CSE 519: Project Proposal- Fantasy Premier League

Better Late than Never - Suboptimal Player Transfer Strategy

Fall 2021, Stony Brook University

Abstract— **The problem of determining the optimal game play strategy to win the FPL league is challenging considering factors like the order of player combinations for optimal team building and the uncertainties like injuries, penalties etc. In this project we relax the problem statement to a "better late than never" player transfer to suggest the next better player profile. An explainable "recruiter" model will use a lazy approach of finding a suboptimal player with critical analysis of player performance. The profile is then used by a "scout" model to statistically estimate an actual player from rankings table. The recruiter model's "player selection polices" are evaluated using a reinforcement learning approach for cumulative points projection. We aim to analyze the criticality of decisions by simulating the projected scores of policies from the model and understand how they differ. We compare the decisions of top players, understand the significance of long term vs short term decisions and explore interesting questions about strategic transfers.**

I. INTRODUCTION

The Fantasy Premier League is a popular online game where millions of players strategically build their own fantasy football team and compete for points. Given the nature of uncertainity, in any EPL match the precise estimation is extremely difficult.

In FPL a total number of 15 players consisting of 2 Goal Keepers, 5 Defenders, 5 Midfielders and 5 Forwards have to be choosen. Out of these 15 players, 11 players are selected to play in a particular match. Every player has to be bought at their current market price and the total price of the team must not exceed 100 million. Out of the 11 players that are selected for a match there has to be atleast 1 Goal Keeper, 1 Forward and 3 Defenders.

Players are given scores based on their performance in the premier league. The scores depend on the type of the player i.e MidFielder, GoalKeeper, Defender and the player's action such as number of goals, assists, clean sheet where they gain points and also can lose points for red cards, yellow cards etc. Incase the player's team doesn't end up playing that gameweek the players get a score of zero creating a 'Blank Gameweek'. A missed match in a gameweek is played later in a 'Double Gameweek' where the player scores points for both the original match of the gameweek and the 2nd match played in place of the missing match previously. The 3 best performing players in every match are awarded bonus points. For every team a captain and vice captain is nominated.

Unlimited transfers can be made between players until the first deadline after which 1 free transfer is available per game week. For every additional transfer after the free transfer 4 points are deducted from the total score. Maximum 1 transfer can be saved at any point. Wildcards allow for unlimited transfers without losing any points but only 2 wildcards can be used in one season. (free hit¿ do we add?) Player's cost can change across the season depending on the player's popularity and performance.

The objective of the project is to predict the scores of the player for a specific game week and find the optimal team according to the player's scores. While choosing the players for the optimal team the constraints in regards to choosing players of a specific type and not exceeding the budget as mentioned above have to be taken into consideration.

A. *FPL Strategy*

The choice of captain is very important to maximize score as the score of the captain doubles every time it's team wins. Avoid Goalkeeper transfers as goalkeepers usually don't get a large amount of points each week so saving the transfer for some other position is a better option. Avoid transfers of consistently performing players if they have a bad gameweek. Fixtures would help analyze if the player has a comparatively easy or difficult match coming up which helps in making the transfer decision. Transfer can used to secure home field advantage,i.e choosing players who score the maximum points on their home fields

II. BACKGROUND RESEARCH

In the article [2] they try to identify the relation between the statistics of the player on the field and the actual FPL Points scored. To what extent can the underlying statistics predict the player's weekly performance. To analyze the impact of fixtures on player's FPL performance. Identifying ways to create the optimal team depending on the value of the player and other alternative player options in the same position. The main approach here was applying linear regression on a group of underlying statistics and FPL Points and then finding correlation between the feature and FPL points.It was found that goals conceded, saves, touches and recoveries were the strongest predictors of FPL points. The analysis to determine the best way to choose a team was done on the basis of value of players in each category and the impact of fixtures on categories.From the analysis the author suggests to invest more in expensive midfielder and defensive positions as rather than pricier goalkeeper and forwards as

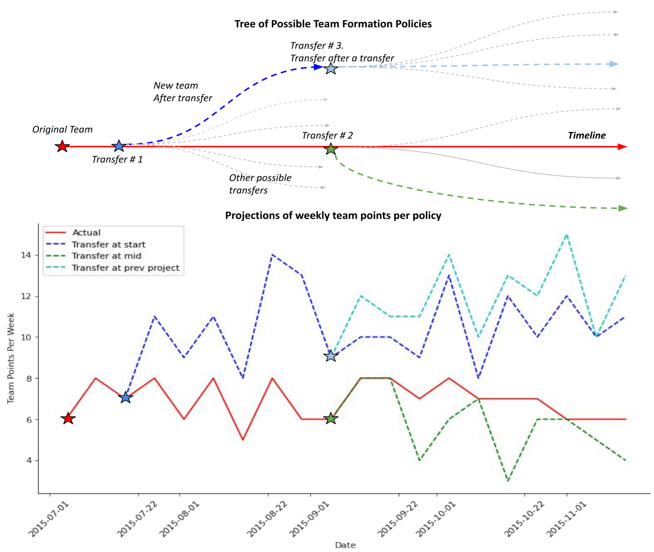

Fig. 1. Point Simulations of suggested policies and Tree Space of Possible Policies

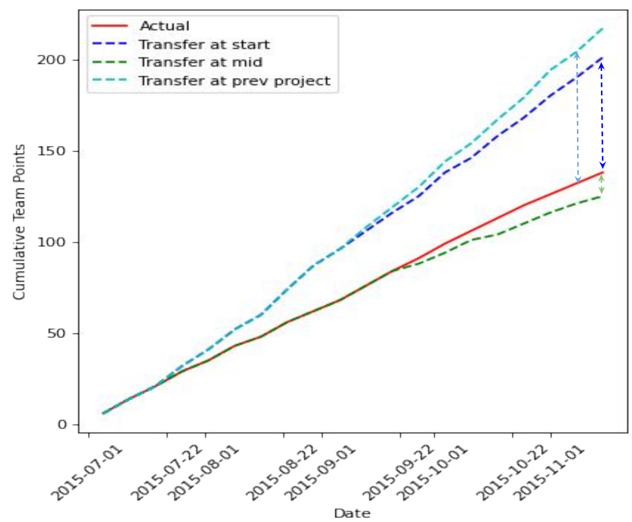

Fig. 2. CDF of points for the player transfer policies

for the later the value for money is lower relative to their less expensive counter parts.

In the article [4], they mention new statics like 'expected goals (xG)' to make player selection. Another is 'Average Formation', which will track the positions of all players when a team is in and out of possession. Other few are 'Live win probability' , 'momentum tracker'.

In the paper [5], they suggest a new method for predicting future player performances by automatically incorporating human feedback into our model so as to handle uncertainties like as injuries, managerial decisions. There are online tools like [6], [7] and available APIs to access datasets and get statistical inferences.

A. Reinforcement Learning

The FPL game week data is turned into a simulator where we use it to calculate the current state of a team and get the points of a team at a particular game week. See sections V-C,V-D, V-B. In the course content [3], a generalized policy iterative method can be used to train a neural network which will learn the optimal policy.

III. Challenges

All the challenges regarding section are listed here with a few major questions that need to answered.

- Ranking of players
- *Point Metrics for Players*: The FPL provides a list of metrics for players like – influence, creativity, ICT index, etc. What weightage to give to each of them?
- *How to measure performance of a player?*. There are a ton of statistics available for a player like – goals, saves, assists, influence, accuracy, etc. How some of the features have been calculated like creativity is totally a black box.
- *How to measure performance of a team / how good is a team?*. To estimate the performance of an actual EPL team, we need a way to estimate the performance of the actual EPL players. The performance not only depends on independent player performance but also on other factors like – other players in the team, the position he plays at in the team, match substitutions or injury, etc
- How to compare two teams?
- Momentum in a game - We need per game match statistics for this.
- Which should be given more importance form vs fixture?
- *Training the Neural Network*: Need to check training for loss convergence. Total runtime and resources needed can be substantial.

IV. Dataset

The official FPL API to be used to get data about 1.Actual Football Players Some of the attributes that can be fetched from FPL API are goals scored, assists, points, minutes, cards, clean sheets etc. Figure 3 shows the points of 3 players(Salah, Max, and Mane) over a season of 38 game weeks. 2.Actual Football Teams The API provides data for 20 EPL teams with details of each of the 38 fixtures in the season. The API provides data like – home/away game, attack, defense, scored, missed etc. 3.Player with the highest points in each of the 38 game weeks for a season. 4.The API provides data for fantasy teams of 8432544 FPL Managers. 5.Game Week 6.Fantasy League Rankings (Classic League) This is needed for validating where does the team suggested by our model land in the league rankings. The FPL API provides a way to get data for all FPL managers and the rankings of their fantasy teams. Some of the attributes provided are – current rank, previous rank, points etc. 7.Squad of FPL Teams for each Game Week This is needed for answering questions related to what players do top FPL Managers choose every week, how do they swap players to maximize their scores, etc. The attributes provided are player, multiplier, is captain, is vice captain, etc. 8.We'll be collecting the dataset using this API for our analysis of 38 game weeks. 9.We'll also be using the FPL API to fetch the scores of all fantasy teams to – Calculate the rank our team. Identify optimal transfer strategy. Quantifying cost of right/wrong player selection/strategy. 10.Getting the dream fantasy team with different constraints like – budget, wildcards, etc. for each gameweek from https://fploptimized.com/week.html . The data can be downloaded as csv.

V. METHOD

Maximization of an FPL team's cumulative points is the objective of a player. The underlying challenges include the estimation of expected goals scored, assists or possible penalties incurred by any player during an FPL gameweek. However to create a player transfer policy, we also look at the long term effects of having a set of players within a team. We build our deck by first statistically constructing a quantifiable profile for the entities- EPL player and EPL team and rank them. From the FPL API data, we categorize the force vectors of a game like goals scored, assist and penalties into groups of positive, negative and neutral forces which have proportional effects on a teams score stated in V-B. We then create a tree space of possible transfers and estimate the goodness of each transfer. The model is trained over this space and estimates atleast the next best policy which can be as good as the optimal strategy in section V-D.

A. Features

We establish a hierarchy of 3 entities - week, team , player. After analyzing the features in our Dataset, we identified and binned columns into positive, negative and neutral forces. For example, form, points, goals, clean sheets, and the ICT(influence, creativity, threat) index are positive. Goals conceded, yellow card , red cards are negative. Features like chances of playing are neutral forces. These are important for player performance evaluation - overall and current in the FPL. Performance metrics are aggregated to get team performance and the general trend of games in the week.

B. Player Selection Scout Model

In player selection, the recruiter policy model would suggest a player transfer with an estimated profile. We use this profile embedding in our player selection model to match with a top player within the budget constraints. In our deck, we will create a rankings table for the players and team using features as stated in Section V-A. The rankings table will be dynamic allowing for player selection by player type or player characteristics like speed, dribbling, strength etc.

$$f(R_t|C_t, p, t, \alpha) = topKmin_p MSE(\{i : C[R_t[i]]] < c\}, p) \quad (1)$$

where f denotes the player finder function, R is the rankings table, C is the budget/cost table and t is the current game week. We set α parameter for function f for other parameters.

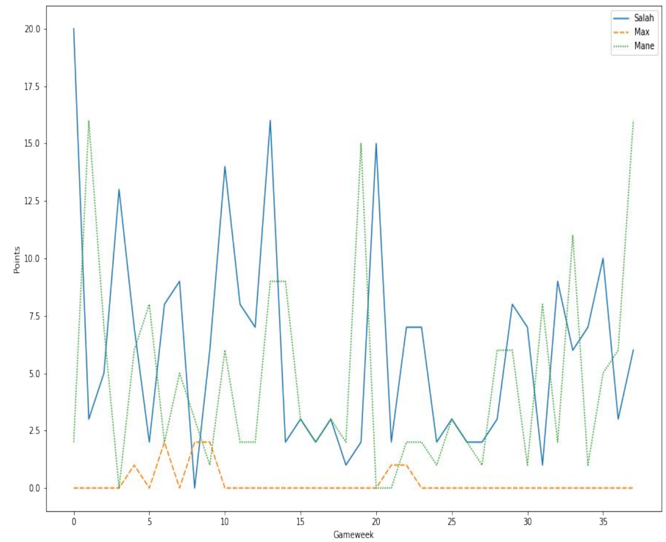

Fig. 3. Distribution of Points for 3 Sample Players over 38 game weeks

We will use mean squared error to find the effective match of the suggested profile with our rankings table. The model can be a decision tree or simple rule based function. Team performance needs to be taken into account for player selection model to accommodate team balance. Swapping out a player shouldn't disrupt the team balance and so player selection must be regularized to not overfit to the policy model suggestions. The game week's momentum is to be explored for player selection.

C. Policy Tree

A policy $\pi(a|s)$ is basically the set of transfers over the starting team and so we get a series of snapshots of the team over time. In Figure 1 each path in the policy tree from root to the leaf is a different policy and each node is a transfer. In the time series data, we have the player FPL scores which can be used to build a team with all possible combinations and thus create our policy tree as shown in Fig 1. The term policy and transfer strategy is used interchangeably here.

$score^t_{team} = \sum_{p=1}^{P} X_{points}(p, t)$ is the team score for set of players P at time t. This is a exponentially growing tree and thus becomes an NP-Hard problem. There are $O(N^W)$ combinations with N actions and W weeks. If this was computationally feasible, the optimal policy can be found by just taking the greedy approach at each game week and finding the CDF. In Figure 2 the current FPL league score of the FPL player is shown for 20 weeks. To cut down on this parameter space, we plan to do predictions for n equidistant points in the whole timeline.

D. Recruiter Policy Model

The problem statement has been modeled using a reinforcement learning approach where the model θ creates a policy $\pi(a|s)$ given the state s and action a. The state s is a matrix of player embeddings which is basically the set of players in the team. We define as discrete action space

A contains the factors in player selection like goals score, assists, clean sheets , form etc. With the Markov assumption of state-action dependencies, we just need to estimate the value function $v(s)$ which gives us an estimate of how good is a state. The value function is given by the Bellman Expectation equation for Generalized Policy Iteration and is given as,

$$v_{k+1} = \sum_{a} \pi(a|s) \sum_{a,s'} p(s', r|s, a)[r + \gamma v_k(s'))] \quad (2)$$

where p is the probability model of state to state changes, γ is a dampening factors for the projected rewards over time. Since the P is given this becomes a dynamic programming problem with the objective of finding the best combination of transfers [1].

The best policy is chosen as one with the better value function. The policy model is trained on simulated transfers starting with a random policy. With an ϵ-greedy approach, the model is trained to learn the best states with a weighted greedy function. After enough policy iterations, the policy updation will converge to the optimal policy. We ensure that the optimal policy will be atleast be better than the current policy.

The model will ideally be a neural network which will learn the value function and is trained with generalized policy iteration. It will take in inputs as the state embedding and action to output the value for the state. The action vector which gives the best value is also the player profile that we need. The player selection is done with this action vector as stated in Section V-B

$$\pi(s) = argmax_a(q(s, a)) \quad (3)$$

where $q(s, a)$ is the value function over action. Example player profile for an attacking player will have higher weightage for the goal feature. For a defender there might be a higher weightage for goals saved. A Midfielder might have a balance for both.

E. Simulations of Transfer Strategies

Figure 1 shows the per week team points. We define 2 transfers each at different points in time of the actual projection. We define 1 transfer over another projected transfer. Even though this is a new team we have the available data to compute the projects score of each formed team. From figure 2 you can see different divergences for different policies with Transfer 2 worse than actual but transfer 1 and 3 being better. Such CDF plots helps to compare policies and do critical analysis of stage by stage transfers.

Plots of FPL manager vs manager can be made to see how top FPL players differ from normal players or find what decision made a huge difference in a managers gameplay.

F. Player Transfer Policy Baselines

The general strategy for winning the FPL League is usually a greedy approach to effectively maximize the team's cumulative points. Though the problem statement is relaxed with no unique solution the problem is still not trivial. We setup baselines for bounding the performance. A random baseline where players are transferred randomly is needed. Another baseline for commonly followed strategies like transfer out least performing player, buy the best performing player is also needed. We plot projections of points for each of these policies.

References

[1] Richard S. Sutton and Andrew G. Barto. 2018. Reinforcement Learning: An Introduction. A Bradford Book, Cambridge, MA, USA.

[2] https://mathematicallysafe.wordpress.com/2019/07/01/fpl-analysis-what-five-seasons-of-data-modelling-have-revealed-about-predictive-analysis-fixture-impact-and-optimal-team-structure-in-fantasy-premier-league/

[3] CSE 525: Robotics, Stony Brook, Micheal Ryoo, http://michaelryoo.com/course/cse525/

[4] https://www.forbes.com/sites/stevemccaskill/2021/05/06/premier-leagues-new-ai-match-stats-will-predict-if-a-team-will-score-in-the-next-10-seconds/?sh=5f06318a6084

[5] Multi-stream Data Analytics for Enhanced Performance Prediction in Fantasy Football, Bonello et al,https://arxiv.org/abs/1912.07441

[6] FPL analytics , https://www.fplanalytics.com/

[7] FPL statistics , https://fantasy.premierleague.com/statistics