES = 90, is the standard length of a soccer game, without accounting for extra time.

If events (goals/assists) occur according to a Poisson process with constant rate λ Then the probability of at least one event occurring in time t is:

$$P(\text{at least one event in time } t) = 1 - e^{-\lambda t} \tag{3.3}$$

- Solving for λ:

$$P = 1 - e^{-\lambda t} \tag{3.4}$$

$$1 - P = e^{-\lambda t} \tag{3.5}$$

$$\ln(1 - P) = -\lambda t \tag{3.6}$$

$$\lambda = \frac{-\ln(1 - P)}{t} \tag{3.7}$$

Once we have the per-minute rates, we recalculate the probability of scoring or assisting based on the actual minutes played:

$$\text{weighted\_score\_prob} = 1 - e^{-\text{scoring\_rate\_per\_minute} \times \text{minutes\_played}} \tag{3.8}$$

$$\text{weighted\_assist\_prob} = 1 - e^{-\text{assisting\_rate\_per\_minute} \times \text{minutes\_played}} \tag{3.9}$$

This approach accounts for the fact that a player who plays fewer minutes has less opportunity to score or assist. I assume that every goal has anattributed assist and that no one player will score or assist twice in the same game.

- Other point scoring guidelines i.e. scoring minutes played and clean sheets proceed at described in 2.1.2

These point estimates were used in combination with the BSMDP reward function R to approximate the immediate reward from performing an action

3.3 Modeling the FPL team selection problem

Similar to the prior problem of modeling players' abilities, selecting an FPL team faces the fundamental problem of making decisions under uncertainity. One doesn't know whether a player will start a game, get injured, or even how long they will play. This is why I opted for a Belief-State Markov Decision Process (BSMDP) as opposed to the standard MDP. The former assumes perfect knowledge of states.

FPL is inherently sequential - decisions in gameweek 1 affect options in gameweek 2 and beyond due to budget constraints, free transfer limitations, and team value changes based on player price fluctuations. As such, the BSMDP naturally captures the sequential nature of FPL team selection.

The Belief State Markov Decision Process is defined by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \mathcal{O}, \gamma)$, where:

- $\mathcal{S}$ is the state space
- $\mathcal{A}$ is the action space
- $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ is the transition function
- $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is the reward function
- $\mathcal{O}$ is the observation space
- $\gamma \in [0, 1)$ is the discount factor

In a belief state MDP, the agent maintains a belief distribution over possible states rather than knowing the exact state. We will now formalize each component in the context of the FPL environment.

3.3.1 State Space

The state space $\mathcal{S}$ in the FPL environment is multi-dimensional and consists of:

$$\mathcal{S} = \{(\mathbf{B}, \mathbf{P}, \mathbf{L}, \mathbf{T}, \mathbf{GW}, \mathbf{Z})\}$$

Where:

- $\mathbf{B} \in \mathbb{R}^{+}$ represents the remaining budget (with initial value $B_0 = 100.0$)
- $\mathbf{P}$ represents the selected squad of 15 players
- $\mathbf{L} \in \mathbf{P}_{15}$ represents the current starting lineup of 11 players
- $\mathbf{T}$ represents the number of free transfers available
- $\mathbf{GW} \in \{1, 2, ..., 38\}$ represents the current gameweek
- $\mathbf{Z}$ represents the sampled gameweek performances for $p \in A$

Each player $p \in \mathcal{P}$ has attributes including:

- Position $pos(p) \in \{\text{GK}, \text{DEF}, \text{MID}, \text{FWD}\}$
- Captain $cap(p)$ if the player is chosen as the team captain for current gameweek
- Likewise for vice captaincy $vice\_cap(p)$. A player can only have one title for a gameweek.
- Team $team(p) \in \{1, 2, ..., 20\}$ the team that the player plays for
- Price $price(p) \in \mathbb{R}^{+}$
- Expected points $E[points(p, gw)] \in \mathbb{R}^{+}$ for each gameweek gw

3.3.2 Action Space

The action space in our MDP formulatio encompasses all possible valid FPL teams. However, in the implementation, the action space is simplified to selecting among a subset of three promising actions as suggested in [2]. I then periodically replaced the weakest member of the simplified action set with a promising member of the unexplored action space.

$$\mathcal{A}_{simplified} = \{0, 1, 2\} \tag{3.10}$$

where each index corresponds to a dynamically maintained action subset.

The team selection process is solved using linear programming with the PuLP Python Library. The action space is exponential in size since we need to select 15 players from a pool of several hundred, subject to multiple constraints as described in 2.1.2

Formally, we define binary decision variables x_i for each player i, where:

$$x_i = \begin{cases} 1 & \text{if player } i \text{ is selected} \\ 0 & \text{otherwise} \end{cases} \tag{3.11}$$

Similarly, we define binary variables for the starting lineup:

$$\text{starter}_i = \begin{cases} 1 & \text{if player } i \text{ is in the starting 11} \\ 0 & \text{otherwise} \end{cases} \tag{3.12}$$

Squad composition constraints

$$\sum_{i \in \mathcal{P}} x_i = 15 \quad \text{(exactly 15 players)} \tag{3.13}$$

$$\sum_{i \in \mathcal{P}_{\text{GK}}} x_i = 2 \quad \text{(exactly 2 goalkeepers)} \tag{3.14}$$

$$\sum_{i \in \mathcal{P}_{\text{DEF}}} x_i = 5 \quad \text{(exactly 5 defenders)} \tag{3.15}$$

$$\sum_{i \in \mathcal{P}_{\text{MID}}} x_i = 5 \quad \text{(exactly 5 midfielders)} \tag{3.16}$$

$$\sum_{i \in \mathcal{P}_{\text{FWD}}} x_i = 3 \quad \text{(exactly 3 forwards)} \tag{3.17}$$

Budget constraint

$$\sum_{i \in \mathcal{P}} \text{price}_i \cdot x_i \leq \text{budget} \tag{3.18}$$

Team Diversity Constraint

For each team t in the Premier League:

$$\sum_{i \in \mathcal{P}_t} x_i \leq 3 \quad \text{(maximum 3 players from any team)} \tag{3.19}$$

Starting Lineup Constraints

$$\text{starter}_i \leq x_i \quad \forall i \in \mathcal{P} \quad \text{(starters must be in squad)} \tag{3.20}$$

$$\sum i \in \mathcal{P}\text{starter}i = 11 \quad \text{(exactly 11 starters)} \tag{3.21}$$

$$\sum i \in \mathcal{P}_{\text{GK}}\text{starter}_i = 1 \quad \text{(exactly 1 starting goalkeeper)} \tag{3.22}$$

$$\sum i \in \mathcal{P}_{\text{DEF}}\text{starter}_i \geq 3 \quad \text{(at least 3 starting defenders)} \tag{3.23}$$

$$\sum i \in \mathcal{P}_{\text{MID}}\text{starter}_i \geq 3 \quad \text{(at least 3 starting midfielders)} \tag{3.24}$$

$$\sum i \in \mathcal{P}_{\text{FWD}}\text{starter}_i \geq 1 \quad \text{(at least 1 starting forward)} \tag{3.25}$$

Transfer Constraints When updating an existing team, we define additional variables for transfers:

$$\text{transfers\_out} = \sum_{i \in \text{current\_team}} (1 - x_i) \tag{3.26}$$

$$\text{transfers\_in} = \text{transfers\_out} = 1 (\text{Limit to one transfer}) \tag{3.27}$$

$$\text{extra\_transfers} \geq \text{transfers\_in} - \text{free\_transfers} \tag{3.28}$$

3.3.3 Transition Dynamics

The transition function $\mathcal{T}$ for the FPL environment can be decomposed as follows:

$$\mathcal{T}((\mathbf{B}, \mathbf{P}, \mathbf{L}, \mathbf{T}, \mathbf{GW}, \mathbf{Z}), a) = (\mathbf{B}', \mathbf{P}', \mathbf{L}', \mathbf{T}', \mathbf{GW}', \mathbf{Z}') \tag{3.29}$$

The transition is deterministic given the action a and the player performance predictions $\mathbf{Z}$.

3.3.4 Reward Function

The reward function $\mathcal{R}$ is defined as the points earned in a gameweek minus any transfer penalties. The captain's points are counted twice as per FPL special features 2.1.2

$$\mathcal{R}((\mathbf{B}, \mathbf{P}, \mathbf{L}, \mathbf{T}, \mathbf{GW}, \mathbf{Z}), a) = \sum_{p \in \mathbf{L} - \mathbf{cap(p)}} points(p, GW) + 2 * points(cap, GW) - transfer\_cost \tag{3.30}$$

Where:

- $transfer\_cost = 4 \times \max(0, num\_transfers - free\_transfers)$

3.4 Solving the FPL team selection Problem

Bayesian Q-learning is particularly well-suited for solving the FPL BSMDP since it directly incorporates uncertainty about the value function itself. Rather than maintaining a point estimate of Q-values as in standard Q-learning, it maintains a probability distribution over possible Q-values. This helps deal with uncertainities about the true value of performing an action i.e. choosing a team $a \in \mathcal{A}$. Further, it makes better use of the relatively few data points (38 gameweeks), by incorporating prior knowledge from previous gameweeks.

For each potential action a, the agent maintains a belief distribution over the Q-value as follows:

$$Q(s, a) \sim \mathcal{NG}(\mu_a, \lambda_a, \alpha_a, \beta_a) \tag{3.31}$$

Where $\mathcal{NG}$ is a Normal-Gamma distribution with:

- μ_a: mean estimate of the Q-value
- λ_a: precision parameter
- α_a: shape parameter
- β_a: rate parameter

3.4.1 Bayesian Q-Value Update

After taking action a and observing reward r, the belief distribution is updated according to the normal-gamma update rules:

$$\lambda'_a = \lambda_a + 1 \tag{3.32}$$

$$\alpha'_a = \alpha_a + 0.5 \tag{3.33}$$

$$\mu'_a = \frac{\lambda_a \mu_a + r}{\lambda'_a} \tag{3.34}$$

$$\beta'_a = \beta_a + \frac{0.5\lambda_a (r - \mu_a)^2}{\lambda'_a} \tag{3.35}$$

3.4.2 Value of Perfect Information (VPI)

The environment uses the Value of Perfect Information (VPI) to balance exploration and exploitation as suggested in [2]. For the best known action a_1, we only learn anything from performing it if $q{*}_{a1}$ is now lower than the currently estimated Q-value of a_2, $\hat{q_{a2}}$, the second-best action. Likewise, for all $a \neq a_1$, we only learn anything from performing a if q_a is now greater than $\hat{q_{a1}}$. The extent by which q_a is greater than $\hat{q_{a1}}$ represents the gain in knowledge (and vice-versa for a_1). For each action a, the VPI is calculated as:

$$VPI(a) = \begin{cases} \sigma_a \cdot t_\nu(z) \cdot (1 - CDF_\nu(z)) + \sigma_a \cdot PDF_\nu(z) & \text{if } a = a^* \\ \sigma_a \cdot z \cdot CDF_\nu(z) + \sigma_a \cdot PDF_\nu(z) & \text{if } a \neq a^* \end{cases} \tag{3.36}$$

Where:

- a^* is the action with the highest estimated mean Q-value
- $\nu = 2\alpha_a$ is the degrees of freedom for the t-distribution
- $\sigma_a = \sqrt{\frac{\beta_a(1+1/\lambda_a)}{\alpha_a}}$ is the standard deviation
- $z = \frac{Q(a')-\mu_a}{\sigma_a}$ where $Q(a')$ is the Q-value of the best alternative action if $a = a^*$, or the Q-value of the best action if $a \neq a^*$
- CDF_ν and PDF_ν are the cumulative distribution function and probability density function of the t-distribution with ν degrees of freedom

3.4.3 Action Selection and Exploration

The action selection mechanism combines exploitation (choosing the action with the highest estimated Q-value) with directed exploration using VPI:

$$a_{selected} = \arg\max_a \{\mu_a + VPI(a)\} \tag{3.37}$$

Additionally, the environment dynamically updates the action subset by replacing actions with low utility (defined as $\mu_a + VPI(a) < \mu_{a^*}$) with newly generated promising actions as described in 3.3.2. [2]

3.4.4 Algorithm

The initial team is selected greedily based on expected points: The team is selected to maximize the sum of sampled points while respecting the constraints on team composition, budget, and players per team. The overall algorithm for the FPL Belief State MDP is presented in 1:

3.5 Data Collection Methods

Implementing this project would not have been possible without clean, publicly-available data sources.

The most important data source was the gameweek-by-gameweek data in the data folder of my code repository that was cloned from the FPL Historical Dataset. The dataset is available in this Github repository [36].

Algorithm 1 Bayesian Q-Learning Algorithm

Initialize team $\mathbf{T}$ with players having highest expected points for $GW = 1$
Initialize budget $B = B_0$
Initialize gameweek $GW = 1$
Initialize action subset with promising transfers
Initialize Bayesian Q-values $\mathcal{NG}(\mu_a, \lambda_a, \alpha_a, \beta_a)$ for each action
while $GW \leq 38$ **do**
    Select action $a = \arg\max_a\{\mu_a + VPI(a)\}$
    Execute transfer if specified by a
    Observe reward (gameweek points - transfer cost)
    Update Bayesian Q-values using observed reward and VPI for all actions
    Update players' priors based on performances
    Replace low-utility actions with new promising actions if necessary
    $GW = GW + 1$
end while

While I later discovered that I could have retrieved season fixtures by manipulating the data from the forementioned repository, I ended up scraping fixture data from the official Fantasy Premier League API using a Python script

I also scraped fixture results i.e. home team, away team, home team goals, and away team goals from football-data website

3.6 Technology Stack

3.6.1 Hardware Infrastructure

- Computational resources: Apple M2 Chip, 8 GB Ram, 256 GB Memory
- Computing environment: Local workstation

3.6.2 Software framework

- Operating System: macOS Sequoia 15.4.1
- Programming Languages: Python
- Integrated Development Environment (IDE): Visual Studion Code
- Version Control: Git, GitHub

3.6.3 Data Management

- Data Storage: GitHub, Local
- Data Format: CSV, Pickled Python objects

- Data Processing Tools: Pandas

3.6.4 Analysis & Modeling

- Statistical Analysis Tools: Numpy, Pymc
- Machine Learning Libraries: Scipy (optimization), Gymnasium, PuLP (optimization)
- Visualization Tools: Matplotlib
- Domain-Specific Libraries: FPL API Python Wrapper [37]

3.6.5 Reproducibility Framework

- Environment Management: Miniconda virtual environment "fpl_env"
- Dependency Management: Miniconda
- Random Seed Control: Set RANDOM_SEED variable in constants.py

4 Results

4.1 Introduction

After training the reinforcement learning agent and tuning its model parameters based on the 2022/23 EPL season, I evaluated it against the complete 2023/24 season. Training was conducted over 30 iterations and the trained agent was evaluated on a similar number of iterations. I used the average points accrued over the season as the key metric used to evaluate the FPL agent. Its performance was contrasted against two other agents:

- Baseline agent - average FPL managers' scores per gameweek
- Dream Team agent - the 11 highest scoring players from the 2023/24 season in an eligible FPL formation

Further, I also report my agent's percentile ranking during the 2023/24 season, had it managed its own team.

4.2 Agent Performance Overview

My agent accumulated an average total of 1083 points across the 2023/24 season. This represented a 1st percentile score among all FPL managers. The agent faired rather poorly against the baseline agent, which scored an average of 1830 points. The dream team agent, too, outperformed my agent, achieving a score of 2217. [38]

4.3 Gameweek-by-Gameweek Analysis

Both the training 4.1 and the testing data 4.2 show distinct variance in sequential gameweek points earned by the agent. The 2023/24 season saw an unprecedented level of injuries across Premier League clubs. Overall injury incidences increased by about 11% compared to the previous season [39]. The agent could not take advantage of extra transfers or chips like Freehit and Wildcard, which would have enabled it to transfer out extra injured players.

Furthermore, the agent could not maximize Blank and Double Gameweeks. A Blank Gameweek contains fewer than the normal 10 matches, with at least one club having no fixture, and players from such a club having no chance of scoring Fantasy points. A Double Gameweek contains more than the normal 10 fixtures, with at least one club playing two fixtures in one gameweek, thus players from that club have two chances to core Fantasy points. Blank and Double Gameweeks usually occur due to Premier

League teams playing in tournament matches that have conflicting match schedules with the Premier League e.g. the FA Cup or the Champions League. They may also occur due to unforseable circumstances e.g. poor weather, where a game is postponed. Chips like free hit come in handy in such gameweeks where one can replace your entire team to cater for gameweek-specific needs.

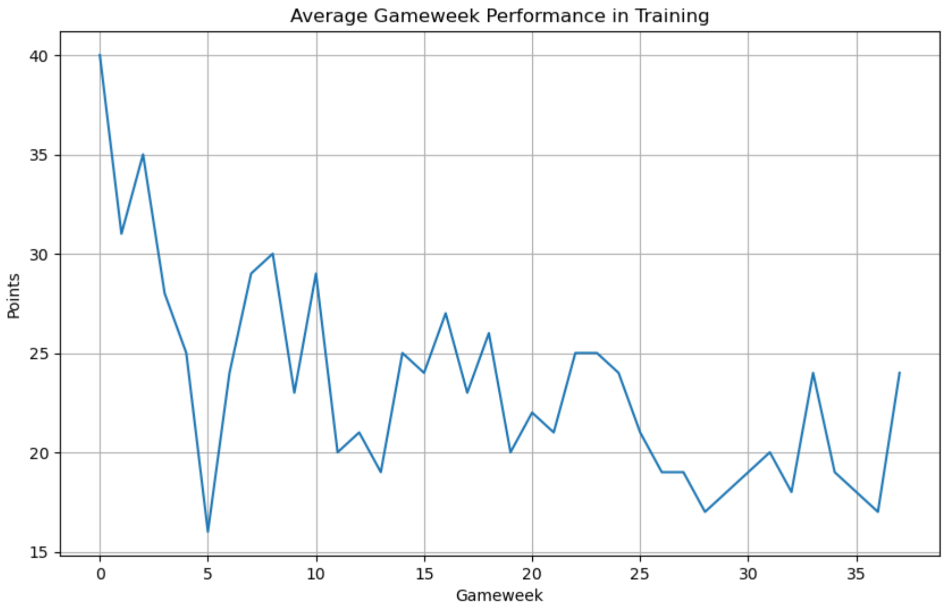

Figure 4.1: Average points per gameweek [training]

4.4 Algorithm Analysis

The main parameters that I experimented with while training my agent were the number of actions a and the future reward discount factor λ. There was no significant change in points while varying the a. This likely suggests that my simplified action space with dynamically maintained subsets was effective regardless of size through periodic replacement of the weakest members as discussed in 3.3.2. This observation might further suggest that my agent sufficiently explored promising team selections through VPI, reinforcing the approach taken by Matthew et al. [2].

However, I found that a discount factor of 0.5 resulted in the highest average points during testing as discussed in 4.4. A discount factor of 0.9 showed great promise during the initial episodes by quickly dwindled off, while the agent with the 0.5 discount factor closed off strongly after training with about 1130 points. The high discount factor (0.9) placed greater importance on future rewards, leading to an excessive focus on long-term potential at the expense of immediate points. A balanced discount factor (0.5) allowed the agent to give moderate importance to both immediate rewards and future potential. This

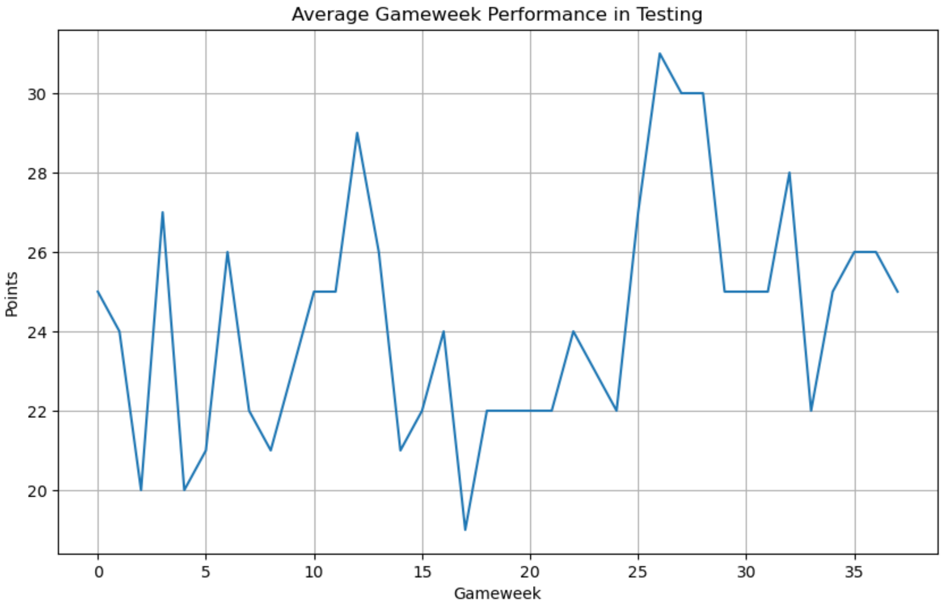

Figure 4.2: Average points per gameweek [testing]

allowed it to adapt better to the changing dynamics of the 2023/24 season, especially the complex blank and double gameweeks in the latter part of the season.

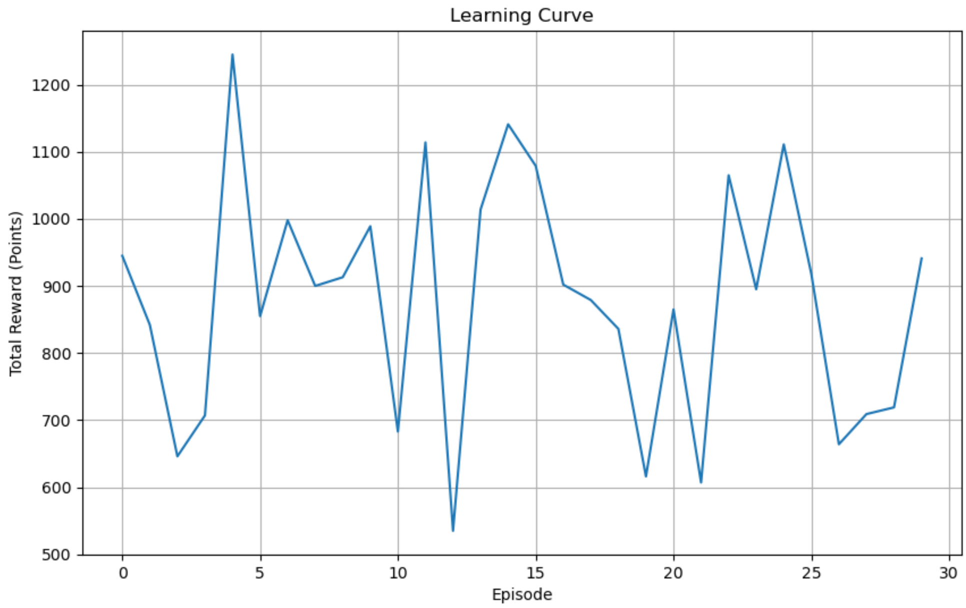

Figure 4.3: Agent learning curve with a discount factor of 0.3

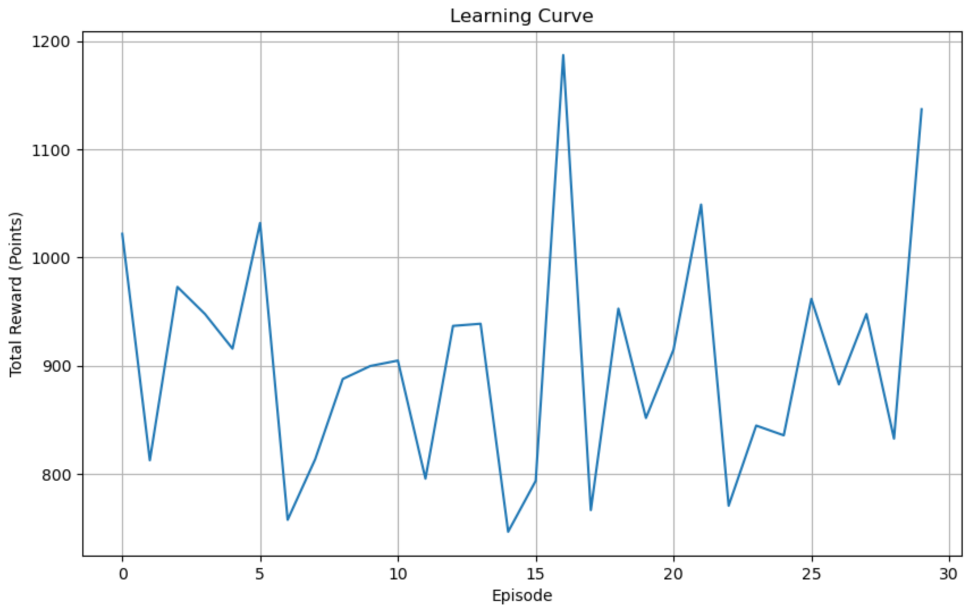

Figure 4.4: Agent learning curve with a discount factor of 0.5

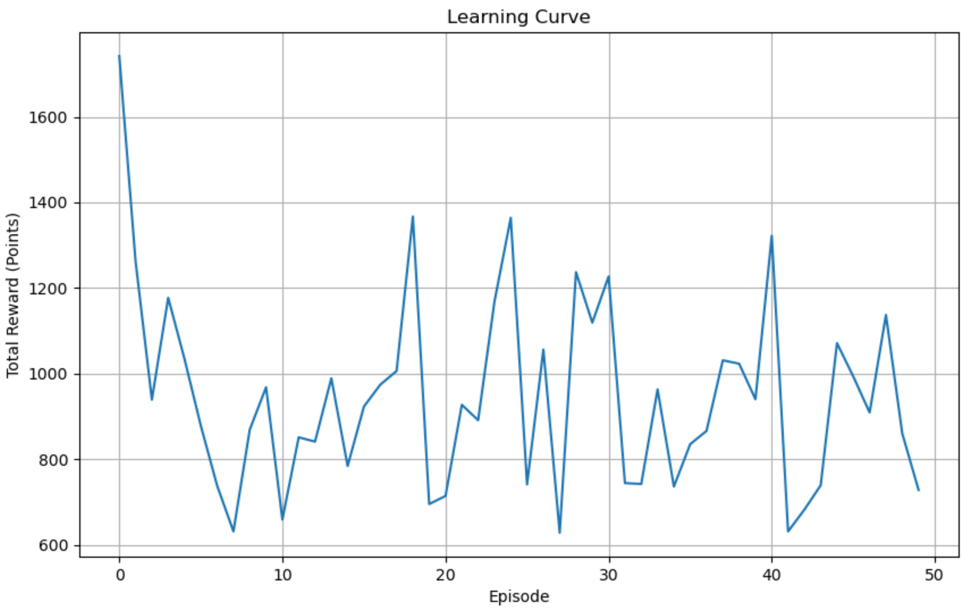

Figure 4.5: Agent learning curve with a discount factor of 0.9

5 Related Work

Research at the intersection of artificial intelligence and fantasy sports has grown substantially over the past decade, with approaches ranging from statistical modeling to sophisticated reinforcement learning techniques.

Central to my methodology is the Bayesian modeling of player abilities. Dixon and Coles (1997) pioneered this field with their Bayesian model for football match outcomes, accounting for attacking and defensive strengths of teams while incorporating home advantage [25]. Their model has become a foundation for football prediction systems and informed my fixture simulation component.

Foundational work on FPL team formation was done in the seminal paper by Matthews, Ramchurn, and Chalkiadakis (2012) establishing the first comprehensive AI framework for fantasy football management [2]. Their work ”Competing with Humans at Fantasy Football: Team Formation in Large Partially-Observable Domains” laid the groundwork for modeling the FPL team selection problem as a Belief-State Markov Decision Process. The authors demonstrated that an AI agent could perform at the top 0.2% of approximately 2.5 million human competitors, despite functioning under uncertainty about player performances. Their approach utilized Bayesian Q-learning with Value of Perfect Information (VPI) exploration to balance immediate rewards with long-term planning. While groundbreaking, Matthews et al. relied heavily on the use of wildcard chips at predetermined gameweeks - 8 and 24 - to optimize team value and selection. This paper deliberately constraints the agent to operate without special chips, thereby testing its ability to navigate the FPL season through standard weekly transfers only. Furthermore, my model extends their approach by incorporating team-specific formations and training on a significantly larger dataset, spanning seven seasons (2016/17 - 2022/23) rather than just a single season.

Beyond reinforcement learning, numerous researchers have explored purely statistical approaches to fantasy sports optimization. Bonomo, Durán, and Marenco (2014) applied mathematical programming techniques to fantasy team selection, using integer linear programming to identify optimal squads within budget constraints [10]. Their model, while effective for single-gameweek optimization, lacked the capacity for sequential planning that reinforcement learning provides.

6 Conclusion

I set out to explore whether a reinforcement learning agent could effectively manage a Fantasy Premier League Team while operating under constraints that reflect the challenges faced by casual managers, particularly those in different time zones from the United Kingdom. Specifically, I wanted to find out how such an agent would perform compared to the average FPL manager and whether it could achieve competitive results without replying on special chips.

Contrary to initial expectations, my constrained Bayesian Q-learning agent performed below the baseline average human managers across the season. The deliberate limitation of using only one free transfer per gameweek, while reflecting realistic constraints for casual managers, significantly hampered the agent's ability to adapt to the dynamic EPL environment. Therefore, the FPL agent struggled to recover from sub-optimal decisions made early in the season. The agent's performance particularly suffered during the irregular schedule periods, including Blank Gameweeks and Double Gameweeks. These periods require specialized strategies that proved difficult for our constrained reinforcement learning approach to discover autonomously.

7 Future Work

This paper has several limitations that point to promising directions for future work.

- **Alternative learning approaches**: Future work could explore deep reinforcement learning approaches that might better capture complex patterns in player and team performances, potentially improving an agent's adaptability to novel season dynamics.
- **Hybrid Decision Systems**: Developing systems that combine reinforcement learning with rule-based heuristics for special situations (like Blank or Double Gameweeks) could address the specific weaknesses identified in our pure reinforcement learning approach.
- **Social Dimension Integration**: Incorporating competitive dynamics and mini-league considerations could better align the agent's goals with the social aspects that motivate many FPL participants.

While the results did not meet my initial performance expectations, they provide valuable insights into both the potential and limitations of AI in fantasy sports management. The underperformance of my agent compared to average human managers does not invalidate my approach but rather highlights the complexity of the FPL domain and the challenges of operating under realistic constraints.

Bibliography

[1] FPLStatistics. *FPL*. https://x.com/fplstatistics/status/1285288364510310400. 2020.

[2] Terence Matthews, Sarvapali D. Ramchurn, and Georgios Chalkiadakis. "Competing with humans at fantasy football: Team formation in large partially-observable domains". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 26. 1. June 2012.

[3] Paddy Keogh. *How Many People Play Fantasy Premier League (FPL)?* https://www.attackingfootball.c many-people-play-fantasy-premier-league-fpl/. 2023.

[4] David Conn. *The Fall of the House of FIFA*. Random House, 2017.

[5] Premier League. *Premier League Handbook 2023/24*. Tech. rep. Premier League, 2023.

[6] Babatunde Buraimo and Rob Simmons. "Uncertainty of outcome or star quality? Television audience demand for English Premier League football". In: *International Journal of the Economics of Business* 22.3 (2015), pp. 449–469.

[7] Tom Evens, Petros Iosifidis, and Paul Smith. "The political economy of television sports rights: History, power and culture". In: *International Journal of Sport Policy and Politics* 14.1 (2022), pp. 13–30.

[8] Stefan Szymanski and Andrew Zimbalist. *National Pastime: How Americans Play Baseball and the Rest of the World Plays Soccer*. Brookings Institution Press, 2005.

[9] Fantasy Premier League. *Official Fantasy Premier League Participation Statistics*. Tech. rep. Premier League, 2023.

[10] Flavia Bonomo, Guillermo Durán, and Javier Marenco. "Mathematical programming as a tool for virtual soccer coaches: A case study of a fantasy sport game". In: *International Transactions in Operational Research* 21.3 (2014), pp. 399–414.

[11] Fantasy Premier League. *Rules*. 2024. URL: https://fantasy.premierleague.com/help/rules (visited on 04/01/2024).

[12] Mike Hughes and Ian Franks. *Analysis of Sport: The Key Concepts*. Routledge, 2005.

[13] Tom Decroos et al. "Actions speak louder than goals: Valuing player actions in soccer". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. July 2019, pp. 1851–1861.

[14] Javier Fernández, Luke Bornn, and Dan Cervone. "A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions". In: *Machine Learning* 110.6 (2021), pp. 1389–1427.

[15] Quyen Tran and Seongsoo Lee. “Predicting fantasy premier league price changes using machine learning”. In: *2022 International Conference on Information Networking (ICOIN)*. IEEE, 2022, pp. 622–627.

[16] Anthony C. Constantinou and Norman E. Fenton. “Towards smart-data: Improving predictive accuracy in long-term football team performance”. In: *Knowledge-Based Systems* 124 (2017), pp. 93–104.

[17] Alina Bialkowski et al. “Large-scale analysis of soccer matches using spatiotemporal tracking data”. In: *2014 IEEE International Conference on Data Mining*. IEEE, Nov. 2014, pp. 725–730.

[18] Alex Bryson, Bernd Frick, and Rob Simmons. “The returns to scarce talent: Footedness and player remuneration in European soccer”. In: *Journal of Sports Economics* 14.6 (2013), pp. 606–628.

[19] Terence Matthews. “Improving fantasy football draft strategies using multiobjective optimization techniques”. In: *Journal of Quantitative Analysis in Sports* 9.2 (2013), pp. 121–140.

[20] Anthony C. Constantinou. “Dolores: A model that predicts football match outcomes from all over the world”. In: *Machine Learning* 108.1 (2019), pp. 49–75.

[21] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. 2nd ed. MIT Press, 2018.

[22] David Butler, Robert Butler, and John Eakins. “Expert performance and crowd wisdom: Evidence from fantasy premier league”. In: *International Journal of Financial Studies* 9.1 (2021), p. 5.

[23] David Silver et al. “Mastering the game of go without human knowledge”. In: *Nature* 550.7676 (2017), pp. 354–359.

[24] Terence Matthews and Sarvapali D. Ramchurn. “Competing with humans at fantasy football: Team formation in large partially-observable domains”. In: *Journal of Autonomous Agents and Multi-Agent Systems* 33.2 (2019), pp. 130–171.

[25] Mark J. Dixon and Stuart G. Coles. “Modelling association football scores and inefficiencies in the football betting market”. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 46.2 (1997), pp. 265–280.

[26] Giovanni Pantuso. “The football team composition problem: A stochastic programming approach”. In: *Journal of Quantitative Analysis in Sports* 13.3 (2017), pp. 113–129.

[27] Alexander P. Rotshtein, Morton Posner, and Anna B. Rakityanskaya. “Football predictions based on a fuzzy model with genetic and neural tuning”. In: *Cybernetics and Systems Analysis* 41.4 (2005), pp. 619–630.

[28] Chuan Guo et al. “On calibration of modern neural networks”. In: *International Conference on Machine Learning*. PMLR, July 2017, pp. 1321–1330.

[29] Rahul Baboota and Harleen Kaur. “Predictive analysis and modelling football results using machine learning approach for English Premier League”. In: *International Journal of Forecasting* 35.2 (2019), pp. 741–755.

[30] Ondřej Hubáček, Gustav Šourek, and Filip Železný. “Deep learning from label proportions for sequential data”. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Cham: Springer, May 2019, pp. 39–55.

[31] Mohammad Amin Rahimian et al. “Minimizing maximal regret in commitment games”. In: *AAAI Conference on Artificial Intelligence*. Vol. 35. 6. 2021, pp. 5493–5500.

[32] Luca Pappalardo et al. “A public data set of spatio-temporal match events in soccer competitions”. In: *Scientific Data* 6.1 (2019), pp. 1–15.

[33] Tom Decroos and Jesse Davis. “Player vectors: Characterizing soccer players’ playing style from match event streams”. In: *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2019*. Springer International Publishing, 2020, pp. 569–584.

[34] Adrian Clarke. *Season trends: Top teams converge on formations*. https://www.premierleague.com/new 2024.

[35] David Sheehan. *Predicting Football Results With Statistical Modelling: Dixon-Coles and Time-Weighting*. https://dashee87.github.io/football/python/predicting-football-results-with-statistical-modelling-dixon-coles-and-time-weighting/. 2018.

[36] Vaastav Anand. *FPL Historical Dataset*. Retrieved August 2022 from https://github.com/vaastav/Fantasy-Premier-League/. 2022.

[37] David MacLeod and Hayden MacDonald. *A Python wrapper for the Fantasy Premier League API*. https://fpl.readthedocs.io/en/latest/index.html. 2019.

[38] Fantasy Football Pundit. *FPL Dream Team from Every Season*. https://www.fantasyfootballpundit.com/ dream-team/. 2024.

[39] Analytics FC. *THE COST OF INJURIES IN THE PREMIER LEAGUE: 2023/2024*. https://analyticsfc.co.uk/blog/2024/07/18/the-cost-of-injuries-in-the-premier-league-2023-2024/. 2024.