FPL AI:
A REINFORCEMENT LEARNING AGENT FOR FANTASY FOOTBALL

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
Dec 12, 2024

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.

Lastly, but not least, I would like to acknowledge myself. For the unending grit to perservere 3 am sessions debugging C program seg faults and memory leaks. For learning how to code from Sratch (literally). For completing this passion project. And for everything in between.

Contents

Title of Your Senior Thesis

Your Name

Abstract

The abstract is a brief description of the paper. Write it in 100-150 words in general. For the thesis, it can be longer. Present your research questions, your contribution to the computer science literature, a concise description of your work, and how you answered your research question. Add a statement of the main results. This part can be revisited when the paper is complete.

Then, you can revisit the title of the paper. If you are revisiting the title of your paper, ask this question: Which one best explains the main point of the paper?

- Microfinance
- Microfinance and Economic Development
- Microfinance as a Strategy for Economic Development: Realizing its Potential for Improving the Standing of Women

Now, examine your title. Does it show the main point of the paper? If you wanted to rewrite it, how would you do it? Review Lecture 5: How to Research (II), Senior Thesis Structure for details in writing your thesis.

1 Introduction

Essentially, the introduction is a summary of the paper. A great introduction is one that makes the reader excited to read the rest of the paper. A good introduction encourages readers to read your work with interest and prepares them to understand it better.

Remember the three-pass reading that we learned in Lecture 2 of the class, CPSC 490 Senior Project? When you write a paper, you can expect most reviewers (and readers) to make only one pass over it. (i.e. 5 min, max 10 min.) Then, decide whether they will read more or put it aside.

A strong introduction does the following in this general order:

- Motivates why the subject of the research is important
- States the research question
- Discusses where the paper fits in existing literature
- Describes the contribution of the paper
- States the research methods
- States the main results

In this template, section titles are provided as suggestions, and you are more than welcome to change them, except for Abstract, Introduction, Background, Related Work, and Conclusion, which are must-have.

Please share the document with your advisor with edit previllage. Comments can be provided *sk: this way, and comments can be removed when addressed.

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 [1]. 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 [2].

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 [3]. Its commercial success is unprecedented, with the 2022-2025 broadcasting rights valued at approximately £10 billion [4]. 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 [5].

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 [6]. 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 [7].

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, and 1 forward)
- Limited free transfers between gameweeks (typically 1 per week, with additional transfers costing points) [8]

Scoring System

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

- Appearance (playing at least 60 minutes): 2 points
- Goals: 6 points (midfielder), 4 points (forward), 6 points (defender/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) [8]

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 per season:
  - 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
  - Assistant Manager: Add a manager to your team to score points for three consecutive gameweeks
  - Wildcard: Unlimited free transfers that permanently change the team [8]

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 [9]

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 [10, 11]

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 [12]

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 [13].

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 [14].

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 [15, 16]

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) [17]

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 [18]

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, fixture schedule
- **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 [19, 20]

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 [21]

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 [22]

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 [23]

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 [24, 25].

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 [26, 27]

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 [28, 29]

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 [30, 31]

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

3 Methodology

This section might vary depending on the field of research. Consider this the approach, design, or architecture of the proposed system. You can also include implementation if it applies to your research.

Include at least one figure. Figure that shows the architecture of the proposed system, concepts of the proposed algorithm, etc. One figure that you could put into a slide that tells your idea.

Include actual implementation details. Describe the implementation language, platform, location, and dependencies on the packages.

4 Results

Organize the result section according to major topics.

4.1 subsection

The reader will scan through the section headers at the first pass. Subsection headings help organizations and help the reader find the parts of interest.

Make section headings as specific and information-rich as possible. Make sure to interpret the results. Summarize the collection of experimental results as clearly and as economically as possible.

4.2 How you should prepare figures

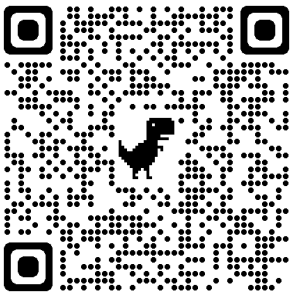

Figure 4.1: QR Code for Poster Session Program, Spring 2024.

The figure and table captions should contain enough context so that a reader can understand the content of the figure or table without having to refer to the text.

Any labels or uncommon abbreviations need to be explained in the figure or table caption.

4.3 In the Results section..

- Organize this section according to major topics.
- Subheadings to make the organization clear and to help the reader
- scan through the text to find the parts of interest.

- Make these section headings as specific and information-rich as possible but try to cover the major common points.
- Interpret your results. (e.g., How do your results compare to related studies? Do your results accord with theory or are they surprising in some way? What are the underlying mechanisms that may explain what you found?)
- Summarize the collection of experimental results as clearly and as economically as possible.
- The figure and table captions should contain enough context so that a reader can understand the content of the figure or table without having to refer to the text.
- Any labels or uncommon abbreviations need to be explained in the figure or table caption.

5 Related Work

Avoid simply summarizing the findings of each paper, one paragraph per paper. [32] discusses..

Instead, synthesize the papers by identifying themes and use the papers to reference for the themes.

6 Conclusions

A good conclusion leaves readers with a clear statement of your point and a renewed appreciation of its significance.

Start with a brief statement with your research questions. Describe the main results. Provide a summary of the conclusion. Emphasize the contribution of your work.

A title is the first thing readers read and the last thing you should write. It is not just a few words to suggest the topic of your papers. A title is useful when it helps readers understand specifically what is to come. Put into your title the keywords in your main point.

7 Future Work

Bibliography

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

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

[3] 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.

[4] 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.

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

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

[7] 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.

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

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

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

[20] 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.

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

[22] 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.

[23] 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.

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

[25] 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.

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

[27] 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.

[28] 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.

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

[30] 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.

[31] 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.

[32] Mohammad Hosseini and Viswanathan Swaminathan. “Adaptive 360 VR video streaming: Divide and conquer”. In: *2016 IEEE International Symposium on Multimedia (ISM)*. IEEE. 2016, pp. 107–110.

A Appendix Title

Instructions on how to reproduce your experiment. Anything that is not covered at the main but you want to add. Additional plots/results, etc.