Electronics and Computer Science Faculty of Physical Sciences and Engineering University of Southampton

Thesis Report Submitted for the Award of PhD in Computer Science

Artificial Intelligence in Team Sports

Author:
Ryan Beal

Co-Supervisors:
Prof. Timothy Norman
Prof. Sarvapali Ramchurn

January 10, 2022

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

Artificial Intelligence in Team Sports

by Ryan Beal

The Sports Analytics Market is growing rapidly, in 2020 it was valued at over $1 billion and is expected to reach over $5 billion by 2026. However, even with this level of growth the use of Artificial Intelligence (AI) techniques have yet to fully be explored. The sports analytics domain presents a number of significant computational challenges for AI and Machine Learning. In this thesis, we propose a number novel methods for analysing team sports data to help sports teams utilise AI to improve their strategic and tactical decision making. By doing so, we present a number of contributions to the AI and sports analytics communities. In particular, we present a model for the tactical decisions that are made in football and show how game theoretic techniques can be used to optimise these. We focus on both the short-term decisions made for individual games, as well as longer-term decisions to maximise performance over a season. We show that we can increase a teams chances of winning individual games by 16.1% and can increase a teams mean expected finishing position by up to 35.6%. We also, introduce a new model for valuing the teamwork between players in sports teams by assessing the outcomes of chains of interactions between the players in a team. We then present a novel model for forming teams based on this value and maximise teamwork by assessing the overlapping pairs in a team. Our model is shown to better predict the real-world performance of teams by up to 46% compared to models that ignore inter-agent interactions. Finally, we show how we can use natural language processing techniques to improve the traditional statistical methods for prediction sports match outcomes. We use domain expert written articles from the media to train our models and we show that by incorporating the features learned from the text, we can boost the accuracy of the traditional statistical methods by 6.9%.

Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I would like to express my sincere gratitude to my PhD supervisors Professor Gopal Ramchurn and Professor Timothy Norman for their help, advice and support throughout my studies. In particular, thank you to Gopal for inspiring me back in 2014 during the first year of my undergraduate degree, by showing me that it is possible to combine computer science with football, and supervising my third-year project where I first explored the use of AI in the sports domain. This started my journey to my research career allowing me to combine my passions.

Secondly, I would like to thank my co-authors and collaborators who have helped with the publications throughout my PhD. These include Dr. Georgios Chalkiadakis, Dr. Stuart Middleton, Narayan Changder and Charbel Merhej. I would also like to recognise the support offered by all my colleagues within the Agents, Interaction and Complexity (AIC) research group, particularly as half of my PhD being completed from home due to the COVID-19 pandemic.

Next, I would like to thank Dr. Tim Matthews for his advice and input into my research throughout my PhD and for being a co-founder of SentientSports. Thanks to Ramm Mylvaganam and AI Abacus for helping my research have the greatest impact by connecting me with sports teams across the world. Also, thank you to the Empati and Future World teams for their guidance on transitioning research into real-world applications.

A special thank you to my partner Ffion for spurring me on to start my PhD, for her love and encouragement throughout, and for consistently correcting my inability to use apostrophes. Finally, I would like to thank my parents and all my family for their love, support and for always pushing me to be the best I can be and to follow my passions.

Chapter 1

Introduction

Sports is a domain that has grown significantly over the last 20 years to become a key driver of many economies, while at the same time, impacting our social and cultural fabric. According to a recent report,[1] the estimated size of the global sports industry is \$1.3 trillion, and has an audience of over 1 billion people, who may attend matches to support their favourite teams, bet in various online or offline markets, or watch games on the television for pure entertainment. Sport accounts for an estimated 1 million jobs in the UK alone, with those involved in either playing games, managing teams or looking after the health and fitness of players. At the core of these economic and social impacts, are the individuals, players, and teams involved. Indeed, as we will demonstrate in this chapter, predicting and optimising the performance in sports are challenging problems but, so far, such problems have largely been dealt with by domain experts (e.g., coaches, managers, scouts, and sports health experts) who rely on basic statistics. Specifically, we focus on *team sports* as they present the most difficult challenges, and tend to have the greatest audience and economic benefit.

We define a team sport as a game that typically involves two teams playing against each other, each composed of a set of players with their individual roles and abilities. There are many uncertainties in team sports that affect the final outcome and performance of the teams. These decisions range from team selection, tactics (e.g., choosing where players should be placed on a football field), player transfers (e.g., choosing which players should be sold to or bought from another team) and planning training sessions (e.g., to help players recover from injuries or improve the collective performance of a team). The results of such decisions can sometimes be quickly obtained and learnt from (e.g., tactics may fail or succeed during a live game) or come through over a long period of time (e.g., a player may recover differently based on different long-term training regimes or preparatory matches).

[1] https://www.plunkettresearch.com/statistics/Industry-Statistics-Sports-Industry-Statistic-and-Market-Size-Overview.

In recent years, the field of team sports (teams, governing bodies, academies etc.) has adopted a range of technologies that collect large amounts of data from training and matches that include the movement of players during games, their health statistics, and their performance during games. Players train and compete while being monitored by various sensors to gain more information about performances.[2] Data acquired in this way helps coaches and managers optimise training sessions and further improve performance. For example, companies such as StatsBomb[3] and STATSPerform[4], specialise in collecting and distributing sports data to teams and media outlets. Major teams around the world already use a variety of datasets to make decisions and improve their on-field performances. This tends to lead to an increase in prize money, higher proportions of TV rights and more sponsorship deals. For example, the promotion of an English Championship football team to the English Premier League is worth £200 million in extra revenue.[5] Professional betting companies also use such datasets to exploit inefficiencies in the sports betting markets to generate profits. Some hedge funds use the sports gambling markets as a way to make investments and exploit these sports betting market inefficiencies.[6]

From a scientific point of view, the availability of such datasets presents a unique opportunity for the Artificial Intelligence (AI) and Machine Learning (ML) communities to develop, validate, and apply new techniques in the real world. Indeed, several works attempt to solve real-world challenges e.g., in disaster response (Ramchurn, Huynh, et al. 2016; Ramchurn, Rogers, et al. 2008), and in autonomous unmanned aerial vehicle (UAV) (Valckenaers et al. 2007; Pujol-Gonzalez et al. 2013) with AI techniques, tend to rely on synthetic environments that ascribe standard probability distributions to the behaviours of entities involved or the external phenomena that impact on their behaviours. For example, simulating the spread of fires in disasters and the ability of fire brigades in extinguishing them, or the changes in energy consumption in power grids due to changes in energy pricing. In contrast, real-world team sports data is available over long periods of time, about the same individuals and teams, in a variety of environmental contexts, thereby creating a unique live testbed for AI and ML techniques. Recent works such as (Matthews, Ramchurn, and Chalkiadakis 2012) and (Vilar et al. 2013) have proposed and validated novel performance prediction and combinatorial optimisation solutions that have advanced the multi-agent and machine learning state of the art. Recently, Tuyls et al. (2021) explored the use of AI in sports strategy with a focus on football. Here, the authors stress that football analytics offers tremendous value for both advances in football itself, but also for the field of AI.

[2]https://www.forbes.com/sites/bernardmarr/2015/03/25/big-data-the-winning-formula-in-sports.
[3]https://statsbomb.com.
[4]https://www.statsperform.com.
[5]https://www.telegraph.co.uk/football/2016/05/28/play-off-final-how-much-is-premier-league-promotion-really-worth.
[6]https://www.businessinsider.com/inside-story-star-lizard-tony-bloom-2016-2.

One key area where AI can benefit team sports is in the decision-making processes made at all sports organisations. Often decisions are made using subjective information and opinions of those in leadership roles within sports organisations. Examples of the decisions made include player recruitment, opposition analysis, tactical selection and training procedures. These decisions are often made under high pressure due to time constraints (college drafts, transfer deadlines, etc.) and could be worth significant sums of financial gain/loss (Massey and Thaler 2013). In Figure 1.1, we propose a decision-making framework of typical team sports.

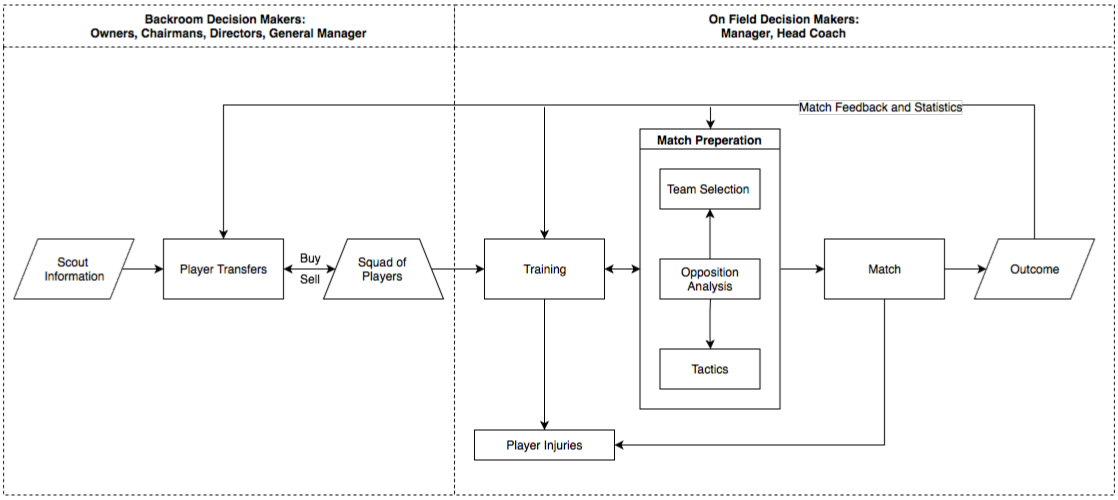


FIGURE 1.1: The Decision Making Process in team Sports.

This framework enables us to identify key research areas and questions. While research in AI for team sports has grown over the last 20 years, it is unclear how the field of AI and sports relate to each other or build upon each other as to date work tends to either focus on specific types of team sports or specific prediction and optimisation problems that are but one part of the whole field. Hence, in this thesis, we propose a number of research questions for AI in team sports and evaluate the performance of key models in the area. In Chapter 3 we focus on the "Match Preparation" section of Figure 1.1 and present new methods that allow decision-makers to optimise their tactics (e.g., team selection, formation, team style, etc.) by using game-theoretic techniques. We also show how we can learn from games played and factor in the longer-term impact of decisions. In Chapter 4 we propose a model to value teamwork, which is able to be used across the decision-making framework. We discuss uses for selecting the best teams to maximise the expected teamwork value but our methods could also be used for player recruitment by predicting how well players will work together or in training to build better teamwork amongst the current set of players. Finally, in Chapter 5 we focus on the "Match" section of the framework. We propose a Natural Language Processing (NLP) model on human-expert text to improve on the traditional statistical methods of prediction. This can help decision-making by allowing teams to predict performance and see the potential impact on the pitch of any decision made.

In the next subsection, we discuss the contributions of this thesis, their impacts on sports games and how this relates to our decision-making framework.

1.1 Contributions

This thesis presents a number of models that have contributed to both the AI and sports analytics communities. These can be summarised as follows:

- **Tactics Optimisation:** We proposed a novel mathematical model for the game of football and the tactical decision-making process. Using real-world data from 760 real-world football games we can learn the payoffs for different team actions and learn state transitions. We show that we can predict game-state transitions with an accuracy of up to 90%. We also show we can accurately predict opposition tactical decisions. By learning action payoffs, we can optimise pre- and in- match tactical decisions to improve the probability of winning a game. This work has been peer-reviewed and can be found in the following publication "Optimising Game Tactics for Football". Ryan Beal, Georgios Chalkiadakis, Timothy J. Norman and Sarvapali D. Ramchurn. (2020). In: *Proceedings of the 19th International Conference on Autonomous Agents and Multi Agent Systems*, pp. 141–149.

- **Long-Term Planning:** We propose a mathematical model for optimising the long-term performance of human teams and apply this to the game of football. We introduce a *fluent objective* to model the moving goals of the team over long periods. This is based on accurate league simulations and further improves individual game payoffs by using knowledge from prior games. In particular, we show that we can increase teams finishing position on average by up to 2.9 ranks (out of 20). By using a fluent objective and prior game knowledge we are able to show an increased probability of improved long-term performance in real-world football teams (by up to 35.6%). This work has been published and can be found at "Optimising Long-Term Outcomes using Real-World Fluent Objectives: An Application to Football". Ryan Beal, Georgios Chalkiadakis, Timothy J. Norman and Sarvapali D. Ramchurn. (2021). In:*Proceedings of the 20th International Conference on Autonomous Agents and Multi Agent Systems*, pp. 196–204.

- **Valuing Teamwork:** We propose a number of network metrics to capture the contributions of individuals and sets of agents. We show that by using machine learning models we can extract the value of teamwork which can be learnt from data and then applied to the prediction of team performance. This is applied to both football and basketball and, we argue, can be applied to team formation more generally. This work has been peer-reviewed and is published at "Learning the Value of Teamwork to Form Efficient Teams". Ryan Beal, Narayan Changder, Timothy J. Norman and Sarvapali D. Ramchurn (2020). In: *Thirty-Fourth AAAI Conference on Artificial Intelligence*,pp. 7063–7070.

- **Team Formation:** We propose a novel approach to team formation based on the value of inter-agent interactions. Specifically, we propose a model of teamwork that considers the outcomes of the chains of such interactions. This was also covered in the above at AAAI-20 paper.

- **NLP in Sports:** We propose a novel combination of Open Information Extraction, Sentiment analysis and supervised ML methods for predicting the outcome of games of football using human opinions from domain experts in the media. Our approach uses a previously unexplored feature set in terms of football match outcome predictions which can factor in human knowledge which is overlooked in traditional statistics. We test and validate our approach by predicting the outcomes of 1770 football games over 6 seasons showing that we can boost the accuracy of statistical approaches by 6.9% when predicting the outcome of events. This work can be found at "Combining Machine Learning and Human Experts to Predict Match Outcomes in Football: A Baseline Model". Ryan Beal, Stuart Middleton, Timothy J. Norman and Sarvapali D. Ramchurn (2021) In:*In the Proceedings of the Thirty-Third Conference on Innovative Applications of Artificial Intelligence*, pp. 15447-15451.

- **Other Contributions to Sports Analytics:** As well as the contributions to the AI community outlined above, during the thesis we have also contributed to the sports analytics community with 2 publications (Beal, T. Norman, and Ramchurn 2020a; Beal, T. Norman, and Ramchurn 2020b) to the International Journal of Computer Science in Sport (IJCSS) and a co-supervised paper (Merhej et al. 2021) with a masters student at Knowledge Discovery and Data Mining (KDD-21). We also organised the *AAAI-20 Workshop on Artificial Intelligence in Team Sports* which bought together academics from both AI and sports analytics as well as key industry leaders.

1.2 Structure of Thesis

The remainder of this thesis is organised as follows:

- **Chapter 2:** provides a full literature review of work to date in team sports and the different areas of research which can be broken down to match prediction, decision-making, fantasy sports and injuries. This chapter also outlines the research questions that the rest of this thesis will focus on.
- **Chapter 3:** In the first research chapter we explore the ability to use game theoretic techniques to optimise tactics both pre-match and in-match. We then expand on the model to factor in the longer-term impacts of the decisions made in football.
- **Chapter 4:** Next, we discuss how we can value the teamwork between players and introduce a new algorithm to form teams based on this teamwork to maximise the value between the selected pairs and the overlapping pairs within the team.
- **Chapter 5:** The final research question explores the use of natural language techniques to improve match prediction and shows how this can be used to boost the predictive power of traditional statistical methods.
- **Chapter 6:** discusses our overall findings from the thesis and the impact of the results from each of the chapters we have presented. We also discuss the future work and the future of AI in team sports.
- **Chapter 7:** summaries the conclusions from our research.

Chapter 2

Literature Review

In what follows, we elaborate on the four key areas that we have identified where decisions and predictions can be optimised due to the significant performance and financial benefits that they may have:

- **Match outcome prediction:** Predicting the outcomes of sporting events is an important factor for a number of stakeholders. According to a BBC report, the global sports betting market is estimated to be worth around \$244billion with millions of bets placed all over the world.[1] This means that the prediction of match outcomes is key to the bookmakers who set the odds, and the punters who place their bets. Match outcome prediction is also an important factor for teams that affects their tactical decisions and overall recruitment and game strategy during a season. There are many uncertainties that may affect the result of a given game and we will elaborate on these in the rest of this chapter.

- **Strategic and tactical decision-making:** Many key decisions in the team sports process affect performance both in-game and behind the scenes. These decisions include player recruitment, tactics, team selection, developing youth players and managing injuries. Player recruitment is one of the costliest parts of team sports due to the price of purchasing new players and the wages that they demand. The world's highest transfer fee in football is £198million for Neymar in 2017 and the highest salary per season in the NFL is \$76million for Aaron Rodgers. The enormous value placed on these players is usually based on subjective measures by the clubs. This means that often, large sums of money can be paid for a player who never lives up to the expectations of their price tag.

[1] https://www.bbc.co.uk/news/business-44362134.

- **Fantasy Sports Games:** Fantasy sports are games (online[2] or via newspapers) open to the general public where competitors are challenged to predict the performance of real-world sports teams and players, and to choose artificial or "fantasy" teams composed of such players. The players in these fantasy teams are then awarded points based on their real-world statistics. It is estimated that over 50 million people play fantasy sports in the US while over 5 million people regularly play the Fantasy Premier League in the UK alone. Fantasy sports presents a number of interesting computational challenges that can be addressed using AI methods. These challenges include but are not limited to: prediction of individual player performance, forming optimal teams based on player performances, predicting the fluctuating player values, and creating betting strategies when entering fantasy teams into competitions.

- **Managing Injuries:** Injuries to professional players can have a huge impact on their careers. Injuries also cause the performance of a team to decline as well as costing teams large sums of money in wages to a player who cannot play. In an annual report by JLT[3], all the injuries in the English Premier League were assessed and it was shown that in the 2017-18 season, £217million in wages were paid to injured players. Due to this, teams in all professional sports are now investing significant efforts into predicting the risk of injury and helping prevent them. The predictability of injuries in sport is discussed in (Lysens et al. 1984) which suggests that injuries may be an area where AI could help benefit teams and players due to the success observed when predicting health issues in the past (Srinivas, Rani, and Govrdhan 2010).

In this review chapter (and the rest of this thesis), we focus our attention on the six most popular team sports in the world: Association Football[4], Rugby Union, One-Day Cricket, American Football, Baseball and Basketball. We explore the existing relevant literature, provide new insights based on our analysis of key statistics, provide a number of frameworks to structure the computational challenges involved, and highlight open areas of research. In this way, we will justify the novelty of the research questions that are explored in this thesis.

[2]For example: www.draftkings.co.uk, www.fanduel.com, fantasy.premierleague.com.

[3]https://www.jlt.com/our-insights/our-insights/how-injuries-have-affected-the-english-premier-league.

[4]Referred to as just "football" throughout this thesis.

2.1 Match Outcome Prediction

Prediction of sports match outcomes is a complex computational problem due to the range of uncertainties that can influence match results. These include: the team configurations, the health of players, the location of the match (home or away), the weather, and team strategies.

Typically match outcomes consist of up to three possible outcome classes: home win, away win and a draw/tie. The draw/tie is a more common result in football, but it is still possible in all the team sports we focus on. When predicting these outcomes, probabilities are assigned to each possible state that the game could end in. Some models also focus on predictions for the scoreline or spread. The scoreline is the number of points/goals scored by each team and the spread is the difference between the number of points scored by each team. These are typically more challenging to predict due to the increased number of possible outcomes. By assigning a probability to each possible scoreline in a match, we are able to solve different prediction problems.

The multiple sources of uncertainties that exist when predicting match outcomes are typically very difficult to characterise. In what follows, we highlight the accuracy of the bookmakers in team sports and elaborate on the approaches that have been applied to predict match outcomes, scorelines and points spread. We explore the earlier literature that exists in statistics as well as literature outlining ML approaches.

2.1.1 Bookmaker Accuracy

Some of the match outcome prediction problems considered in this thesis are more challenging than others. Bookmakers use sophisticated pricing models that assign "odds" to an outcome (which reflect the likelihood) to maximise their chances of making a profit (Graham and Stott 2008). By comparing who the bookmakers made favourite (shortest odds) and the actual match outcome, we calculate a percentage accuracy[5] and use this to evaluate how predictable each sports outcome is. This provides an estimation of the predictability of each sport. Bookmakers price markets based on their predictions of the match as well as using the bets that are placed as an indicator of the likely match outcome.

To demonstrate the variability across team sports, we focus on the prediction of match outcomes (see Figure 2.1). As can be seen, Football has the lowest accuracy showing it is the least predictable. This is to be expected due to the frequency of goals being far less than the frequency of points scored in the other sports (see Appendix A). A draw/tie is

[5] Odds that are given across the whole of the past season (2016/17) with historic odds and results data taken from https://www.oddsportal.com.

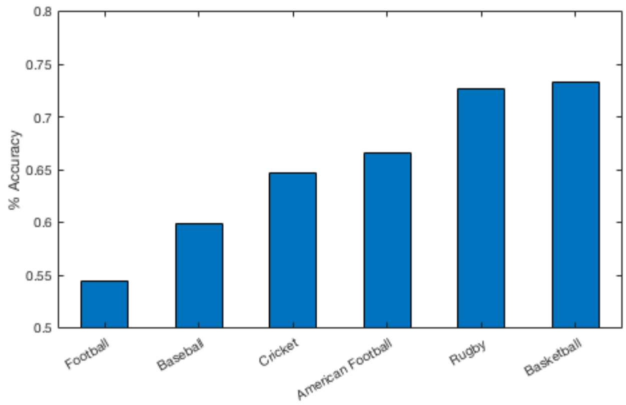


Figure 2.1: Bookmakers Accuracy Across 2017/18 Season.

also much more common in football meaning there are 3 possible outcomes to consider instead of just 2. Basketball is shown to have the highest accuracy by the bookmakers. This may be due to the high number of points scored in a game or a smaller playing area with fewer players.

2.1.2 Statistical Approaches

Some studies have focused on finding ways that the game of football could be modelled and to find inefficiencies in the UK football betting market. Dixon and Coles (1997) set out to exploit the inefficiencies and bias in UK football betting markets. Building upon the seminal work by Maher (1982), they developed an initial model to assign probabilities to each of the different game outcomes (home win, away win and draw/tie). Using this they are also able to form a new betting strategy. The model is based on the different abilities of both teams, calculated from prior matches. These abilities are broken into attack and defence and normalised based on the abilities of the opponents. Their model also takes into account a home advantage as discussed in (S.R. Clarke and J. Norman 1995). They can gain positive returns in a betting strategy. They use a technique based on a Poisson regression model, modifying Maher's basic bivariate Poisson model to give the equation shown in (3).

$$Pr(X_{i,j} = x, Y_{i,j} = y) = \tau_{\lambda,\mu}(x, y)\frac{\lambda^x exp(-\lambda)}{x!}\frac{\mu^y exp(-\mu)}{y!} \tag{2.1}$$

Where, $\lambda = \alpha_i\beta_j\gamma$ and $\mu = \alpha_j\beta_i$ and τ is a parameter for low scorelines fully defined in (Dixon and Coles 1997). In these equations x and y represent the goals scored by the home and away team respectively ($[x, y] \in \mathbb{N}$), α_i is the attacking parameter, β_i is the defensive parameter and γ_i represents the home advantage of team i ($[\beta_i, \gamma_i, \alpha_i] \in \mathbb{R}$). Finally i represents the ID of the home team and j the ID of the away team.

Dixon and Robinson (1998) studied the effect of the scoring rate changing depending on the current score of a game of football. They found that the scoring rate generally increases for both teams throughout the match, most likely due to the tiredness of players that leads to mistakes in defending. They also found the scoring rates of home and away teams depend on the current score. Each scoreline is modelled as a different game-state. When the scores are level, the scoring rates are similar to those at 0-0. If the home team is leading, the home and away rates generally decrease and increase respectively. If the away team is leading, the rates of both home and away teams tend to increase. Their findings can be used to find match outcome probabilities and to attempt to improve on Dixon and Coles (1997). This is done by finding the probability of each state and integrating over all the possible times and for each possible route to arrive at the final game state (x, y).

Crowder et al. (2002) again builds upon the work of Dixon and Coles (1997) by changing the original models' calculations of attack and defence efficiencies. The new framework assumes that the efficiencies evolve through time (rather than remaining constant) according to some unobserved bivariate stochastic process. The original stochastic process model is replaced with an approximation that yielded a more tractable computation without comprising the predictive power. Dixon and Pope (2004), evaluate the value and significance of the statistical forecasts from their earlier work in relation to betting market prices. They performed a detailed re-examination of match outcome odds and correct score odds across a number of years between 1993 to 1996. They suggest that the football betting market (at the time) remained inefficient[6] and the earlier models discussed in (Dixon and Coles 1997) could still be used effectively to earn positive returns when used with a strict trading rule to select the games to place a bet on.

More recently, (McHale and Scarf 2011) focuses on international matches instead of English League games. The authors present a new model for the number of goals scored by each team in a match and can be used for match outcome predictions. The model used in this paper is based on Copula functions (Nelsen 2006) which generate bivariate dependent discrete distributions which are used to forecast the match scorelines. As this paper is based on international football matches, it may not be as successful if used for domestic leagues due to significant differences between international and league football. In comparison to the domestic leagues, there is a gulf in quality between teams that could play against each other internationally. Furthermore, international teams do not play as often. Therefore, datasets detailing the performance of international teams

[6]Where inefficiency in the match outcome betting market will be indicated if returns are greater than the return on an uninformed, random betting strategy.

may not be as reflective of the current ability/form of the team and players. In a different study, Karlis and Ntzoufras (2008) use the Skellam distribution (Skellam 1948) to predict the winning margin of games during the EPL 2006/07 season. The Skellam distribution models the difference between two independent Poisson distributed variables. Using these distributions, probabilities are assigned to the possible goal differences and therefore the match outcomes.

Turning to American football, there are many applications of statistical techniques to predict match outcomes and scoreline predictions. A birth-process model (Harville 1980) uses a linear approach to create a baseline for NFL predictions in American Football, building on work that he had originally tested on college and high school American Football (Harville 1976). More recently, (Boulier and Stekler 2003) compare their model with human prediction and the bookmakers in the NFL between 1994-2000. They evaluate the use of "Power Scores" (published in the New York Times) as a predictor by creating forecasts generated from probit regressors. A probit model is a type of regression where the dependent variable can take only two values, e.g., home win or away win (Cappellari and Jenkins 2003). This model was able to improve the accuracy of the predictions made by human experts. However, it was unable to improve on the bookmakers' accuracy. In turn, Leung (2014) uses the teams' current ability based on other rating systems such as "Elo Ratings"[7], which were initially designed to rank chess players (Coulom 2007). Leung (2014) makes predictions on the outcomes of college American Football matches using historic results and a sum of other metrics (e.g., historic power indexes, Pythagorean wins, offensive strategy and turnover differential), the highest total sum is the predicted winner. The paper states that the model achieves high accuracy, but it does not detail how this was tested. Finally, (Baker and McHale 2013) looks to predict the exact scores in a game of American football. The authors use similar methods which were used for football in (Dixon and Coles 1997). The model takes each team's attacking and defensive abilities and finds the probabilities of the final state of the game scoreline using a Chapman-Kolmogorov forward equation (Gardiner 2009). This achieves an accuracy of 66.9% outperforming Boulier and Stekler (2003) who achieved 61%.

In basketball, (Zak, Huang, and Siegfried 1979) calculates the production efficiency of points scoring for each team and using the "Richmond" technique (Richmond 1974) they are able to estimate the potential scoring output of teams. Therefore, this could be used to make match outcome predictions. They also evaluate the basketball home-field advantage. Finally, "Yoopick" (Goel et al. 2008) outlines a different approach to create a sports prediction market. The market they create directly allows estimation of the entire point spread probability distribution within a single unified market. Punters bet on the outcome of the points difference of a game landing in a given interval with the interval prices determined by Hanson's logarithmic market scoring rule market maker (Hanson

[7] These ratings are a measure of strength based on head-to-head results and quality of opponent.

2007). This paper has yet to be tested against the accuracy of the more traditional betting markets.

In the next section, we will explore the ML methods that have been applied to the match outcome prediction problem for the sports this thesis focuses on.

2.1.3 Machine Learning Approaches

Many of the ML works involve Bayesian approaches and we focus on such approaches primarily. We also generally cover other approaches that have most recently come to the fore.

2.1.3.1 Bayesian Methods

Bayesian methods have been particularly popular as they can be used to express hypotheses (potentially by experts in the game) and then learn the parameters that can lead to more accurate predictions. As well as this, their ability to naturally quantify uncertainty makes Bayesian methods particularly useful in sports where it is likely there are relatively few observations to draw conclusions from. Rue and Salavesen (2000) apply a dynamic Bayesian linear model to estimate the time-dependent skills of all teams in the English Premier League (EPL). These skills are used to predict the outcomes of the matches. The model uses a Markov-Chain Monte Carlo (MCMC) method to make estimations on the attack and defence abilities of teams. The MCMC method is particularly useful to model the changing abilities of the teams across the season and therefore the abilities need to be updated after each game week. Previous results between teams are used to aid the predictions alongside the attack and defence abilities. They achieved an accuracy of 54%. At the time this was slightly better than the bookmakers' accuracy for the English Premier League and Division One results.

Joseph, Fenton and Neil (2006) compare a Bayesian approach to other machine learning approaches for predicting football outcomes. They test a number of algorithms on Tottenham Hotspur Football Club over the 1995–1997 seasons. The methods they compare are naive Bayesian Network (BN), a Data-Driven BN (learns the structure of the network by using the correlation between the attributes), a K-nearest neighbour implementation and a Decision Tree. The results confirm the potential of Bayesian Networks when they are built by a reliable domain expert. The advantage of this is the model is able to provide accurate predictions without requiring large datasets. However, this work is focused solely on predicting the outcomes of a single team's results which means it would have to be re-implemented for every team if used on a wider scale.

Following on from this, (Constantinou, Fenton, and Neil 2012) apply Bayesian Models to football match outcomes across two Premier League seasons. Their model (known as "pi-football") uses a selection of variables such as team strength, team form, team psychology and fatigue for both teams in a match to generate the outcome prediction. Some of their parameters are more subjective compared to team strength and form which can be calculated using the number of points a team has accumulated and goals scored/-conceded. The "pi-football" model can be used to generate profits against maximum, mean, and common bookmakers' odds. This model was improved further in (Constantinou, Fenton, and Neil 2013) by identifying the key features (e.g., team strength, form, and fatigue with motivation) to reduce the inputs into the model. The number of features are reduced from 21 in total (10 for each team plus one representing discrepancy) to 10 (5 for each team).

Other examples of Bayesian approaches to football match prediction include a study to predict results in the 2006 Germany World Cup (Suziki et al. 2009) and a Bayesian hierarchical model that was used to predict games in Italian football (Baio and Blangiardo 2008). In more recent work there have been studies for using Bayesian approaches to predicting games in-play (Robberechts, Van Haaren, and Davis 2021). In this paper, the authors introduce a Bayesian statistical framework that predicts the probability of the match ending in a win, draw and loss by using a set of contextual game state features as the game evolves over 90 minutes.

Moving away from football, Bayesian methods have been applied to other sports prediction problems. In American football, (Glickman and Stern 1996) use a state-space model with Bayesian diagnostics to predict games in the NFL (tested on 1993 season). This paper focuses on predicting the points spread, as this is the main betting market in the NFL. They produce good results when compared against the "Las Vegas betting line"[8] but were unable to outperform it. Thus, their model achieves an accuracy of 58.2% whereas the Las Vegas accuracy mwas 63% (at the time). When comparing the mean squared errors of the point differences the model achieved 165.0 which was better than the Las Vegas result of 170.5. In Baseball, (Yang and Swartz 2004) use a two-stage Bayesian model to predict the winners of games in Major League Baseball (MLB). Data from the 2001 season and an MCMC algorithm is used to carry out Bayesian inference and to simulate outcomes of future games. This model performs well and can accurately predict the winning percentage of an MLB team across a season but it does not state the accuracy when used for individual match outcomes. Finally, there is an example of Bayesian models being used for cricket outcomes in (Kaluarachchi and Aparna 2011). They test a number of methods to predict the winning team and their final model (known as CricAI) uses a Naive Bayes Classifier. On average they achieved an accuracy of 0.593 when using the Naive Bayes approach.

[8] https://www.complex.com/sports/2015/01/how-betting-lines-work.

The Bayesian methods that we have discussed in this section have produced some good results. However, they rely heavily on expert knowledge and also can be extremely intensive computationally for complex models. In the next section, we will explore other ML methods that have been applied to the match outcome problem and the results they achieved.

2.1.3.2 Other ML Methods

Many other machine learning methods have been applied to the sports match outcome prediction problem. In what follows we elaborate on these methods and summarise their key properties in terms of outcome prediction.

Jayalath (2018) considers ODI Cricket prediction and focuses on quantifying the significance of important features using classification and regression tree (CART) and logistic regression approaches. The study identifies that the key feature to improve a team's chances of winning is home advantage. Building on this, (Jayantha et al. 2018) creates a model for predicting ODI games using machine learning techniques and also outlines a team recommendation system. The prediction model in the paper uses an SVM model with linear, poly and RBF kernels. They use features such as batting and bowling averages to create power rankings for each player. The model takes the line-ups of the two teams and the player statistics in these line ups. The SVM models are trained with historic win or lose percentages. When tested with the linear, poly and RBF kernels they achieve an accuracy of 70.83%, 68.75% and 75% respectively.

As discussed in the previous sub-section, (A. Joseph, Fenton, and Neil 2006) apply other machine learning techniques beyond Bayesian approaches for football. A decision tree and a K-nearest neighbour model were developed. The MC4 Decision Tree achieved an overall average test percentage result of 41.72%. The K-nearest neighbour method uses a likeness approach, where the model finds similar instances to the test case and then a voting mechanism is used to predict the outcome. This performed better than the MC4 Decision Tree by achieving a test accuracy of 50.58%, the Bayesian approach in the same paper achieved an accuracy of 59.21%. Baboota and Kaur (2018) again looks at applying machine learning techniques to football match outcomes and compares the results to bookmakers. They use feature engineering and exploratory data analysis to find the feature set with the most important factors for predicting match outcomes. They use a number of features with different weightings such as form, shots on target, goals and more. They model the ternary classification problem to a binary classification one, and a prediction is made for whether a team will win the match or not. The methods that are tested by the authors are: Gaussian Naive Bayes, SVM (with RBF and linear kernels), random forest and gradient boosting. They use training data from 2005-2014 in the EPL and they find that the best performing algorithm was the gradient boosting method (56.7%), followed by the random forest (56.4%), SVM models (RBF 54.5%, linear 54.2%) and then finally the poorest performing was the Gaussian Naive Bayes method (52.6%). Similarly, (Hucaljuk and Rakipović 2011) test a number of features and classifiers. The features they use are the form of the team, previous meetings of the teams, current league position, number of injuries and average number of goals scored and conceded in a game. Six different learning classifiers are tested using these features: Naive Bayes, Bayesian Networks, LogitBoost, k-nearest neighbours, random forest and artificial neural networks. Datasets from the UEFA Champions League[9] (a cup competition as mentioned in Section 2.1) are used in this paper, focusing on only 96 games. They achieve an accuracy of up to 68% when using Neural Networks. This is considerably higher than results in the EPL. This may be because, in the Champions League, the best teams in Europe's top leagues compete against weaker teams from smaller football nations in the earlier stages of the competition meaning the match outcomes are more predictable. There are also fewer games played in the Champions League, so there is less data available for testing the models as shown by the test-set in this paper only using 96 games.

McCabe (2002) uses neural networks to predict games of Rugby League[10] in Australia. This work is extended in (McCabe and Trevathan 2008) where again a model is created with a neural network system using a multi-layer perception with a number of different features such as prior performance data, game location, team rankings etc. This model is able to perform well in Rugby League competitions with the average accuracy reaching up to 67.5%. This work was also applied to football results in the EPL. The results from this were compared to top human expert "tipsters" who also make weekly predictions on the same games in the form of a competition called TopTipper[11] and they were able to reach the top percentile with the model against the other human experts.

Shi et al. (2013) consider the problem of predicting college basketball games in the US NCAAB league. Five different machine learning models are developed: decision trees, rule learners, artificial neural networks (multi-layer perception), naive Bayes and a random forest, using data from 2009 to 2013. The methods all achieve between 68.4% to 74.5% accuracy. Their evaluation shows that a high level of accuracy is achieved when using neural networks, and this can be used to beat human predictors. Finally, part of the work performed in (Landers and Duperrouzel 2018) focuses on making predictions on NFL match outcomes and point spreads which they apply to "Pick'em" style[12] online competitions. Their model uses 28 features such as bookmakers favourite, average points (home and away), game location and more team performance-related statistics.

[9] https://www.uefa.com/uefachampionsleague/about.

[10] A similar but different sport to Rugby Union.

[11] http://www.toptipper.com.

[12] Pick'em is a game within Fantasy leagues where competitors guess who will win each American Football game in the NFL game that game week.

These features are used with an average perceptron and a boosted decision tree classifier algorithm to create their model. They tested the model over three NFL seasons and find the decision tree provided the best results achieving an average accuracy of 58%. This work is compared to (Boulier and Stekler 2003) which achieves 61% and to the bookmakers who achieve 65.8% accuracy. Finally, work in (Beal, T. Norman, and Ramchurn 2020a) compares a number of different machine learning methods for predicting match accuracy in the NFL.

In Table 2.1 we summarise the ML approaches that have been used for match outcome predictions.[13] These algorithms mainly use key team performance metrics as their features such as points/goals scored and conceded, league position, form etc. However, some key factors are not yet accounted for by the approaches that we have discussed. These are largely the external factors that can impact the results of sports outcomes (e.g., weather, player moods, changes in coaching, player transfers or impact of injuries).

ML Method	**Sport Application**	**Paper**	**Accuracy**
Neural Networks	Football	Hucaljuk and Rakipovic (2011)	68.8% (UCL)
	Basketball	Shi et al. (2013)	72.2%
	Rugby League	McCabe and Tevathan (2008)	67.5%
Decision Trees	Football	Joseph et al. (2006)	41.7%
	Basketball	Shi et al. (2013)	69.2%
	American Football	Landers and Duperrouzel (2018)	58.0%
KNN	Football	Joseph et al. (2006)	50.6%
SVM	Football	Baboota and Kaur (2018)	54.5%
	Cricket	Jayantha et al. (2018)	75.0%
Random Forest	Football	Baboota and Kaur (2018)	56.5%
	Basketball	Shi et al. (2013)	62.2%
Gradient Boosting	Football	Baboota and Kaur (2018)	56.7%
	Football	Hucaljuk and Rakipovic (2011)	68.7% (UCL)

Table 2.1: ML Approach Summary

In this section, we have evaluated the different approaches that have been used to make sports outcome predictions. Across all the different forms of predictions that we have discussed, all appear to reach a “glass ceiling” which we discuss further when highlighting our research questions. The papers we evaluated also show that football is the hardest game to predict due to the low scoring nature of the game.

There are many decisions that occur before and during a game in team sports. All of these have an impact on the outcome of sports matches. Therefore, in the following section, we explore some of the decision making processes that exist in team sports.

[13] UCL = UEFA Champions League, KNN = K Nearest Neighbours, SVM = Support Vector Machine.

2.2 Strategic and Tactical Decision Making

In this section, we turn our attention to the key decisions that arise when managing sports teams both on and off the pitch. In particular, to structure our discussion we propose a new framework discussed in the Introduction of this thesis (see Figure 1.1). This framework captures the key processes that operate in team sports and the interconnection among these processes that create a number of feedback loops. Using such a framework it is then possible to understand the importance of both machine and human decision making throughout. In more detail, player transfers presents a recruitment problem where teams want to ensure that they purchase the best possible players within their budgets. The squad of players then train to prepare for matches and develop their skills. During the training process, we can optimise the development of youth players to ensure they reach their maximum potential. The next stage focuses on decisions that are made to improve teams chances of winning games. This includes opposition analysis which supports the team selection and tactical decisions made by managers/coaches. Finally, these decision-making processes have feedback from the match outcomes and the in-game team performance. All of these will be expanded upon within this section.

2.2.1 The Recruitment Problem

The recruitment of new players is a different process in every sport and usually involves decisions from managers/coaches alongside the directors higher up in the sports organisations. In football, players are bought and sold between clubs (as discussed in Section 2) whereas, in many American sports players are drafted[14] and traded. In most cases, clubs gather information on players (scouting), therefore the amount teams pay for a player, relates to how well they think that the player will perform in the future and how much they will impact the team. Many elements add uncertainty to the process, namely concerning whether a player will continue performing well if the player will fit into their new team, if the player will settle into a new environment and if the player will stay free of injuries. These uncertainties are discussed when drafting a college player into the NFL in (Hendricks, DeBrock, and Koenker 2003). Here it is suggested that statistical discrimination and option value, influence choices in this market meaning that some players could be over-valued. Modelling the uncertainties that exist in future performances of players and predicting how well they will impact a team would provide huge benefits to sports teams. This will allow the decision-makers to evaluate the risk of the player before paying large sums of money. These types of predictions can also help assign a monetary value to a player, so that a fair price is paid. There are a number of factors that affect the price of a player some of these are explored in (Dobson and

[14] In a draft, teams take turns selecting from a pool of eligible players. Usually from a college or high school system.

Gerrard 1999). In Figure 2.2 we show the generic recruitment process that sports teams follow when investing in new players.

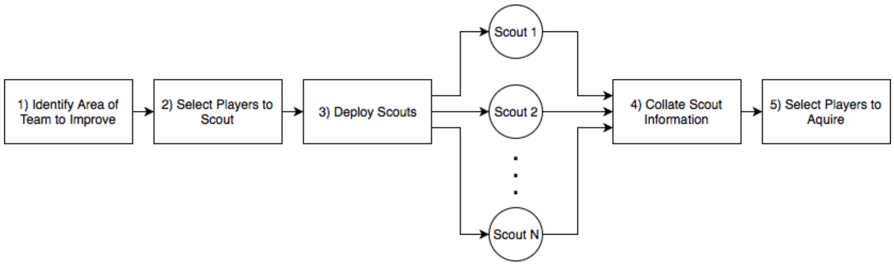


FIGURE 2.2: The Player Recruitment Process.

All of the stages within the player recruitment process present different challenges that can be improved through the use of AI methods. These are discussed below and correspond to the numbered processes in Figure 2.2.

1. The first stage is identifying which areas in the team need to be improved. We can go about this firstly by looking at the statistics of the team performance to identify what in particular needs improvement (e.g., more goals in football or more wickets in cricket). We can also highlight which individual players are not pulling their weight in the team and look to improve these.

2. The next process is gathering intelligence on a large set of players, this can be done in a number of ways. Teams have access to league datasets (statistical, event-driven and tracking) where they can find information regarding player info. These statistics are becoming more detailed and could be used alongside AI to efficiently evaluate current ability and potential. This is an important inexpensive stage of the process as it can save money further down the line by avoiding sending scouts to watch players who are not right for a team. This can also help to identify players that are overlooked by other clubs and help find the best value players.

3. Once we have basic data on players, scouts are deployed who will gather more subjective information which may not show in the statistics. However, most teams have a limited number of scouts (N) and a limited scouting budget. Therefore, we must optimise this process so that the scouts time is not wasted and as many players are watched as possible.

4. The information that the scouts collect is collated alongside the datasets collected in process 2. Once all this information has been gathered, a team can use the statistics, scouts data and scouts opinions to rank the players they have watched. Using this teams can identify the players they would like to sign to improve their team and estimate the costs involved for transfer fees and/or wages.

5. Usually in a transfer or trade window teams will want to buy and sell multiple players to improve the squad. This presents a budget optimisation challenge as we want to purchase as many highly-rated players from the information gathering process who can positively impact the team. Therefore, the objective of this optimisation is to maximise the quality of the players that are purchased while staying within the constraints of the transfer/wage budgets. There are also other constraints set by the leagues such as squad sizes and wage caps. Finally, if a team is to sell their current players they can increase their transfer/wage budgets and create room for more new players. This is something that would need to be treated with caution though, as it could ruin the cohesion of the players within the team if we were to sell/buy too many players.

The processes we have discussed aim to improve the probability that a team will be successful in the transfer market and presents interesting computational challenges that are yet to be addressed by AI. The scouting process relates to AI literature which focuses on learning from imperfect classifiers. An example of this is shown in (Simpson et al. 2013) where human decisions, with prior knowledge about the ability of that human's decision making, are combined with Bayesian approaches to make decisions. This can be applied to scouting as we can use the teams' scouts opinions on players, with the knowledge of their prior scouting performance, alongside AI methods to rate players. The challenge of deploying the scouts relates to optimisation literature such as (Dang et al. 2006; Ramchurn, Polukarov, et al. 2010) as we aim to maximise the number of high-quality players the scouts assess while meeting the time and budget constraints. The transfer budget optimisation problem discussed in process 5 also relates to this literature as we are aiming to maximise the quality of players that are bought within the transfer and wage budgets where we can also sell current players to increase budgets.

Boon and Sierksma (2003) discuss the scouting of new team members to fill open positions and enhance the quality of teams. They calculate the potential value that new players in a team would have, focusing specifically on football. Their model uses linear programming to form an optimal team based on the quality of the players and their positional weightings that they calculate. Once an optimal team is formed they can use this for scouting purposes. Using a database of scouted players, players can be substituted into the team to calculate the effects that this would have and what value would be bought into the team. This model could be improved by taking into account the multiple positions that players can play in and the different roles players can take in different positions (e.g., a central midfielder could be a defensive player and sit deeper or could be more attacking to push further forward). This section mainly focuses on how scouted players will impact a team rather than looking to identify players that could be scouted and finding players that may have been overlooked by other teams.

In the next sub-section, we explore how teams train youth players in their academies which is another route that teams can take to improve their squad and bring in new players.

2.2.2 Training and Developing Youth Players

Young players can be trained by professional teams from ages as young as six.[15] Thus, teams can play a huge part in how they develop players and how they bring these players into the first team squad once they are old/good enough. The process of bringing players through youth systems can be fine-tuned and optimised at many stages. This can involve making sure that their training is tuned to improve their skills efficiently and ensuring that they are given the right amount of experience at the right times either in the first teams or by being sent out on loan to smaller clubs. The challenge of personalising the training regime of youth players, therefore, involves a number of prediction and optimisation problems that could be addressed by AI techniques. This is particularly so when such training regimes need to cope with significant degrees of uncertainties in player performance (e.g., injuries, variability in mood, or weather conditions). Many studies have explored the effects of injuries to youth players. Price et al. (2004) highlights the nature and severity of injuries that occur at the academy level and (Gall et al. 2010) evaluates the fitness characteristics of young players in youth academies, highlighting which of these characteristics improve players chances of proceeding to higher levels.

De Silva et al. (2018) have also used the player tracking data that is available as a tool for training youth players and for physical performance management in football. They tested their work in a professional Premier League football academy. This research uses standard statistical analysis to compare the activity demands in key playing positions, such as Central Midfielders and Centre Forwards. This study helps to provide insights from an elite performance environment regarding the relationship between player activity levels during training and matches and how they vary by playing position. This is an example of where machine learning based analytics could be used by a leading professional club to extend their knowledge and make changes to some of their training practices.

Finally, (Fister et al. 2015) outlines the challenges for computational intelligence in sport. The authors discuss the problems and current work that exist in sports (not just team sports) domain and in particular training for athletes. They open up a number of research questions in the area of training for sports and showed a necessity for developing an artificial personal trainer to optimise sessions. They also outline the process of sports training, showing the key components and a programming model. The paper mainly

[15] American youth players come through a college and draft system rather than individual teams having youth teams.

focuses on training that is not specific to any sport or skill such as for strength and power. However, it is still a useful tool for us to identify the stages in the team sports training process that can be optimised using artificial intelligence.

Next, we turn our focus to the team selection problem where managers/coaches select the players to play in games.

2.2.3 Team Selection

Team selection is a key tactical decision in team sports which has to factor in many uncertainties. In essence, the challenge involves picking a set of players to play in a game, which will maximise the chances of winning. There are many different combinations of possible team selections which are different for each sport. For example in football there is a squad of 25 players and need to select a team of 11, therefore there are 4457400 different possible team line-ups. This is calculated using $\binom{n}{r}$ where n is the number of players and r is the size of the team.

The team selection process also involves thinking about developing younger players. This is a balancing act between selecting a team that will win against thinking about using youth players. In most cases these players are bought into games as substitutes or are selected to be used in less important games such as pre-season friendlies or cup games.

In American football, cricket and baseball it is generally easier (compared to football and basketball) to identify which players have been performing well and therefore the challenge of finding a team that maximises the chances of winning is slightly easier. This is due to the statistical element of the scoring to see the contribution (e.g., yards gained, runs scored etc). That said, there is a lack of academic work which has focused on solving this problem. In football and basketball, it can typically be a challenge to attribute each player's contribution to a team. In these sports, there are a number of other factors that make a good performance other than just scoring or creating goals.

Deep learning has also been applied to model the behaviours of players in both Basketball and Football (Le et al. 2017; Seidl et al. 2018). Here, deep imitation learning has been used to "ghost" teams so that a team can compare the movements of its players to the league average or the top teams in the league. A simulation is run to see how an AI team would move in certain situations with the AI team created by "ghosting" the characteristics of average and top teams. This helps to identify where teams can make changes to their players' movements and change events to improve the probability of scoring a basket/goal or reduce the probability of conceding. Le et al. (2013) is also an example of multi-agent approaches to imitate and learn the movements of players in a game of football. The authors show that having a coordination model for the roles of players gives substantially improved imitation in comparison to conventional baselines.

Other factors that may need to be considered in this area involve predicting what an opposition will do: their line-up, their formation, their set-pieces, what style they will play, what areas of the pitch they target, where a player will aim a penalty and many more. An example of work that forms teams based on an opposition is shown in (Jayantha et al. 2018) where the authors create a team recommendation system for cricket teams which is based on selecting players who increase the probability of the team winning. This is also explored in more detail in (Gürpınar-Morgan et al. 2020) where the authors present a model to predict shot type using deep learning.

The team selection problem in sport relates to team formation literature in the multi-agents domain such as (Chalkiadakis and Boutilier 2012) which proposes new methods for coalition team formation. Coalition formation is the analysis of one or more groups of agents, called coalitions, that together jointly determine their actions. They integrate decision making during repeated coalition formation under type uncertainty using Bayesian reinforcement learning techniques. Matthews, Ramchurn and Chalkiadakis (2012) form optimal teams for fantasy sports games under the constraints that the fantasy sports problem presents. They do this by predicting the performance of football players (in terms of how many fantasy points they will score) and then form a team that maximises the number of expected points. This could be extended to aid team selection for sports teams and improve teams chances of winning. Vilar et al. (2013) discuss the complex social systems that are presented by team sports. The authors focus on the pattern-forming dynamics that emerge from collective offensive and defensive behaviours. They evaluate the differences in strategies and formations of two teams in a single game of football to understand the successful and unsuccessful relationships in the teams. This type of study provides significant results to demonstrate how complex systems analysis can help to better understand performance in football, by assessing team behaviour as a collective rather than individually.

One new approach to team formation in sports has been shown in the work in this thesis (Chapter 4). Work published in (Beal, Changder, et al. 2020) aims to value the inter-agent teamwork between players in sports teams. This teamwork value can then be used to form the optimal team of players that will work best together. The paper introduces a novel team formation method that considers the *interactional alignment* which allows us to form a team based on agent pairs that not only work well as a pair but overlap to create many other strong pairs in a team. Work in (Bransen and Van Haaren 2020) also explores the notion of teamwork in football and presents a new method to value teamwork between players based on the "Expected Threat" (xT) model (discussed in the next subsection).

There have also been Game Theoretic approaches to optimising teams of agents in other domains. These approaches have shown success in real-world applications. An example of this is shown for Stackelberg Security Games (SSG), the success of SSG is discussed in (A. Sinha et al. 2018). In an SSG a defender must defend a set of targets using a number of resources, whereas the attacker is able to learn the defender's strategy and attack after planning. Fang, Stone and Tambe (2015) use game theory and the application of an SSG to optimise the protection of endangered animals and fish stocks.

An important factor to be considered in the team selection process is to ensure that the players selected in the team are right for the team tactics. The approaches to tactical decision making, made by the manager/coach, are discussed in the next section.

2.2.4 In-Game Tactics

The in-game tactics used by teams to enhance their chances of winning games vary a lot from sport to sport. When creating tactics many factors must be considered such as the opposition team and their weaknesses as well as the ability of the players available. Getting tactics right can give teams a huge advantage and can allow weaker teams to win games that they are not expected to. In football, tactics cover the formation that the team will use, the "style" that they play in, set-piece selection and many more. American Football tactics cover the plays that are selected by the coaches and coordinators.

There have been a number of studies that aim to better understand the tactics in sports. Fernandez and Bornn (2019) provide a model to assess the expected ball possession that each team should have in a game of Football. The expected possession value (EPV) assigns a point value to every tactical option available to a player at each moment of a possession, allowing analysts to evaluate each decision that a player makes. In this paper, machine learning is used to estimate the parameters which are used in the model, such as pass and turnover probabilities which are estimated using logistic regression. A similar model is also applied to Basketball (Cervone et al. 2014) where points are predicted and player decisions are valued. Similarly, (Yue et al. 2014) focus on play prediction in Basketball developing models for anticipating near-future events given the current game state. These models are validated using 2012/13 NBA data and show that their model can make accurate in-game predictions. Building on this, (Zheng, Yue, and Hobbs 2016a) studies the problem of modelling spatiotemporal trajectories of the players using expert demonstrations. In particular, they look to see how a basketball player makes decisions with long term goals in mind, such as moving around opposition players or scoring points. They propose a model that uses both long-term and short-term goals and instantiates this as a hierarchical neural network trained using a large dataset of tracking data from professional basketball games. They show that this model generates more realistic trajectories compared to non-hierarchical baselines as judged by human expert sports analysts. This work could be improved by modelling the defensive team as well as the offensive team to give a more accurate simulation of the plays. Finally, for basketball, (K. Wang and Zemel 2016) focuses on offensive play call classification. This helps teams to understand the strategies of the opposition to influence the final match outcome. They apply variants of neural networks to SportVU[16] tracking data and find they are able to label play sequences quickly with high precision. Using a Recurrent Neural Network (RNN) they can achieve a precision of 90% and recall rate of 59% (when making predictions if the probability of the classifier is above 70%).

Other papers that focus on tactics in team sports include (Bojinov and Bornn 2016) which evaluates how in football a "pressing" tactic affects performance and disrupts the opposition's defences. By doing so, they are able to define and learn a spatial map of each team's defensive weaknesses and strengths which is useful for coaches when preparing to face an opposition. In a similar fashion, (Hobbs et al. 2018) aims to quantify the value of transitions[17] in a game of football. They aim to explore how teams create goal-scoring opportunities based on their transitions and find that if a team counter-attacks immediately rather than looking to maintain possession, the chances of scoring rises by 4.4% and chances of having a shot rises by 24.4%.[18]

Power, Hobbs et al. (2018) focus on set-pieces in football (e.g., corners, free-kicks, penalties). They discuss a number of "myths" regarding set-pieces and then prove/disprove these myths. For example, they show that a team is more likely to score from set-piece than in normal possession (1.8% chance of scoring from set-pieces vs 1.1% in open play). They also find that the type of delivery and the defensive set-up of the oppositions can significantly affect the chances of scoring. Work in (Shaw and Gopaladesikan 2020) also explores corners in football and aims to identify the best strategy for these. Finally, (Lucey et al. 2012) models team behaviours in football using entropy maps, created from team ball movements, which give a measure of predictability of team behaviours across the field. This provides a useful tool for coaches and decision-makers to be able to analyse opposition teams.

We have highlighted the studies that focus on the tactics within team sports. These approaches aim to decompose and break down how teams play which give more interesting insights for coaches. This can be useful for teams when setting up their own tactics to maximise their chances of winning against another team. The concept of using game theory to optimise tactical decision making is explored further in this thesis (Chapter 3) as well as in (Beal, Chalkiadakis, et al. 2020; Beal, Chalkiadakis, et al. 2021). In the next sub-section, we explore work that focuses on individual players performances.

[16] https://www.stats.com/sportvu-basketball.

[17] overloading the opposition when they have just lost the ball.

[18] This study was run over the 2016/17 EPL season.

2.2.5 In-Game Player Performance

Measuring player performance is an important factor in the decision making processes in team sport. This helps to provide the feedback to identify when changes need to be made to team line-ups, tactics and when transfers need to be made. A number of papers focus on ways to measure performance objectively with data. Whitaker, Silva and Edwards (2017) use a Bayesian approach to determine the abilities of players using different event types. They implement a Poisson model for event types and can then infer player abilities from this. These inferences allow English Premier League players to be ranked and for differences between players to be visualised. Power et al. (2016) focus on measuring the risk and reward of passes in a game of football. This gives new methods to evaluate player passing performance and identify key players in a team who execute the key passes consistently.

More recently, there has been a rise in models which aim to extract the value of actions that players make within sports games. The first of these is "Expected Goals" (xG) in football (Spearman 2018) which aims to add a probability of scoring from a shot using models trained from thousands of shots in historic data which are trained based on shot location and the context of the opposition players.[19] Other models have built on this concept to value actions away from shots, these include the "Expected Threat" (xT)[20] and "VAEP" models (Decroos et al. 2019). These aim to assign values to passes or dribbles based on the contribution to increasing the likelihood of your team scoring or reducing the likelihood of the opposition scoring. These have also been extended in (Merhej et al. 2021) which aims to value defensive actions based on the predicted xT of what has been stopped. Similarly, (Stöckl et al. 2021) looks to better understand defenders performances using graph convolutional networks and (Llana, Madrero, and Fernández 2020) aims to create more off-ball metrics for exploiting an opponent's spatial weaknesses. Finally, (Van Roy et al. 2021) aims to evaluate the decisions made by players in games and if they were optimal.

Power, Cherukumudi, et al. (2019) focus on the performance of football goalkeepers. They simulate each goalkeepers performance when facing a number of example shots and compare which goalkeeper would concede the least number of goals. They do this by using a "spatial descriptor" for each goalkeeper which is made up of features such as clean-sheet percentage, win percentage and save percentage for in different thirds of the goal. This type of player performance modelling was also explored within baseball to evaluate how a batter will perform against certain pitchers (Alcorn 2018). As well as evaluating player performances, this type of analysis could be useful to simulate the

[19] xG models have begun to be widely accepted across football and are now features as a stat on BBC Match of the Day - https://www.bbc.co.uk/sport/football/41822455.

[20] https://karun.in/blog/expected-threat.html.

performances of new signings into a team. Also in Baseball, (Bouzarth et al. 2021) presents a mathematical approach to defensive positioning.

Turning to Basketball, (Felsen and Lucey 2017) evaluates NBA players' body pose and shooting styles to find any correlations between the player body shape and shooting success. They find statistically significant differences in distributions of attributes describing the style of movement of different phases of the shot. In American Football, (Burke 2019) uses Deep Learning to quantify Quarterback decision-making again allowing us to identify the NFL quarterbacks who have the best decision-making skills which is a vital part of this position in a game of American Football. Their model correctly identifies the targeted receiver in 60% of cross-validated cases. They find when passers target the predicted receiver, passes are completed 74% of the time, compared to 55% when the QB targets any other receiver. Their approach gives a new way for teams to quantitatively assess quarterback decision-making performance. Finally, Correia et al. (2011) assess players' decisions when making passes in Rugby Union based on the positions of oppositions and teammates.

The majority of the work that we have discussed in this section focuses on finding new insights into tactical analysis in sport. These studies help identify the strengths of different tactical processes and find new ways of evaluating player and team performances. There are good examples of work in football and basketball however not as many in American football or rugby where tactical decisions are also key to winning games. There still remains a number of areas where AI could impact tactical decision making. This work would mainly be focused on how individual agents (the players) perform in different teams, with different tactics and how much impact they have on the game outcomes. This type of analysis could benefit all of the processes that we showed in Figure 3 as AI could improve player transfers, match preparation and help to gain better feedback from the outcomes of games.

Next, we turn our focus to fantasy sports games and the computational challenges that these present.

2.3 Fantasy Sports Games

In America alone, an estimated 32 million people take part in fantasy NFL games (American Football) with an average spend of \$467 per person, per season totalling to around \$15billion across the season and in the UK over 5 million people take part in the Fantasy Premier League for football.[21] There are fantasy sports games for nearly every professional team sport and there are many different sites and leagues ranging from competitions with millions of competitors to small leagues run between friends.

[21] http://www.forbes.com/the-70-billion-fantasy-football-market.

In fantasy sports games competitors select a team of real-life players, who are assigned a value/salary, within a given budget. Dependent on how well the players perform in real life they are given corresponding fantasy points (e.g., for a goal/assist in football or a touchdown in American football). The aim of the game is to maximise how many points the selected team can obtain under the constraints of the fantasy game. Figure 2.3 shows the process of fantasy sports games. Initial values, based on knowledge of the players' ability, are set for the players before the season starts and the fantasy competitors select an initial team. This team of players is then awarded points each game-week based on their real-world performance (if the player does not play they receive no points). The values of the players can also be updated throughout the season so that the players who have performed better than expected will then cost the reflective amount. The fantasy league standings are updated each week and once all the N game-weeks have been completed, prizes are awarded based on the standings.

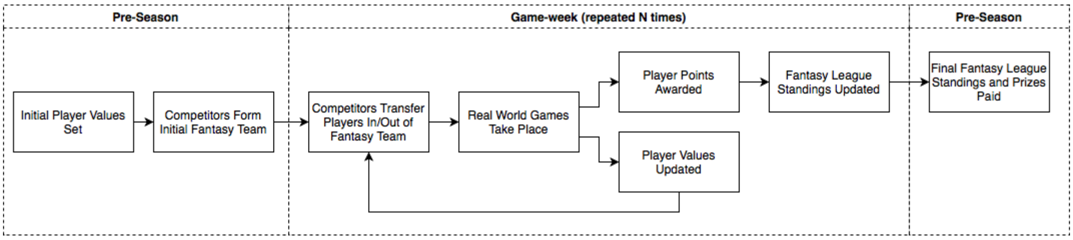


Figure 2.3: The Fantasy Sports Game Process.

Traditionally people compete in a league across the season, with the aim to accumulate the most points whilst having restrictions on the number of changes/transfers that can be made to a team. In these games the "game-week" processes shown in Figure 2.3 are repeated for every set of matches (usually once a week). Leagues such as these have been running for many years. A good example of this is the Fantasy Premier League (FPL) for Football in the English Premier League, which is extremely popular in the UK and worldwide. In the FPL there are 38 game-weeks therefore $N = 38$. This means that in fantasy games such as this the transfer stage is key to success as, dependent on the fantasy league, the number of transfers is limited (e.g, in the FPL there is a limit of 1 transfer per game-week). Due to the rules on transfers, when selecting the initial team and making changes it is important to consider the players' future performances as well as just the next game-week.

More recently Daily Fantasy Sports (DFS) sites, such as FanDuel and DraftKings, have seen a large growth in popularity. These sites offer leagues that only run for one game-week rather than across the whole season, meaning that a new team is selected each week instead of making transfers. In these types of fantasy games new teams are formed every week from scratch, therefore in Figure 2.3 the value for $N = 1$. This means that only an initial team is set and only one future game needs to be considered when predicting the players' performance.

Although there has been significant growth in fantasy sports, there is a lack of research focus on ways that AI could be used to improve competitors performances or using AI automated teams to compete against humans. There are a small number of studies in fantasy sports. The seminal work in this area (Matthews, Ramchurn, and Chalkiadakis 2012), provides the first real-world benchmark for sequentially optimal team formation in this domain. More recently, (Landers and Duperrouzel 2018; Beal, T. Norman, and Ramchurn 2020b) use machine learning techniques for NFL fantasy leagues. When forming teams to enter into fantasy sports leagues, two key computational challenges are presented. These are:

- **Player Performance Prediction:** Predicting how well a player will perform in the real-world and therefore the number of points that a player will obtain both in a single game and over a given time period.
- **Team Formation Optimisation:** Selecting an optimal team using the performance predictions so that the constraints of the fantasy league are met. This also includes the challenge of making effective transfers in longer running fantasy leagues.

As well as using AI to form teams and aid competitors fantasy performances there are other challenges highlighted in Figure 2.3 that can be addressed using AI. Examples of these challenges are: player price forecasts, opponent modelling after every match (as competitors are able to see other competitors teams), draft strategies and betting strategies to maximise the chances of winning cash in DFS fantasy games.

2.3.1 Player Performance Prediction

Predicting the player performance is key to selecting which players are worth having in a fantasy team. If a player (and the team he plays for) performs well, more points will be accumulated for the fantasy team. There are a number of factors that need to be considered when predicting how well an individual is likely to perform and this varies significantly based on the sport being played and the position that the player plays in. For example, in football when a striker's performance is predicted, the aim is to predict the number of goals that he/she may score whereas for a defender we focus on predicting if his/her team will keep a clean sheet (concede no goals). Also, it must be considered how likely a player is to play in a given game, as players may not play due to injuries and tactical decisions. If a player does not play, they receive no fantasy points. These examples show a small number of the many uncertainties that need to be considered when making predictions on future performances.

When predicting player points, a feature set for each player (in a given game week) is taken as an input, this is referred to as X where n is the number of features (shown in Equation 2.2). Example features for an American Football player would be yards gained per game, number of touchdowns scored and games played. In football, example features would be goals scored, number of assists, number of clean sheets and minutes played. The feature set could also be made up of previous fantasy points scored from a number of prior weeks. Once a feature set is formed, a target vector Y represents the points scored by the player in the given game week, corresponding to the feature data in X. Equation 2.2 shows X and Y for n features and i players where $x_{i,n}$ is the n^{th} feature for the i^{th} player and $y_{i,w}$ is the i^{th} player's points in game-week w. Next, a machine learning algorithm can be trained using X and Y to produce a function ϕ, that can output the prediction of Y using the features in X. A row of X, X_j can be used with this to make a prediction on the corresponding row in Y. This gives $\phi(X_j) = Y_j$. Different ML methods can be tested across the different sports.

$$X = \begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \dots & x_{1,n} \\ x_{2,1} & x_{2,2} & x_{2,3} & \dots & x_{2,n} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{i,1} & x_{i,2} & x_{i,3} & \dots & x_{i,n} \end{pmatrix} Y = \begin{pmatrix} y_{1,w} \\ y_{2,w} \\ \vdots \\ y_{i,w} \end{pmatrix} \tag{2.2}$$

When making player performance predictions in a daily fantasy contest we can be much more precise as the predictions only need to focus on how that player will perform in the specific game. Whereas, in more traditional leagues future performances must be considered as well as just the next game. Thus, a player who will perform well in multiple future games will be selected and not just one who will perform well in a single game (which may be against a poor opposition). This is considered by the model shown in (Matthews, Ramchurn, and Chalkiadakis 2012) where predictions for a players performance are made for a number of future game-weeks.

Matthews, Ramchurn and Chalkiadakis (2012) predict a player's performance and the number of points that a player will score based on the prediction of a given game using the (Dixon and Robinson 1998) framework. The authors choose to use this approach as the Dixon and Robinson model treat football as a dynamic situation-dependent process as well as its proven success in the football betting domain. Dixon and Robinson's score predictions work by taking each club's attacking and defending efficiencies from past results and then using these to derive the probabilities of the side scoring different numbers of goals in a given match. Using the match outcome prediction, Matthews attributes the probability of a player scoring points based on the 4 most significant point-scoring categories in the FPL.

1. Player appearance in a game (worth 2 points per player playing over 60 minutes and 1 point for any player playing under 60 minutes). For this a three-state categorical distribution is used, the states are: starting, substitute or unused. Three probabilities are computed for each category and the highest probability is assigned to the player to obtain the predicted points.

2. Clean sheet[22] (worth 4 points to a defender/goalkeeper and 1 point to a midfielder). For clean sheets, the probability of a team not conceding can be calculated using the scoreline distributions from the Dixon and Robinson model and points can be predicted based on these probabilities.

3. Goals scored (worth 6 points for a defender/goalkeeper, 5 points for a midfielder and 4 points for a striker). This is calculated using a Bernoulli distribution (Gelman et al. 2004) or a Binomial distribution over a single trial, describing a players probability of scoring the goal given he was playing at that time.

4. Goals created (worth 3 points per player). Another Bernoulli distribution is used for this, again describing the players' probability of creating the goal given he was playing at the time.

Turning to American Football, the NFL study (Landers and Duperrouzel 2018) for player point prediction starts by engineering features from FanDuel that their model uses. In their paper they test two different feature sets (FD_1 and FD_2). They also test two different methods and compare using the different feature sets. The methods they test are a least-squares with averaged perceptron (Lehtokangas et al. 1995) and a boosted decision tree (Schapire 2003). These methods are evaluated by using the coefficient of determination to measure the accuracies (R^2). The boosted decision tree achieves average accuracy of 0.417 using FD_1 and 0.250 using FD_2 while the least-squares approach achieves 0.401 on FD_1 and 0.146 on FD_2. Other related work has shown similar results to those in (Sugar and Swenson 2015). It is noted that different methods could be tested for different positions in American Football. This is because the roles of different positions differ more than in other sports (e.g, Quarterback vs Running Back).

The current studies for player predictions show applications of machine learning techniques providing good benchmarks in both football and American Football. Further techniques could be tested in these sports to improve on the current benchmarks. These could also be tested with new feature sets such as time-series data which would analyse the form of the players over a given number of game-weeks. Matthews, Ramchurn and Chalkiadakis (2012) focus on the FPL fantasy game, it may be useful to test this work on a DFS site to see if it also performs well in these fantasy games. Landers and

[22] When a team concedes no goals.

Duperrouzel (2018) only focus on DFS so it would also be interesting to see how their predictions would work in a more traditional multiple-week league. There also remain gaps in the literature for predicting performances of players in other sports where fantasy games are popular such as Basketball and Baseball.

The player score predictions become more valuable when combined with a good team optimiser, allowing the maximum points to be accumulated each week. The methods for team formation that currently exist in the literature are outlined in the next section.

2.3.2 Team Formation Optimisation

Selecting the best team of players within the given constraints of the fantasy site is vitally important, as shown in Figure 2.3 this process is repeated sequentially across a season with constraints for changes that can be made to the team. This process involves choosing players with different abilities, prices and risks that need to be considered. As discussed in the previous section we may also need to consider the future performances of these players over a number of given game-weeks. In formation problems, there are some actors, with their own abilities and characteristics that have formed a team to perform a task to achieve a common objective in a real-world domain. Within the fantasy sports team formation problem, there is a set of players (the actors), with their own abilities (predicted points) and we are looking to maximise the number of these points (the common objective). Therefore, it is key to get the best players possible into a high performing fantasy team. The fantasy problem relates to other works in the AI community. In particular, (Dang et al. 2006) focuses on choosing the best sensors to surveil an area, (Ramchurn, Polukarov, et al. 2010) focuses on dispatching optimal teams of emergency responders and (Chalkiadakis and Boutilier 2012) looks at the appropriate set of agents to work within a coalition formation problem.

In fantasy sports, the main aim is to select a number of players, all of whom are assigned a value, in different positions within a given budget. For example in the FPL, there are around 500 players available to select from, all with a position and a given value and 15 need to be selected (2 goalkeepers, 5 defenders, 5 midfielders, 3 forwards). The total value of the selected team most not exceed £100m. Eleven of those fifteen players are put up as your 'starting 11' (in a selected formation) and they will be the ones who will earn your fantasy points. The remaining players are 'subs' and will automatically come into the starting 11 if one of the current players does not play.[23] This means that the players are selectable in over 10^{25} ways. One team is selected at the start of the season and then one transfer is permitted per game-week. In DFS competitions a fresh team is created each week, therefore the optimisation is simpler. Figure 2.4 shows example team formation set-ups from FPL and DraftKings.

[23] Rules: https://fantasy.premierleague.com/a/help.

FIGURE 2.4: Example Fantasy Team Set Ups.

(a) FPL (b) DraftKings

Formally, a generic fantasy sports team involves selecting a team of players to fill N number of positions places within a team (e.g., a fantasy NFL team requires 10 players). There may be given positional constraints depending on the sport in questions (e.g., 1 goalkeeper, 3-5 defenders, 3-5 midfielders and 1-3 strikers to fill 11 spots in a football fantasy team). The fantasy sports optimisation problem is defined in Equation 2.3 (excluding other different constraints that the different leagues have in place).

$$\begin{aligned} & argmax(\sum_{n=1}^{N} points_n) \\ \text{s.t.} \quad & \sum_{n=1}^{N} selected_n = teamSize \\ & \sum_{n=1}^{N} value_n \leq budget \end{aligned} \tag{2.3}$$

Where n represents the ID of a player, which is used to identify if that player is selected ($selected_n \in \{0, 1\}$) and what the player value is ($value_n \in \{\mathbb{Z}\}$). Our objective is to maximise the total number of points ($points_n \in \{\mathbb{R}\}$) while staying within the given team size and ensuring that the combined salaries/values of the players are below the given budget constraint.

Matthews, Ramchurn and Chalkiadakis (2012) approach the FPL team optimisation as a sequential team formation problem which is formalised as a Markov decision process (MDP). The model in this paper considers the limited transfers that could be made meaning that the future performances of players are considered when making changes. A reinforcement learning approach is used, working under uncertainty regarding the underlying MDP dynamics. In an MDP the agent (in this case the fantasy competitor) and the environment (fantasy game) interact continually, the agent selecting actions (by selecting a team each game-week) and the environment responding to these actions (in the form of points) and presenting new situations to the agent. Using an MDP means that the model is able to assess the possible rewards that come from the possible actions that are made (in FPL this would be the action of making a transfer). The problem can then be treated as an optimisation problem. Matthews et al., set up the problem as a Multi-Dimensional Knapsack Packing (MKP) (Korte and Vygen 2012) and solve the optimisation using IBM ILOG's CPLEX 12.3. A number of different setups of the MKP were tested. These include Q-Learning and Bayesian Q-Learning with different parameters. By doing this Matthews was able to obtain a model that ranked in the top percentile and give a benchmark for how AI and machine learning models can perform in the FPL.

When forming fantasy teams for the NFL in DFS competitions, Landers and Duperrouzel (2018) compare four different approaches for team optimisation, these are:

1. Selecting a completely random team to loosely model the behaviour of a human with no knowledge of the sport.
2. Selecting a random team from a filtered dataset with just the higher-performing players. This is to loosely mimic the behaviour of a player with general knowledge about the sport.
3. A filtered optimised approach picks a team from the filtered set and uses a brute force algorithm to select the best team based on maximising the predicted points, that fits the constraints.
4. Filtered actual best selects the best possible team from the filtered data set to give a view of what the best performing team would be.

Using the player points prediction models discussed in the last section, 100 teams are selected using each of the above methods for each prediction method. The authors are able to evaluate the performances using the following metrics: The average points return of the 100 teams, the maximum points return and the percentage of teams that produce a profit.[24] This is compared against the work in (Sugar and Swenson 2015) which achieves a success rate[25] of 71.4% whereas (Landers and Duperrouzel 2018) achieve a success rate of 82% for weeks 3-9 in the 2016 season. However, this would need more testing due to the small sample size (7 game weeks) meaning the results are not statistically significant.

[24]The amount of points needed to break even is set to 111.21.
[25]Success rate = number of weeks that the model would earn a profit.

Matthews, Ramchurn and Chalkiadakis (2012) were able to achieve good results by solving their optimisation problem as an MKP. As a brute force method for team optimisation is used on a filtered dataset in (ibid.) there may be more efficient ways for this to be done, such as a similar approach using CPLEX in (Matthews, Ramchurn, and Chalkiadakis 2012). Using this type of approach would ensure an optimal team is selected and improve the run-time efficiency in comparison to a brute force approach. This is explored in (Beal, T. Norman, and Ramchurn 2020b). Another area to explore would be to model the uncertainties in the players' performances rather than predicting their points. This would allow teams with different levels of risk and reward to be selected which may be useful in a DFS competition multiple teams can be entered. This could be achieved by using a stochastic optimisation approach (Ermoliev and Wets 1988). There also remain gaps in the fantasy sports process where AI could be used to forecast the player prices, assess opponents fantasy teams and create AI betting strategies to maximise the chances of generating profits in DFS fantasy games. DFS strategies are discussed in (Haugh and Singal 2018) where a portfolio of teams is used to maximise the chances of generating profits.

In the next sub-section, we will discuss the impact that injuries have on sports teams and how AI can be used effectively in this domain.

2.4 Injury Prediction and Prevention

Contact team sports have a high risk of injury (Drawer and Fuller 2002b). If a team is missing their star players the probability of winning matches decreases significantly. Moreover, it is not financially beneficial to pay large wages to players who are unable to play. The economic impact of injuries is highlighted in the annual report by JLT[3] who evaluate the injuries that occur in the Premier League every season. They found in the 2016-17 season over £175million of wages were paid to injured players. The impact in Football is also discussed in (Drawer and Fuller 2002a). Being able to predict when these injuries are likely to occur and change real-world variables to reduce the likelihood of the injury, presents an interesting computational problem. Lysens et al. (2012) discuss the predictability of sporting injuries.

Vast amounts of data is now collected in relation to individual players in both competitive matches and in training. All professional sports are now collecting this data in real-time for both competitive matches and training. Companies such as Catapult Sports and STATSports sell GPS trackers that sports teams in multiple sports across the world use to monitor their players. The type of features that are collected by these trackers include but are not limited to:

- Distance covered
- Meters per min
- Speeds reached
- # sprints
- Sprints distances
- Intensity
- Heart rates
- Impacts
- Stress load

This data alongside the historical medical data that is collected by physios and club doctors can give a feature set for players that have yet to be studied by the AI community. As well as collecting other features from the club doctors and physios, we can form a list of injuries that have occurred to the players. An example of a standard machine learning formulation for predicting injuries could involve training a model based on a list of injuries that have occurred. Using this as training data, we can extract a relevant feature set (from a given time frame) such as Acute:Chronic Workload and the number of competitive minutes played, alongside historic data such as previous injuries to that player. To avoid bias in injury models, we would also need to consider examples of where no injury occurred to ensure the model has awareness of different injury rates from player to player. Using this type of formulation, standard machine learning classification methods could be used to calculate the probabilities of a player getting injured given the feature data and identify what real-world changes can be made to reduce the chances of injury. Highly accurate models for this would provide huge real-world value to both the athlete/player who can extend his/her career if better protected and injuries are avoided, as well as the team who can plan more effectively so they do not lose top players to injuries.

There has been a number of studies in relation to injuries in sport that have mainly focused on the medical domain. Some relevant literature that supports the theory that AI could have a part to play in this domain are summarised in this section. There are a number of computational challenges that injury prediction presents for AI. The current literature is broken down into medical research and research that has been focused on the use of wearable sensors in sports teams.

2.4.1 Sports Medicine Research

There have been many studies in the Sports Medicine community focused on the causes of injuries in sports and studying a number of large datasets. Firstly, (Hägglund, Waldén, and Ekstrand 2006) evaluates how previous injury is a risk factor for future injuries at the top level of football. The study compares two seasons (2001-2002) worth of injury data from 12 elite Swedish male football teams. They use a multivariate model to determine the relationship between a previous injury and the risk it causes. They found players who were injured in the 2001 season had a greater risk of injury in the following season compared with non-injured players. Particularly, players with previous hamstring, groin, and knee joint injuries were two to three times more likely to suffer the same injury in the following season. This work was extended in (Hägglund, Waldén, and Ekstrand 2009; Hägglund, Waldén, and Ekstrand 2013), looking at 14 football teams across Europe between 2001-2012. This study focused on the injury characteristics and variation of injuries during a match, season and consecutive seasons over the time period discussed. It found that the rate of some injury types decreased over the last 11 years. However, training and match injury rates and the rates of muscle and severe injuries remain high. It was also found that the risk of injury increased with time in each half of matches. Clearly, such works point to patterns in datasets that could be used in machine learning algorithms to determine the risk of injury and to optimise the recovery process.

Fuller (2018) models the effect of the player workloads on the injuries in the English Premier League. The study shows how a team's injury burden varies from day to day during a season, based on a team's match and training schedules. It also compares successful and unsuccessful Premier League teams and how their training loads affect their number of injuries. They find that a successful team undertakes fewer training sessions each week so there are fewer opportunities in which to influence training activities and to reduce injuries. A similar study (Bowen et al. 2016), investigates the relationship between physical workload and injury risk in elite youth football players. They also found higher workloads were associated with greater injury risk. This highlights that workloads can be used as a metric in an AI model for injury prediction. Workloads have been seen as a significant factor in injury prediction, (Hullin et al. 2016) evaluates the Acute:Chronic Workload Ratio (ACWR) in Rugby League players and find that a greater ACWR increases injury risk. The ACWR is calculated by dividing the acute workload (fatigue) by the chronic workload (fitness). This is defined as:

$$ACWR_t = L_{t-7}/L_{t-28} \tag{2.4}$$

where L is the players load[26], t is the current day. Therefore, it is a ratio of the last seven days load to the last twenty-eight days load. This, along with other related variables, could be key in using machine learning to predict injury.

There are examples of injury studies in other sports such as Basketball and American football. In Basketball (Podlog et al. 2015) surveys the injuries in the National Basketball Association over a 25-year period and identifies the relationship of injuries to team performance. In American Football, Ward et al. (2018), finds that regardless of

[26]A common method for calculating workload is by multiplying the athletes' perceived exertion (sRPE) by session duration (e.g, if an athlete reports an sRPE of 5 and trained for 90 minutes, the athlete's workload for the day would be 450 arbitrary units (AU)).

the position, training days with high amounts of volume and intensity share an association with increased risk of injury while training days of a high amount of low-intensity training share a relationship with a decreased risk of injury.

In the next sub-section, we discuss the applications using the data from the wearable sensors that we discussed at the start of this section.

2.4.2 Wearable Sensor Research

Due to the amount of data that is collected from the wearable sensors, some sports teams have begun to look at ways that this can be used to benefit sportsmen and women. Firstly a study by Kelly et al. (2012) researched how wearable sensors can be used to automatically detect collisions in Rugby. As we discussed in Section 2.3 Rugby is a very high impact sport and tackling is the most common cause of injury in Rugby, and therefore having a way of automatically modelling tackles can improve the work of the medical staff. The work was compared against collisions which were manually labelled using data from elite club and international level players. The paper tested a number of different algorithms for this problem such as Support Vector Machines, Neural Networks and Convolutional Neural Networks (CNN). The results show the model is able to identify collisions to a high level of accuracy, achieving a recall and precision rating of 0.933 and 0.958, respectively using CNNs. The model can give coaching and medical staff tackle-specific measurements, in real-time, which can be used in injury prevention and rehabilitation strategies. Following on from this, (Cust et al. 2018) reviewed the ways that machine learning and AI can be used to classify certain movements in sport.

In the next section, we will discuss the findings from this chapter and highlight the computational challenges and open areas that exist for AI in the team sports domain that will be explored throughout the thesis.

2.5 Identifying Open Areas of Research

Here, we discuss our findings from the literature review and highlight the open areas of research that exists in this domain. These are organised into the key sections that we have explored in this chapter. We identify a number of novel research questions that are yet to be addressed by any academic literature. We then target the ones that will be focused on in the chapters of this thesis.

2.5.1 Match Outcome Prediction

As we highlighted in Section 2.1, there is a large amount of existing academic literature which focuses on statistical methods for predicting sporting outcomes. Relatively speaking there has been less work that applies machine learning, deep learning and AI approaches. The existing literature that we have evaluated appears to have reached a glass ceiling, this is highlighted in Table 2.2 where we show the prediction current top (published to our knowledge) match outcome accuracy within each individual sport.[27]

Table 2.2: Current Best Accuracy.

Sport	**Accuracy**	**Paper**
Football	56.7%	Baboota and Kaur (2019)
American Football	66.9%	Baker and McHale (2013)
Rugby League	67.5%	McCabe and Trevathan (2008)
Cricket	75.0%	Jayantha et al. (2018)
Basketball	72.2%	Z.Shi (2013)

Fewer studies focus their attention on solving the match outcome problem in American Football, Baseball and Rugby Union. Algorithms that have shown to be successful in other papers for predicting match outcomes and scorelines, could be applied to these sports to compare the results to the statistical approaches outlined in the current literature. Papers such as (Baboota and Kaur 2019) and (Hucaljuk and Rakipović 2011) apply and test a number of ML algorithms to solve the match outcome problem in football. The approach of applying, testing and comparing different ML techniques could also be applied to American Football, Rugby Union and Baseball to find which techniques are successful in these sports with the available datasets. In particular, neural networks have shown to be successful when applied to Rugby League in (McCabe and Trevathan 2008).

There are many factors that make it challenging to accurately predict sports match outcomes (e.g., uncertainties in team form, injuries to players, players mood, team morale, weather, playing conditions and changes in management). Finding ways to quantify how these uncertainties will impact a game is also a challenge. What makes sports games so entertaining is the element of randomness and luck that play a big part in the outcome. This is especially the case in a lower scoring game such as football. For example, the underdog team may get some luck in the form of a deflected goal or a refereeing decision going their way, this then allows them to pull off an upset. The majority of the current literature focuses on the final events of the game such as outcomes (win/loss/draw) and scorelines. In turn, there is less work that focuses on the more granular predictions which can aid the final outcome forecast. Examples of how more granular approaches can help are outlined as follows:

[27] Football focused on domestic league games.

- Modelling the problem using match processes to train a model that is able to recognise what outcome occurred when similar set-ups and conditions have been observed before. Deep learning techniques could be used for this approach, specifically Long Short Term Memory known as LSTMs (Hochreiter and Schmidhuber 1997), a type of recurrent neural networks which have shown to be successful when using historic datasets. Due to the vast amounts of historic datasets that are available for team sports, this approach may prove to be successful.

- Create models based on attack and defence player movements in order to gain a better understanding of their performances when they are facing each other in a game. This may improve the current work as some teams attacks may be more effective against some types of defences than others (e.g., a blitz defence may be more successful against a running attack in American Football). Applying multi-agent systems and coalition formation techniques to imitate the player movements and simulate their performances is one example of how to achieve this.

- Basing predictions on tactical set-ups of the teams. This approach would factor in the formation of the two teams, the style of plays and teams set pieces (e.g., in football, a team set up in a 4-4-2 formation may have an advantage against a team in a 5-3-2 formation). For this, an application of Game Theory approaches could be used to assess different teams tactics against other teams (e.g., Stackelberg games).

- Analyse the players' personality, moods, and mental state, in the build-up to games when making match outcome predictions. This would allow models to consider how players perform in certain games. For example, certain players may perform better in the finals of a competition or in fighting against relegation to a lower league, while other players may "choke" and their performance will deteriorate in these types of game (Beilock and Gray 2007). Natural Language Processing techniques (Sebastiani 2002; Manning and Schütze 1999; Collobert and Weston 2008) could thus be used to monitor pre/post-match press conferences and interviews that players give. This could also be used to analyse the managers' confidence levels in their scheduled pre-match interviews. Another approach could be to monitor the players' social media accounts by using sentiment analysis on their online public posts (Bo Pang, Lillian Lee, et al. 2008). There exist numerous approaches that aim to leverage social media by means of natural language processing (Farzindar and Inkpen 2015; Pak and Paroubek 2010). However, some issues that arrive from this type of analysis such as: not all players show up in press conferences, press conferences are often pre-prepared and players' social media accounts are often controlled by media managers/agents.

Stekler, Sendor and Verlander (2010) explore a number of problems with prediction models in team sports and test current models against alternative forecasting methods (such as experts and the betting market) as well as examining existing biases in the models. Furthermore, (S. Ganguly and Frank 2018) focus their attention on problems when predicting teams' win probabilities (many of the works we have evaluated use win probabilities). Firstly, the authors suggest that win probabilities lack sufficient context, and the models should respond to in-game factors such as injuries. Next, it is suggested that the current win probability models should incorporate a level of uncertainty due to the many possible events and scorelines that make up the match outcome. Finally, they explore the "what-ifs" in sports games and use these to model the "alternative outcomes". For example, what if a teams star player gets injured in the first half of a game - how would this affect the outcome? Modelling match outcome problems in this way allows for the discussed issues to be addressed and could lead to more accurate predictions.

2.5.2 Strategic and Tactical Decision Making

There are many key decisions and complex tasks in team sport, these decisions must consider the many existing uncertainties. Figure 1.1 outlines a number of decision-making processes, how they link together and how they are evaluated through feedback. This process diagram gives a framework for the areas where future work to improve the decision making in team sports can occur. The first of these processes is the recruitment of players through transfers and trading. Clubs make large investments into players who they believe will improve their teams. However, an issue with this is that when a new player signs for a team there is no guarantee he/she will perform to the levels expected for the cost. This presents us with the challenge of being able to use AI to successfully predict what impact new players will have on the team. This could save teams millions on transfer fees and wages for players who fail to live up to expectations as well as helping clubs to identify undervalued players. We also explored how AI could be used to improve and aid the scouting process, where information is gathered about players. Clubs have limited resources both for scouts time and money, therefore this presents an optimisation problem where AI techniques could be used to improve this process for teams. We can use techniques that learn from imperfect classifiers to improve scouting, such as those shown in (Simpson et al. 2013). These techniques provide a platform to combine humans scouting ability with machines to provide more accurate reports on players.

Some questions remain in team selection and tactical decision-making. Some more recent work has explored how individuals impact a team in papers such as (Alcorn 2018; Power, Cherukumudi, et al. 2019). In these papers, players performances are simulated in different scenarios to give a comparison. This type of analysis could be built on in all sports to identify the impacts that the individuals have within the complex team systems. Coaches could use this to identify how changes in their style of play, with their current set of players, would affect the performance of the team. Simulations could also be used for player recruitment as potential new players could be simulated to show how they would perform in a new team.

Applications of Game Theoretic approaches such as Stackelberg Games could be used for tactics in team sports, to create strategies and tactics that maximise the rewards in a game and improve the chances of exploiting weaknesses in opponents teams. Finally, AI techniques could be used to optimise the training process to help players reach their full potentials. This is especially the case for youth players who may be at a club from an early age and where there are a number of decisions that need to be made to help them (e.g., should he/she be sent out on loan and should he/she play in the first team). Using AI to support these decisions could ensure that the best development path is selected to find the optimal results of that individual player.

The decision-making processes within team sports present interesting computational challenges for AI due to the number of existing uncertainties at every stage. These challenges can help us to understand and model the complex systems within team sports and apply novel techniques using AI.

2.5.3 Fantasy Sports

To be successful in fantasy sports there are many real-world and in-game factors that must be considered. These include: team formation and player point prediction, changing strategies based on other competitors teams, predicting the values of players so the total value of the fantasy team is maximised, and modelling the risk of the fantasy team so that profit-making strategies can be created for DFS.

Matthews, Ramchurn and Chalkiadakis (2012) and (Landers and Duperrouzel 2018) are two examples of the small amount of academic literature which use ML techniques and team formation optimisation in fantasy sports. The team formation methods that are used in (Matthews, Ramchurn, and Chalkiadakis 2012) can ensure that optimal teams are formed based on their player predictions whereas, in (Landers and Duperrouzel 2018) a brute force method is used for team formation. Therefore, the application of the techniques used by Matthews and his co-authors for team formation would provide value in other fantasy competitions (especially DFS where the team formation is vital for the single game-week). This could have a big impact on fantasy sports in American Football due to the size of the market and the amount of extra performance data that is available. As well as applying new team formation methods, other machine learning feature sets and techniques could also give new results and comparisons. For example, using time-series analysis on prior game performances would be a novel feature-set to test which may give an accurate representation of the players form and give more accurate points predictions.

As well as the methods that we have described for player points prediction and team formation optimisation, there are a number of other open areas within fantasy sport. Firstly, creating and running fantasy teams presents human-machine interaction challenges where a mixture of human knowledge and AI-based optimisation could be used to improve human performance. For example, it may be worth studying the effects of using an AI-based team formation approach with human player performance predictions and vice-versa. This would allow us to evaluate what parts of fantasy sports games are skill and what is down to chance and luck. This is especially important in the USA due to restrictions on gambling (Boswell 2008).

Finally, in DFS competitions, multiple teams can be entered into the same league by a single competitor. Meaning that a portfolio of teams can be entered to maximise the chances of winning. By modelling the risk and uncertainties of players points different teams can be formed at different levels of risk, meaning that the fantasy team should contain "higher risk" players that other competitors may have been overlooked by other players in that league. By entering a portfolio of teams at different risks, competitors can create strategies that maximise their chances of generating a profit while avoiding selecting teams that other competitors in the league may have picked. This type of strategy is discussed in (Haugh and Singal 2018) and could be supported by AI work using a stochastic optimisation approach (Ermoliev and Wets 1988) and ideas from portfolio risk management strategies (Dunis et al. 2016).

2.5.4 Injury Prediction

Sports injury prediction is an area where almost no literature exists in the AI community. Therefore, further work should be focused on applying machine and deep learning techniques to the data that can be obtained from sports teams. The developed models should highlight players who are likely to get injured in a given training session or a competitive match. To build these models it would be important to use the expertise of the doctors and physios at the clubs to identify what new approaches and features could be used when predicting injury as well as collecting as much data as possible even if this needs to be shared amongst clubs to make players safer. It would be key for these studies to focus on muscular and similar injuries that could be prevented by changes in training load and other variables. It is worth noting that an approach to predicting injury may need to be more granular as different features cause different injuries. Therefore, some methods may be better at identifying different injuries to different body parts (e.g., a single model to predict hamstring tears in footballers would be very beneficial).

There are many papers that suggest Acute:Chronic Workload Ratio (ACWR) and injuries are related and this variable could be used in a predictive model. It is worth noting however that Bornn et al. (2019) suggest that the value of ACWR may have been over-estimated by other papers. It would be almost impossible to predict injuries inflicted by opposition players in contact. However, models such as (Kelly et al. 2012), are used for contact injuries to aid the medical staff in sports teams and help improve their responses by providing information quickly from sensors. If prediction models for injuries were proved to be successful they would become a vital tool in the sporting world due to the economic and performance-based benefits that they would lead to (e.g., the financial benefits discussed in the JLT annual injury report).[3]

2.6 Research Questions

In the previous section, we discussed the open research areas that have been exposed from our review of the current literature. From this we identify the following research questions that will form each chapter in this thesis:

- **Optimising Short-Term and Long-Term Team Strategy in Football:** Chapter 3 will explore how AI and game theoretic techniques can be used to optimise the tactics selected by the manager in football games that can maximise the chances of winning a game or avoiding defeat. We identified this as a research topic using the framework presented in Figure 1.1 and saw an opportunity to research the use of game theory for the selection of tactics in football which has yet to be explored in great detail.
- **Learning The Value of Teamwork to Form Efficient Teams:** In Chapter 4, we research how we can extract the value of teamwork from sports teams and present a novel model for forming teams based on maximising the overlapping teamwork between pairs of players. This area was identified again from the decision making aspect of sports as we aim to drive more insights into human teams and then give ways of using this information to improve the chances of winning games. Our work in AAAI-20 Beal, Changder, et al. 2020 was the first to extract the value of teamwork in sports games and has helped to build a new area of research already being utilised by professional teams.
- **Combining Machine Learning and Human Experts to Predict Match Outcomes in Football:** Our final area of research in Chapter 5 is focused on the use of NLP techniques to improve the match prediction accuracy. In the previous

section, we discussed the "glass-ceiling" that exists in sports prediction and we prose a new method using text data from the media and social media to break through and learn from previously unexplored datasets. This helps learn more human context that may be missed by traditional data.

The first of these is expanded on in the next chapter which will focus on "Optimising Short-Term and Long-Term Team Strategy in Football".

Chapter 3

Optimising Short-Term and Long-Term Team Strategy in Football

In this chapter, we present a novel approach to optimise tactical and strategic decision making in football both for individual matches and in the long-term across a season. We model individual matches as a multi-stage game compromised of a Bayesian game to model pre-match decisions and a stochastic game to model in-match state transitions and decisions. We then model teams' long-term objectives for a season and track how these evolve to give a *fluent objective* to guide decision-making. We also present a method to predict the probability of game outcomes and predict the final outcome of a season to monitor long-term objectives. Empirical evaluation using real-world datasets from 760 matches shows that using optimised tactics through Bayesian and stochastic games, increases the chances of a team winning an individual game by 16.1% and 3.4% respectively. We also show that by using optimised tactics with our fluent objective and prior games, we can increase a team's mean expected finishing distribution in the league by up to 35.6%.

3.1 Introduction

Many real-world settings can be modelled as games involving opposing teams of players. In these types of games, each team optimises its tactical decisions to maximise its expected outcome (e.g., to win points, trophies, or control over resources). Examples include politics where teams of politicians aim to win an election as a party (Snidal 1985) and security where teams of agents schedule their rota to protect facilities against attackers (Paruchuri et al. 2008; Shieh et al. 2012). In the simplest case, such games can be simple one-shot interactions (e.g., elections or a quiz competition) while in more complex cases (e.g., league games or military conflict scenarios), such encounters are repeated with teams of agents aiming to optimise their long-term performance over a series of inter-related multi-step games (i.e., each game may involve multiple rounds of interaction), where one game's outcome feeds into the next one and so on. In this chapter, we focus on a model for such games and the long-term optimisation of decision-making in team sports, specifically on games of football. Here, the availability of real-world datasets presents a unique opportunity for game theoretic techniques to be developed, validated, and applied in the real-world. Across team sports, real-world data is available over long periods of time, about the same individuals and teams, in a variety of environmental contexts, thereby creating a unique live testbed for our techniques.

In football, tactical decisions may include assigning positions to players, composing a team, and reacting to in-game events. Such decisions have to be made against significant degrees of uncertainty, and often in very dynamic settings where a range of factors can influence performance (e.g., health of players, mood of the team, weather, location of games). Football presents us with an interesting challenge where a team of human agents compete against other teams of agents across long periods (usually up to 8 months) and the success of teams is not only judged in individual games but how they perform over a season in a league format. Leagues are made up of a set of teams that play every other team twice, both home and away. Teams are awarded points based on winning, losing or drawing and at the end of the season teams are awarded prize money and other incentives based on their points gained in comparison to all other teams in a league rankings/standings.

Prior AI research in team sports, and specifically football, has focused more on the contribution of individual agents within a team (Decroos et al. 2019; Beal, Changder, et al. 2020). However, to date, there is no formal model for the tactical decisions to improve a team's probability of winning and how these can be optimised in the longer term. There are a number of tactical decisions that are made both pre-match and during the match that are often made through subjective opinions (Andrienko et al. 2019).

In this chapter, we propose a formal model for the game of football and the tactical decisions that are made in individual games. We model the decisions made in football as a 2-step game: a Bayesian game into a stochastic game (Shoham and Leyton-Brown 2008). Our Bayesian game is used to represent the pre-match tactical decisions that are made due to the incomplete information regarding the tactical choices of the opposition. We then use a stochastic game to model the in-match tactical decisions based on the different states that unfold during a game of football. This to our knowledge is the first model to exploit AI decision making that is trained and tested on real-world datasets over long periods of time.[1]

We also propose a formal model for optimising the long-term performance of football teams. We assess how they can add more context to their decisions and learn from other games that unfold in the league. We introduce the novel notion of a *fluent objective* which is a long-term objective of the agent that may change over time. We stress that these variables can take the form of a broader goal (e.g., initially a team aims to finish in the top 4 of a league but after a good start this changes to aim to win the league). We use Markov chain Monte Carlo simulations to set achievable objectives that add more context to the tactical decision-making process within individual games. We also take inspiration from observational learning literature (Borsa et al. 2019; Bandura 2008; Jang and Cho 1999) to learn which tactics are most effective from other games that happen in the league.

Together these models for both individual matches and long-term decision making allows us to learn the impact of given decisions on increasing the probability of winning a game. This allows us to optimise the decisions that are made by a team and to identify tactical decisions that boost the chances of winning. As the season progresses, teams learn more about their upcoming opponents as they play more games — we encapsulate this into our model. We validate and test our models and algorithms on data from real-world matches. We show that our pre-match and in-match tactical optimisation can boost a team's chances of being successful and finishing higher in a league. We also show that we can use machine learning effectively to learn the match outcome probabilities based on the tactical decisions and accurately predict in-play state changes (the game scoreline). Thus, this chapter advances the state of the art in the following ways:

1. We propose a novel model for the game of football and the tactical decision-making process for optimising the long-term performance of human teams.

2. Using real-world data from 760 games from the past two seasons of the English Premier League (EPL), we propose a model for different team actions and learn scoreline transitions. In particular, we show that we can predict goals being scored, which leads to transitions in game-states, with an accuracy of up to 90%. We also show we can accurately predict the decisions that the opposition make which allow us to pick better tactics that are more likely to be successful.

3. We set a fluent objective based on accurate league simulations and further improve individual game payoffs by using knowledge from prior games. In particular, we

[1] This chapter expands on the models of two earlier papers (Beal, Chalkiadakis, et al. 2020; Beal, Chalkiadakis, et al. 2021), showing how to couple the two models and with additional experiments to demonstrate the value of the combined model.

show that we can increase teams' finishing position at the end of a season on average by up to 2.9 ranks (out of 20).

4. By learning payoffs, we can optimise pre- and in-match tactical decisions to improve the probability of winning a game by 16.1% and 3.4% respectively.

5. Using a fluent objective and prior game knowledge we can show an increased probability of improved long-term performance of football teams (by up to 35.6%).

When taken together, our results establish benchmarks for a computational model of football and data-driven tactical decision making in team sports. They show that by looking ahead and thinking about long-term goals, teams can add more context to the tactical decisions that are made for individual games and thus are more likely to achieve the long-term objectives that they have set. Furthermore, our work opens up a new area of research into the use of these techniques to better understand how humans make decisions in sport.

The rest of this chapter is organised as follows. In Section 3.2 we outline the background to tactical optimisation. Section 3.3 defines the model for single-game optimisation and in Section 3.4 discusses the model for long-term optimisation across a season. Section 3.5 discusses how we solve our models for the game of football and we perform a number of experiments in Section 3.6. Finally, in Section 3.7 we discuss our findings and Section 3.8 summarises.

3.2 Background and Related Work

In this section, we review key-related literature around applications of game theory to real-world problems, we also give an overview of football tactics and show how they influence the outcomes of games and performance across a season.

3.2.1 Modelling Real-world Strategic Interactions

Several works have modelled strategic decisions in a number of real-world applications using Bayesian and stochastic games. For example, Synnaeve and Bessiere (Synnaeve and Bessiere 2011) use Bayesian modelling to predict the opening strategy of opposition players in a real-time strategy game, StarCraft. Work in (Paruchuri et al. 2008) focuses on Bayesian Stackelberg games, in which the player is uncertain about the type of the adversary it may face and in (Shieh et al. 2012) the authors present a game-theoretic system deployed by the United States Coast Guard in the port of Boston for scheduling their patrols.

In (Chen et al. 2013), a model is developed for strategies in two-player stochastic games with multiple objectives explored for applications in autonomous vehicle stopping games. This paper shows one of the first applications of multi-objective stochastic two-player games. Another example of a model for stochastic games is shown in (Kardeş, Ordóñez, and Hall 2011) which shows the use of discounted robust stochastic games in a single server queuing control. Finally, work in (Avsar and Baykal-Gürsoy 2002) models the problem of inventory control at a retailer, formulating the problem as a two-person nonzero-sum stochastic game.

The work we present in this chapter uses real-world data for a real-world problem creating a model that can feed optimised strategies from a pre-match Bayesian game into an in-match stochastic game (Shapley 1953); to the best of our knowledge, this is the first time such an intertwining is proposed in the literature.

3.2.2 Modelling Long-Term Decision Making

Here we review previous work that has looked at optimising strategic decisions with long-term objectives. Our model of *fluent objectives* is inspired by these works. The model described in (Ranganathan and R. Campbell 2003), enables context awareness to help build context-aware applications. Similarly in our model, we aim to gain context of performance in terms of league standing which is determined by the performance of our team and the others in the league competition. This can help us make decisions based on the rewards in the league standings which determine prize money and winnings at the end of the competition. An example of a similar model is presented in (Sim and Choi 2003). Here, agents react to the situations presented by the ever-changing variables in the stock market.

In our work, we consider how agents learn from prior games to gain a better understanding of what tactics work against their opponents. This is closely related to the work presented in (Borsa et al. 2019), where the authors explore the notion of “observation learning” which is a type of learning that occurs as a function of observing, retaining and imitating the behaviour of another agent. This applies to football as if we observe another team perform well against another opponent then we may want to imitate their tactics to help us to win.

3.2.3 Decision-Making in Sport

Past work in the sports domain focus on tactics and looking at game-states in football. Firstly, work in (J. Jordan, Melouk, and Perry 2009) explores different risk strategies for play-calling in American Football (NFL). Although game theory has not been applied to football tactics in prior work, in (Tuyls et al. 2021) the authors discuss the use of game theory for penalty-taking in football (a free shot against a goalkeeper). Secondly, (Fernández, Bornn, and Cervone 2019) provides a model to assess the expected ball possession that each team should have in a game of football, this can be used to identify where teams can make changes to their styles and to improve their tactics. Another application of learning techniques in football is shown for fantasy football games in (Matthews, Ramchurn, and Chalkiadakis 2012) and for the RoboCup competition in (Stone and Sutton 2001; Stone, Sutton, and Kuhlmann 2005).

To give more background around these papers and the problem we are looking to solve, in the next subsection, we provide an overview of football tactics and their importance to the game.

3.2.4 Football Tactics Overview

The basic foundations of sports and football tactics are discussed in (Gréhaigne, Godbout, and Bouthier 1999; Bate 1988) and applications of AI is discussed in (Beal, T. Norman, and Ramchurn 2019). There are multiple tactical decisions that a coach or manager must make before and during a game of football. These can have a significant impact on the overall result of the game and can help boost the chance of a team winning, even if a team does not have the best players. Managers and coaches prepare for their upcoming matches tactically to the finest details, usually by using subjective opinions of their own and opposition team/players. Example pre-game decisions that are made by teams include:

- **Team Style:** A team playing style is a subjective concept that relates to the team's overall use of different playing methods. There are many different styles that a team can use but these can be analysed using game statistics and similar styles can be identified.

Style	Description
Tika-Taka	Attacking play with short passes.
Route One	Defensive play with long passes.
High Pressure	Attack by pressuring the opposition.
Park The Bus	A contained defensive style.

Table 3.1: Example Playing Styles.

- **Team Formation:** The formation is how the players are organised on the pitch. There is always 1 goalkeeper and 10 outfield players who are split into defenders (DEF), midfielders (MID) and forwards (FOR). An example of a formation is 4-4-2, this represents 4 defenders, 4 midfielders and 2 forwards. Figure 3.1 shows how this is set up on the pitch (attacking in the direction of the arrows).

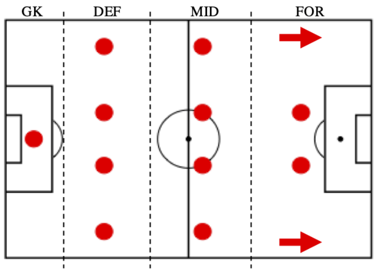


FIGURE 3.1: Example Team Formation (4-4-2).

- **Selected Players:** The selected players are the 11 players that are selected to play in the given starting formation or selected to be on the substitute bench (between 5-7 players). Some players may perform better in different styles/formations or against certain teams.

In terms of the in-game decisions that are made, one change that can be made is a substitution (other examples include tweaks to the style and formation). To optimise how this substitution is made, we can model the in-game decisions as a stochastic game and look to make optimised substitutions that increase the probability of scoring a goal. This can help teams to improve their chances of winning games by making the right decision at the right time.

Due to the number of decisions that can be made by teams in football, there are many uncertainties both in what the opponent may do and on how the decisions made may affect the overall outcome of the game. In this chapter, we aim to address these uncertainties by predicting what the opposition will do and how we should respond. In the next sub-section, we discuss how these tactics are important in the long term as well as for individual games.

3.2.5 Long-Term Football Tactics

In football, each game has an impact over a long period of time and on the overall league standings. The final league standings is the final position of all teams in a league based on the points they have gained over an N game season. In a standard football league (e.g., English Premier League (EPL) or German Bundesliga), across a season, each team plays each other twice (once home and once away) a win is worth 3 points, a draw 1 point and a loss no points. There are significant intrinsic and financial gains to be made by finishing higher up the table and there are certain milestones that teams aim for to boost their success such as qualification for European competitions.[2]

[2]http://eightyfivepoints.blogspot.com/2018/03/show-me-money-how-much-is-each-premier.html.

The season is often broken down into given "game-weeks" where all teams play a game within the week. We can therefore break down the season into these game weeks as incremental steps in a game. In each week our team plays a game and some other games also take place. Therefore, we want to maximise a team's performance in their game and learn from other games for the future when we play those teams (see Figure 3.4).

In this chapter, we aim to model teams tactical decisions not only in single games but also based on the overall league environment and use *fluent objectives* to add context to our decisions and prior games' knowledge to replicate the successful tactics used by other teams. This allows us to optimise the long-term performance of teams and improve their league outcome.

3.3 A Formal Model for the Game of Football

We model the tactical decisions that are made in individual football matches into two parts (shown in Figure 3.2). First, we model the pre-match tactical decision-making process as a Bayesian game, taking into account the fact that each team has incomplete information regarding the opposition's tactical decisions before the game begins. Second, we model the in-match decisions as a stochastic game due to changing states of a game of football as the game progresses (see Section 3.3.2 for more details on the states of the game). We use the Bayesian into a stochastic game model as a framework to conduct learning in, rather than to solve for equilibria. This is because we have a dynamic setting with multiple sources of uncertainty, where opponents change for every instance of the game. In this context, solving for some kind of equilibrium is impractical, but opponent types and probabilistic transitions among states that represent scorelines given formation and style are natural in this problem. As such, our Bayesian and stochastic game framework provides a natural model to facilitate learning in this domain.

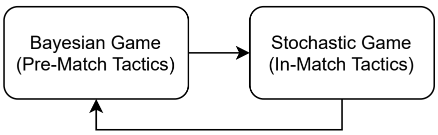


FIGURE 3.2: Football Game Model

3.3.1 Pre-Match Bayesian Game

We define a Bayesian game consisting of two teams $T = \{T_\alpha, T_\beta\}$ where T_α is the team whose actions we are optimising and T_β is the opposing team. Each team has a corresponding action set $a \in A_\alpha$ and $a \in A_\beta$ that are sets of one-shot actions describing tactical choices before the match (e.g. some action a may represent selecting the team formation or selecting the starting team).[3]

Each team may select a type Θ. We define this as θ_α and θ_β where $\theta \in \Theta$. These types correspond to the style of football that an opposition is likely to use (e.g., tika-taka, route one or high pressure). To select better tactics a team must predict an opposition type and actions. This is the teams' prior beliefs about a team which in this case is the probability that the opposition will play a given formation and style combination. Our prior belief probability function is defined as $p(a_\beta, \theta_\beta) \in [0, 1]$ which represents the probability that a team T_β will select a given action a_β (a team formation) and the style θ_β.

The payoff function in our game is used to represent the probability of gaining a positive result from a game based on the selected actions, as well as the type and prior beliefs of the opposition. We calculate the probability of a win, draw or loss for a team and weight these based on their impact on league points. A win is weighted to 3, a draw to 1 and a loss to 0. The payoff utility function is then defined as $u(a_\alpha, \theta_\alpha | a_\beta, \theta_\beta)) \in \mathbb{R}$. This represents the payoff (weighted sum of result probabilities) based on the teams selected actions (a_α,a_β) and their style (θ_α,θ_β) where $a \in A$ and the type is $\theta \in \Theta$. We therefore, define our Bayesian game as:

$$G^B = (T, A, \Theta, p, u) \tag{3.1}$$

We assume that neither team knows the other's tactics, but both have access to data from previous games. This data can be used to predict the likely style and formation that a team will use. A team looking to maximise their chances of winning a game would select the action set of decisions and style which maximises the payoff function and therefore gives the greatest probability of winning a game. However, there are multiple strategies that we can take to optimise the selected decisions depending on the state of the team in the real world (e.g., league position, fighting relegation, a knock-out cup game etc). Therefore, we present three approaches to optimising the selected tactics:

- **Best Response:** maximises the chances of a win.

$$\underset{a_\alpha, \theta_\alpha}{\operatorname{argmax}} \left\{ \sum_{a_1 \in A_\alpha} \sum_{a_2 \in A_\beta} u(a_1, \theta_\alpha | a_2, \theta_\beta) \cdot p(a_2, \theta_\beta) \right\} \tag{3.2}$$

  where, A_α and A_β are the set of actions that team α and β can take respectively. We aim to maximise the sum of payoffs u multiplied by the probability of the opposition selecting the action a_2 and style θ_β. This approach has the highest risk as we are not considering the opposition payoff, we just select the best payoff for ourselves.

[3]A full list of the symbols used in this chapter can be found in the appendix.

- **Spiteful Approach:** minimises the chances of losing.

$$\underset{a_\alpha,\theta_\alpha}{\operatorname{argmin}} \left\{ \sum_{a_1 \in A_\alpha} \sum_{a_2 \in A_\beta} u(a_2, \theta_\beta | a_1, \theta_\alpha) \cdot p(a_2, \theta_\beta) \right\} \tag{3.3}$$

  where, we aim to minimise the sum of the payoffs u for the opposition team multiplied by the probability of the opposition selecting the action a_2 and style θ_β. By reducing the chances of the opposition winning the game, this increases the chances of a draw or a win for our team. This approach is likely to select tactics which are more reserved and defensive to increase chances of a draw as we do not consider our payoff, instead we are selecting the payoff that limits the opposition.

- **Expectimax Approach:** In this approach we find the tactics that maximise the chances of winning the game but also minimise the chances of the opposition winning a game.

$$\underset{a_\alpha,\theta_\alpha}{\operatorname{argmax}} \left\{ \sum_{a_1 \in A_\alpha} \sum_{a_2 \in A_\beta} (u(a_1, \theta_\alpha | a_2, \theta_\beta) - u(a_2, \theta_\beta | a_1, \theta_\alpha)) \cdot p(a_2, \theta_\beta) \right\} \tag{3.4}$$

  where, we aim to maximise the sum of the payoffs u for team α while also minimising the sum of the payoffs u for the opposition team. This is weighted by the probability of the opposition selecting a_2 and θ_β.

The different optimisation approaches allow teams to select the tactics which are best suited to their risk levels which may be dependant on the overall position of a team in a league or the individual game scenario. The pre-match decisions that are made by the team are then used as their pre-match tactics which feed into the stochastic game defined next sub-section.

3.3.2 In-Match Stochastic Game

As a game of football progresses the game changes state in terms of the scoreline, in turn changing the objectives for either team. If a team is winning they may make defensive changes to ensure they win the game and if a team is losing they may make attacking changes to get back into the game. Due to these state changes, and football being a repeated game, we model the in-game tactical decisions as a stochastic game (Shapley 1953). A stochastic game is a tuple, in our case, this is defined in Equation 3.5.

In our stochastic game, we define the two teams as $T = \{T_\alpha, T_\beta\}$ where T_α is the team whose actions we are optimising and T_β is the opposing team. We have a set of states X which represent the different possible scorelines in a game starting at 0-0 (where the left number represents the home team goals and the right number represents the away team goal). Each team has a corresponding set of strategies $\sigma(x)$ at each of the different

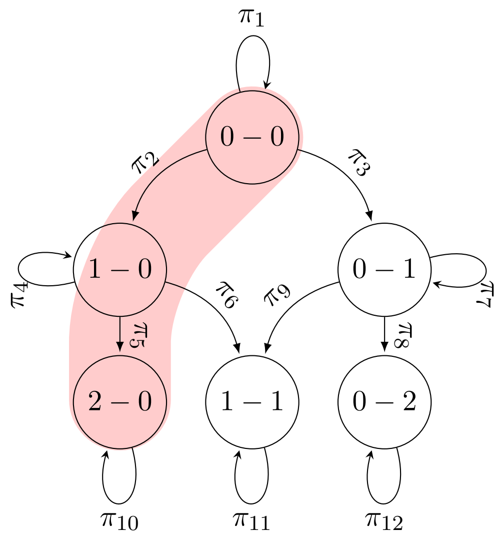


FIGURE 3.3: An example of a state-transition diagram in a match with 2 goals being scored and the different routes that can be taken through the states. The highlighted route shows the transitions for a match ending 2-0 to the home team.

states $x \in X$. The strategies represent the current team formation, players and the style of play. At the starting state, x_0 (0-0) the team strategies correspond to the selected actions from A_α and A_β defined in the Bayesian game in the previous section.

Given the selected strategies of the two teams and the current state (x) we can calculate the probability of a transition (π) into another state x'. This is defined as $\pi(x'|x, \sigma_\alpha, \sigma_\beta)$ where σ is the strategy of each team in state x. In the case of football, from each state, there are only two possible states that can be moved to. These will be transitioned by a goal for the home team or a goal for the away team. The other probability we will have to consider is that the state will not be changed for the remainder of the match. In this problem, the length of the game (t) is known (90 minutes + injury time) and therefore the probability of state changes will change as the remaining time of the game decreases. The utility function $u(x, \sigma_\alpha, \sigma_\beta)$ for this game equates to the probability of a transition into a more positive state (e.g., a team scoring in a 0-0 state to move into a 1-0 state or a winning team (1-0) staying in that state for the remainder of the match time). Given these definitions, we define our stochastic game as:

$$G^S = (X, T, \sigma(x), \pi, u) \tag{3.5}$$

Each team aims to move to a more positive state than they are currently in. They make decisions to improve the probability of moving to the more positive state based on their strategy $\sigma(x)$. The state changes based on goals in the game, Figure 3.3 shows the possible transitions in a game with two goals.

In the next section, we discuss how we model long-term decision making and how we can factor in the consequences of our decisions in a wider environment and how this affects the outcome of the league.

3.4 Modelling Long Term Team Performance

In this section, we discuss how we model the long-term performance of football teams over a season and identify how we can use fluent objectives and learn from games to optimise the long-term performance of a team. At the start of a season or competition, a team will have a target. Across a full season of football in a league competition, there are many objectives that a team can have to maximise their financial rewards and reputation. For example, as discussed in Section 3.2.5, in the English Premier League there is always only one winner but there are also benefits to finishing in the top 4, top 7 and avoiding finishing in the bottom 3. We therefore, model an entire season to optimise a team's long-term performance in any league across the world and at any level.

3.4.1 Sequence of Multi-Games Across a Season

In Figure 3.4 we show the structure of our model for an entire season in football. We build on the multi-step (Bayesian into stochastic) games for optimising single game tactics that we have outlined in the previous sections. These are designed to help teams achieve their objectives in an N game season. There is a sequence of steps that we highlight and show how each one feeds into the next. We also show how a teams' objective can be fed into the first game which informs tactical decisions as well as what can be learned from enacting those decisions (e.g., certain tactics that work well against certain opponents).

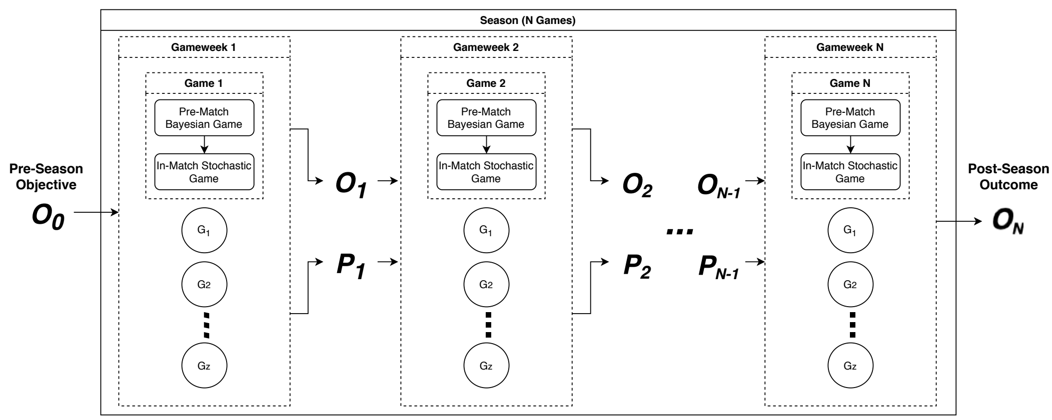


FIGURE 3.4: Sequence of Multi-Games Across a Season

The objective may change over time, and hence it is a *fluent objective*; e.g., a team may initially intend to win the league, but poor outcomes from early games may lead to the selection of a less ambitious objective. As we show in Figure 3.4, the pre-season objective is set as O_0, this then changes each game-week as the environment around the team develops, changing to O_1 after game-week 1, O_2 after game-week 2 and so on until the final in-season objective the week before the final game of the season $N-1$. The final fluent objective, O_N, corresponds to the overall end of season outcome.

We also consider how we can learn from the games that are played as the season progresses. As we play each game we learn something new, both about what works for a team and what works against a given opposition based on the actions/style that they are likely to select. Therefore, we learn parameters from each game that we can carry forward through each game week and similarly to the fluent objective we update each week. For example, we may find that when a team uses a given formation against a certain style of opponent we see better results. As we show in Figure 3.4, this is encapsulated by a *prior knowledge parameter* P, which is updated after each game we play where P_1 is after game-week 1, P_2 after game-week 2 and so on.

Finally, we must consider other games that we observe, G_N is the set of other games in game-week N and $G = \{G_1, G_2, ..., G_z\}$ where z is the number of other games played in that week. Within each game week, all other teams also play one another, so that at the end of the season, each team has played every other team twice (once at home and once away). For example, in the EPL there are 20 teams in the league and each team plays 38 games, meaning there are 342 games that may be observed and learned from. The outcomes of all games affect the relative positions of each team, and hence have an impact on a team's fluent objective O. We can also learn what styles and formations work well or poorly against given teams, informing our prior knowledge P for future games.

3.4.2 Fluent Objectives in Football

At the start of each season, a team will have some objective for what they are looking to achieve in the next season. This is decided based on several factors such as previous season performance and money invested into the team. The goals are usually set by the owners/directors of the team and are based on their subjective opinions of how their team should perform and where they should place in the league against the other teams. The opinions of what the team should achieve then changes over the season which can drive key decisions such as a change in coach/manager for an under-performing team or investing more money into an over-performing team so they achieve a European place which comes with huge financial gains.[4]

Our model for the fluent objective can objectively evaluate how we expect a team to perform over a season and allow teams to change their tactical decision-making based on this. There are two different objectives that can be set: a more granular objective of the expected league position and an objective of what could be achieved in terms of broader incentives in the league (e.g., avoiding relegation or qualifying for European competitions). In this chapter, we focused on the latter and can define the set of possible objectives as $O = \{o_1, o_2, ..., o_k\}$ where k is the number of different objectives. An example of the set of values that an O_x objective variable can take in the EPL would be:

- **Winning the League (o_1):** Team who finishes top of the league.
- **Qualifying for the Champions League (o_2):** Top 4 teams, so in this case, the objective relates to teams finishing 2nd-4th.[5]
- **Qualifying for the Europa League (o_3):** Another European competition usually awarded to teams who finish between 5th-7th.
- **Top Half Finish (o_4):** Due to financial benefit teams often aim to finish in the top half of the table (higher than 10th).[6]
- **Avoiding Relegation (o_5):** The bottom 3 (18th-20th) teams in the EPL are relegated into the second division of the English football leagues.

To set the objective we can simulate how we expect the season to unfold and create a distribution $\mathcal{D}$ that allows us to use a Maximum a Posteriori (MAP) estimation (Gauvain and C. Lee 1994) for the probability of the team finishing in each position. This then allows us to calculate a set of probabilities for of a team achieving each objective $\mathcal{P} = \{p(o_1), p(o_2), ..., p(o_k)\}$. We then set the O_o (for a pre-season objective) as the most likely objective that can be achieved that season.

This process can then be re-run after each game week is completed to give the fluent objective O_1 to O_{N-1}. Our simulation of the league will include the real results which will get more accurate as the season progresses and we learn more about each team. At the end of the season, we can compare O_0 to O_{N-1} to the final outcome O_N that the team achieves. Next, we discuss how we can learn from other games we observe.

[4]In other settings, these type of objectives could be the defence of a given target or the rescue of a person.

[5]https://www.premierleague.com/european-qualification-explained.

[6]https://www.goal.com/en-gb/news/how-much-money-do-premier-league-2019-20-winners-get/19jbauady17cw1ieojo40yextz.

3.4.3 Learning From Prior Games

As well as the fluent objective, we can also improve the tactical decision-making in our Bayesian and stochastic games by adding prior knowledge P that we learn after each game we play and observe. In more general terms we aim to observe and learn from other successful agents and our own actions.[7]

We can learn a set of weights $\mathcal{W}$ that relate to how effective given style/formation pairs (actions that are made in the multi-step games) that we select in our games are against given oppositions style/formation pairs. These weights are initially set to 1 and are then increased if found to be effective and decreased if found to be ineffective. These can be updated after each game week and also updated from the other games that we observe. Our *prior knowledge parameter* (P) is defined in Equation 3.6.

$$P = \begin{pmatrix} w_{11} & w_{12} & w_{13} & \dots & w_{1j} \\ w_{21} & w_{22} & w_{23} & \dots & w_{2j} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ w_{i1} & w_{i2} & w_{i3} & \dots & w_{ij} \end{pmatrix} \tag{3.6}$$

Where $w \in \mathcal{W}$ and i/j is the number of possible style/formation pairs. The columns represent the style/formation pair selected by our team and the rows represent the style/formation selected by the opposition (e.g., w_{ij} is how effective our pair i is against an opposition using pair j).

3.5 Solving the Games and Learning the Models

In this section we discuss the approaches we use to optimise tactics and solve the models discussed in the previous section. These can be summarised as:

- **Pre-Match Bayesian Game:** Predicting the actions an opponent will use and learning the probability of winning a game based on each formation and style a team can select to be able to optimise the decisions made.
- **In-Match Stochastic Game:** Learning the probabilities of moving to new states from each state (e.g., if 0-0 probability of moving to 1-0, 0-1 or staying at 0-0). Then assessing how each possible action affects the probability of moving to a more positive state. Using this, the selected actions are optimised to maximise the chances of positive outcomes of games.

[7] This could also be applicable in swarms of UAVs or imitating other agents trading in the financial markets settings.

- **Long-Term Optimisation:** To consider the long-term impact of decisions, we simulate the season to set the fluent objective which is used in the optimisation of the pre/in-match games. The effectiveness of decisions made are learned based on other games that unfold in the wider environment.

In what follows, we expand on each game and optimisation in more detail.

3.5.1 Solving the Pre-Match Bayesian Game

For the game G^B defined in the previous section, we formulate a model to solve this and to select the optimal tactics which maximise a team chances of obtaining a positive outcome from the game.

3.5.1.1 Predicting the Opponent Strategy

When predicting how an opponent will select their strategy, there is limited historical data for games against them in the past. Therefore, we cluster the teams into different playing style categories so we can look for trends in how the opposition play against similar team styles. To cluster the teams, we use a feature set ($\mathcal{F}$) containing the number of: passes, shots, goals for, goals against and tackles that a team has made. K-means is used to cluster for $|C|$ clusters using Equation 3.7 which aims to choose centroids that minimise the inertia, or within-cluster sum-of-squares criterion.

$$\sum_{i=0}^{n} \min_{\mu_j \in C}(||\mathcal{F}_i - \mu_j||^2) \tag{3.7}$$

where, n is the number of teams and C is the set of cluster means μ.

This allows us to evaluate the style of a team and how each cluster aligns with human expert styles. For example, a team with many passes and many shots may be seen as a "tika-taka" style team which is an attacking team playing a passing style of football (e.g., the World Cup winning Spain team from 2010 or Barcelona), whereas a team with fewer passes and defensive play may have a "route one" style where they look to use long balls over the opposition defence.

Using the clusters of team styles we can learn the strategies that an opposition uses against similar teams to then predict what they will do against our own team. To do this we build a model using a Support Vector Machine (SVM) with a radial basis function kernel (Scholkopf et al. 1997), shown in Equation 3.8. The algorithm learns using features $\mathcal{F}$ which are made up from the tactics from the prior N games against teams from the same style cluster.[8]

$$a_\beta = \sum_{i=1}^{C} \lambda_i \phi(|\mathcal{F} - m_i|) \tag{3.8}$$

where, a_β is the predicted oppostion actions, C is the clusters, m is the cluster centres and λ is the cluster weighting.

3.5.1.2 Learning the Expected Payoffs

To learn the payoffs from historical data we develop a model that uses the team's tactical style, potential formation and team strength to give probabilities of a team winning the game. The set of features that we use in our model are: home team style, away team style, home team formation, away team formation and then team strengths are calculated by using the outputs from the model described in (Dixon and Coles 1997) (likelihood of a home win, draw or away win). The target class ($\mathcal{O}$) is the final result of the game: home team win, away team win or a draw.

Using these features, we train a multi-class classification deep neural network. The neural network is trained using stochastic gradient descent using a categorical cross-entropy loss function and a soft-max activation function ($-\frac{1}{N}\sum_{i=1}^{N} \log p[y_i \in \mathcal{O}_{y_i}]$). Where, N is the number of games that we are using to train the model and $p[y_i \in \mathcal{O}_{y_i}]$ is the models' probability that y_i is in the class $\mathcal{O}$. This model takes the given teams, possible playing styles and possible formations to give a probability of winning, drawing or losing the game. Finding and selecting optimised tactics is discussed in the next subsection.

3.5.1.3 Optimising Pre-Match Tactics

Once we have a model that learns the expected payoffs from the different possible actions (by ourselves and the opposition), we then look to find the best actions/decisions to make, i.e., those which maximise the chances of gaining a positive outcome in the game.

We use the methods that we discussed in Section 3.5.1.1 to predict the actions and style that an opposition is likely to select. Clustering methods are used to find their most likely tactical style θ and then the formation prediction model to give the formation with the highest probability of being selected. The predicted opposition style and formation, are used to explore our possible actions and select the best tactics. Table 3.2 shows the payoffs for the different actions that we can take (when facing a given opposition formation and style). Here, θ corresponds to a given style we are able to play in (x

[8] Our testing found that $N = 5$ prior games was optimal for the accuracy of predicting the true strategies used by the opposition.

possible styles), a corresponds to a given formation (y possible) and then $p(outcome|\theta, a)$ is the probability (output from the model discussed in Section 3.5.1.2) for the outcome of the game given the selected style and formation. The payoff for the team is the weighted sum of win and draw probabilities and these values are pre-computed so that we can then use the three approaches defined in Section 3.3.1 (Best Response, Spiteful and Expectimax) to optimise the tactical decisions that we can take depending on the opposition. In the next sub-section, we discuss how these tactics can be taken into the game and changed depending on how the game unfolds.

	θ_1	$\dots$	θ_x
a_1	$p(outcome\|\theta_1, a_1)$	$\dots$	$p(outcome\|\theta_x, a_1)$
a_2	$p(outcome\|\theta_1, a_2)$	$\dots$	$p(outcome\|\theta_x, a_2)$
a_3	$p(outcome\|\theta_1, a_3)$	$\dots$	$p(outcome\|\theta_x, a_3)$
$\vdots$	$\vdots$	$\dots$	$\vdots$
a_y	$p(outcome\|\theta_1, a_y)$	$\dots$	$p(outcome\|\theta_x, a_y)$

TABLE 3.2: An example payoff table for a team who can have a tactical style of θ_1 to θ_x and a given formation a_1 to a_y.

3.5.2 Solving the In-Match Stochastic Game

Optimised strategies for our in-match stochastic games G^S are computed using historical data of the team tactical setups (style and formation as discussed in the previous section). By using our models for G^S we learn the state transition probabilities (π) and evaluate how certain tactical actions will affect this and therefore learn the action payoff values. This allows teams to make in-match decisions that can boost their chances of staying in a positive state or moving into a more positive state by scoring a goal. An example of this problem is shown in Figure 3.3. This is expanded on in the following subsections.

3.5.2.1 Learning the State Transition Probabilities

Prior work by Dixon and Robinson (Dixon and Robinson 1998) models how the rate of scoring goals changes over the course of a match. Their model incorporates parameters for both the attacking and the defensive strength of a team, home advantage, the current score and the time left to play. They show how the scoring rate tends to increase over the game but is also influenced by the current score. They then use their model to find the probability that the game will end in a given state which can be used for match outcome prediction and goal-time prediction. We take inspiration from the model presented by Dixon and Robinson to learn the state transition probabilities (π) that we need to use in our stochastic game.

To learn state transition probabilities we build a new model at each game-state that will give the probability of each of the possible outcomes from that state (home goal, away goal, no goals). We use a feature set made up from the team strength, the teams' formation and style are taken from the Bayesian game (in this game we know our oppositions tactics and style but not the in-match actions they may take). For our model (ϕ) we use the SVM classification model (with an RBF kernel) described in Section 3.5.1.1 and Equation 3.8. Also, π is the transition probability of moving from state x to state $'x$ and $\mathcal{F}$ is the feature set. This means $\pi_{x,x'} = p(x \rightarrow x') = \phi_x(\mathcal{F})$ giving a prediction model for each of the possible states $x \in X$ the game could be in.

3.5.2.2 Learning the Action Payoffs

We build on the models discussed in the previous section to include new features into the feature set $\mathcal{F}$ to model the effect of in-match decisions such as substitutes. The first feature we use is a measure of a player's contribution to the team, which represents the impact of a new substitution on the pitch. This allows us to calculate the payoff of the action (substitute) so that we can make an optimised decision at a given point in the game. To calculate the contribution of the players on the bench we use the centrality metric that is discussed in (Beal, Changder, et al. 2020). This metric gives the importance of a player to the overall team network. We also use the remaining match time as a feature so we can see how long an action has to make an impact on the game. These features are used to update Equation 3.8. The payoff of each action is the transition probability of moving to a more positive state (e.g., if a team is winning 1-0 it is the probability of making it 2-0 or if a team is losing 3-0 it is the probability of making the game 3-1). The computed payoffs are used to optimise our in-match decisions which are discussed in the next subsection.

3.5.2.3 Optimising In-Match Tactics

Assuming the standard rules of football, each team can make up to 3 substitutions in a game (these can be one at a time or all at once) and has 7 players to choose from, meaning there are 64 combinations of actions (including doing nothing) that we can take at each game-state. We pre-compute the payoffs for each of these possibilities and then select the optimised action to take. Depending on if the team wants to remain in or move to a better state, we can optimise the actions by using two different approaches:

- **Aggressive approach:** Choose the action that maximises the probability of moving to a more positive state.

- **Reserved approach:** Choose the action that maximises the chances of staying in the current state (if winning).

3.5.3 Calculating the Fluent Objective

In this section, we discuss how we simulate seasons, calculate the fluent objective, and how this can be used to optimise game tactics.

3.5.3.1 Simulating Season Outcomes

When we simulate the season outcomes and calculate the distributions of where we expect the team to finish we must predict all remaining games in the season for both our team and all other teams in the league. To do this we use the match prediction model which is defined in Section 3.5.1.2. To simulate the remaining games of the season, we use the real-world fixture list and predict the outcome of each game. We find the probability of a home win, away win and draw and use a Monte Carlo Markov chain simulation (Mooney 1997) to simulate all remaining games and the total points that each team will gain (3 points for a win, 1 for a draw and 0 for a loss). This works well as it emulates the randomness that we see in real-world football games. We repeat this process 100,000 times for each simulation which allows us to derive a distribution for the probability that a team will finish in each place in the league in the final standings. An example distribution is shown in Figure 3.5.

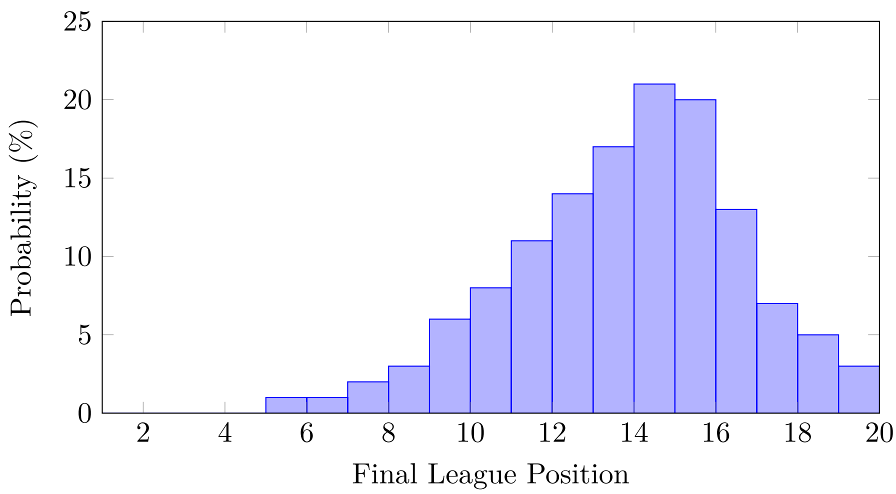


FIGURE 3.5: Example League Outcome Probability Distribution.

3.5.3.2 Setting the Fluent Objective

Once we have calculated the distributions of possible place outcomes from the MCMC simulation, we use this to find the most likely objective outcome that can be used to update our fluent objective. More specifically, we use a Maximum a Posteriori (MAP) estimation (Gauvain and C. Lee 1994) to set the fluent objective. To do this, we can use the posterior distribution to find interval estimates of the final position for the team in the league. We use the position intervals for the objectives discussed in Section 3.4.2 and can find the $o_k \in \mathcal{O}$ that maximises the posterior PDF. This then sets the objective O_n that is used in game-week n and is updated after each game week. In the real world, decision makers may want to set their own objective depending on their own risk appetite.

3.5.3.3 Optimising Tactics using the Fluent Objective

Once we have set the fluent objective we can now use this when optimising the team tactics in the multi-step game for optimising individual game tactics in that game week. In the pre-match Bayesian game outlined in Section 3.3, we select the optimisation method depending on whether the team is on track for their objective or not. This is process is outlined in Figure 3.6.

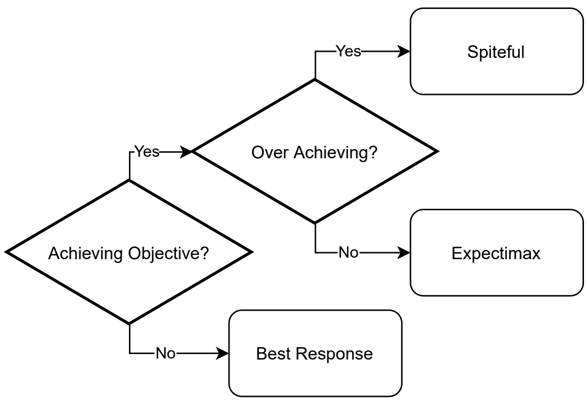


FIGURE 3.6: Selecting Bayesian Game Optimisation Method

In terms of the in-match stochastic game, two options can be selected when making in-match decisions. The *aggressive approach* is used if we know that the objective is to win, and the reserved approach is set if a team is winning/drawing and is happy with their current state.

3.5.4 Learning From Previous Games

In this sub-section, we discuss how we can learn from prior games that all teams in the league play. This allows us to find formation/style combinations that work best against a given formation/style combination that an opposing team may use. To do this we learn a matrix of weights P that corresponds to estimated successes of the formation/style combinations. To estimate each of the weights $w \in P$ we factor in both the games that we have played as well as the games that we have observed. Each weight w corresponds to how effective a given formation/style combination is against a given opposition formation/style. These are computed using a weighted average shown in Equation 3.9 where the games won while using the formation/style (x) against the given opposition formation/style (y), both in games played by a team (first fraction) and in other observed games in the league (second fraction).

$$w_{xy} = \left(\mu_1 \frac{\text{games won}}{\text{games played}} + \mu_2 \frac{\text{observed games won}}{\text{observed games}} \right) \div 2 \tag{3.9}$$

Where, $\mu \in \mathbb{R} | 0 < \mu < 1$. These weights in P are updated after each game week so should become more accurate across the season. In game week 1 all weights can either be set to 1 or be carried over from the previous season. In the next subsection, we outline how P is used to optimise the pre-game tactics in the Bayesian Game and in-match decisions in the stochastic game.

3.5.4.1 Optimising Tactics using Prior Games

The computed weights in P discussed in the previous subsection are used when making our pre-match decisions in our Bayesian game. In the optimisation model, a payoff table is computed for each combination of opposition actions to give the probability of the match outcomes based on the selected style θ_β and action a_β (formation). The payoff for the team is the weighted sum of win and draw probabilities made up from the different decisions that we can make.

We apply the computed weights in P to the payoff table to weigh each payoff depending on how successful these have been in prior and observed games. Therefore, we can optimise the tactical decision based on the weighted payoffs in these tables using either the best approach, spiteful or expectimax approaches which are decided based on our fluent objective. The same approach can be applied when changing the formation and style in the in-match stochastic game and each change made can be weighted by the corresponding element in P.

In the next section, we evaluate our models and assess the performance over a whole season of games. We assess how our models perform in individual games as well as how the inclusion of O and P can be used to help teams improve their performance and meet their long-term objectives.

3.6 Empirical Evaluation

We use a dataset collected from two seasons (2017/18 and 2018/19) from the English Premier League (EPL).[9] The dataset breaks down each of the games from the tournament into an event-by-event analysis where each event gives different metrics including: event type (e.g., pass, shot, tackle etc.), the pitch coordinates of the event and the event outcome. This type of dataset is industry-leading in football and used by top professional teams. Thus, this is a rich real-world dataset that allows us to rigorously assess the value of our model. The experiments[10] performed are as follows:

3.6.1 Experiment 1: Testing the Opposition Action Prediction

In our first test, we aim to evaluate the performance of the style clustering methods and the team formation prediction. This allows us to accurately predict the tactics of opposition and therefore optimise ours around this.

To select the number of clusters we use an elbow approach (shown in Figure 3.7) to find the point where the within-cluster sum of squared errors (SSE) will decrease and find that 4 clusters are the best to use.

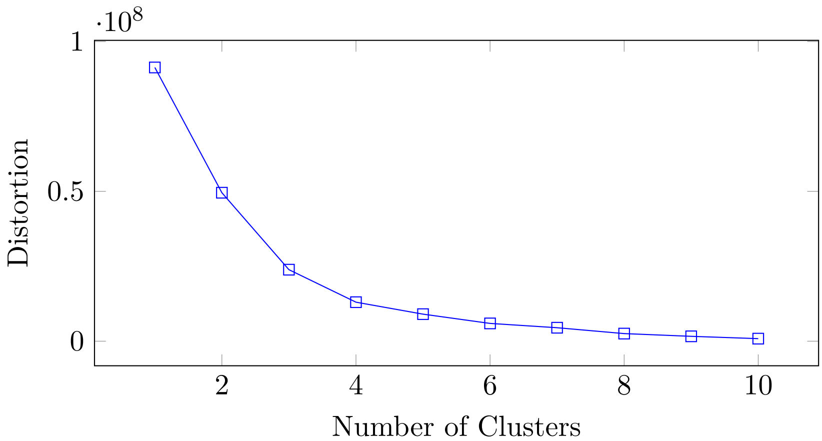


FIGURE 3.7: Elbow Method to Find Number of Style Clusters to Use.

[9] All data provided by StatsBomb - www.statsbomb.com.
[10] Tests have been run using Scikit-Learn and TensorFlow.

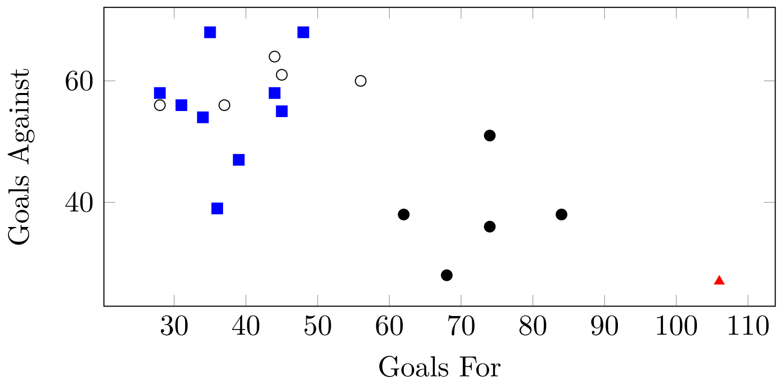


FIGURE 3.8: 2017/18 EPL Team Style Clusters.

We next test our opposition formation prediction model as discussed in Section 3.5.1.1. Using features taken from the prior 5 games against teams in the same style cluster we predict team formation. We predict the correct formation with an accuracy of 96.21% (tested on using a train-test split of 70% to 30% with a cross-validation approach for 10 folds). The model achieved a precision score of 0.9867, recall score of 0.9135 and an F1 score of 0.9441. There were a total of 30 different formations used across the season with the most popular formation being '4-2-3-1' used 21% of the time. In future work, we could use alternative clustering methods such as "Principle Component Analysis" (PCA) to find the style clusters (Wold, Esbensen, and Geladi 1987).

3.6.2 Experiment 2: Learning the Payoffs

We build the deep learning model defined in Section 3.5.1.2, that takes the actions of the teams and the team strengths into account, the model then assigns a probability to the predicted outcome of the game (home win, draw, away win). We test the outcome probability model by evaluating the accuracy in our dataset and comparing the results to the following standard approaches: Logistic Regression classifier (Dreiseitl and Ohno-Machado 2002), Dixon and Cole (1997) model and the bookmaker's pre-match odds (sourced from *oddschecker.com*). The results are shown in Figure 3.9 (train-test split of 70% to 30% with a cross-validation approach for 10 folds).

This shows that by having extra information regarding the team formation and style clusters (used by the pay-off and logistic models), we can more accurately predict the match outcomes. Therefore, the pay-off model shown in Figure 3.9 produces better predictions which are used to optimise our actions in the Bayesian game.

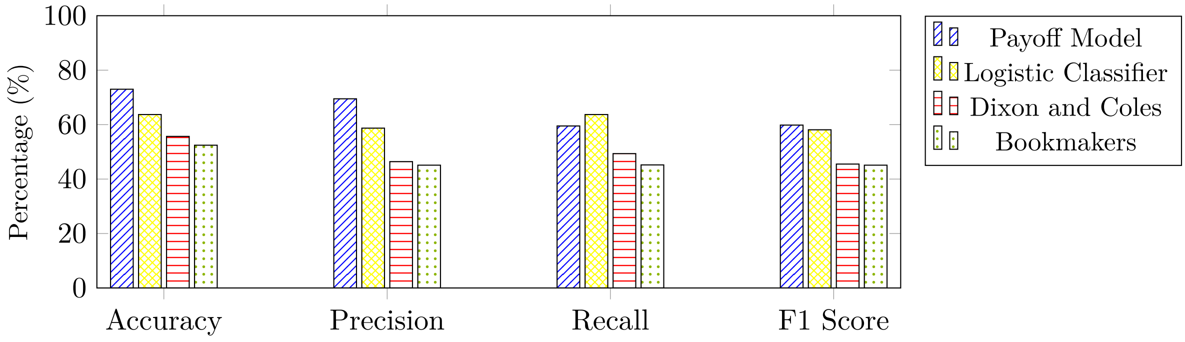


FIGURE 3.9: Payoff Model Performance Comparison.

3.6.3 Experiment 3: Pre-Match Optimisation

To test the pre-match Bayesian game we iterate through all games in our dataset (760 games) across the two EPL seasons and find the optimised tactics using the 3 different optimisation approaches discussed in Section 3.3.1.

By calculating the optimised tactics we can compare our approaches and validate our models using real-world data. Firstly, we look at how "close" our optimised tactics are to what was selected for the real-world game. We define "closeness" as a formation that is equal to our recommendation or is only 1 change away (e.g., 4-4-2 is close to 4-5-1 as you can move a striker to midfield to give the "close" formation). Using this metric we evaluate the optimisation methods for tactic recommendations and find that the best response method has a closeness of 35.3%, the spiteful approach has a closeness of 59.7% and the expectimax approach is at 44.6%. This shows that the spiteful approach is the closest representation to the real-world. However, when this is split into home and away (50% and 69%) tactics we see that this number is skewed by teams that aim to minimise the chances of losing (using the spiteful approach) in away games.

We next look at how the team performed in the real world when the selected tactics were "close" to our recommendation. In Figure 3.10 we show how the results of teams who use our recommendation in terms of the win, draw and loss percentage. This shows that when teams take the expectimax approach they are more likely to win a game in comparison to the other approaches (0.2% more than the best response approach). Although their results are similar, in comparison to the best response, expectimax boosts the chance of a draw by 1.1% and reduces the chance of a loss by 1.2%.

Finally, we assess the difference between the payoff of the recommended tactics and the actual tactics used across all 760 games. We find that taking the best response approach boosts a teams probability of winning on average by 16.1% and the expectimax approach boosts by 12.7%, while the spiteful approach reduces the chances of losing a game by 1.4%. This shows that, as expected, the best response gives the biggest boost to the

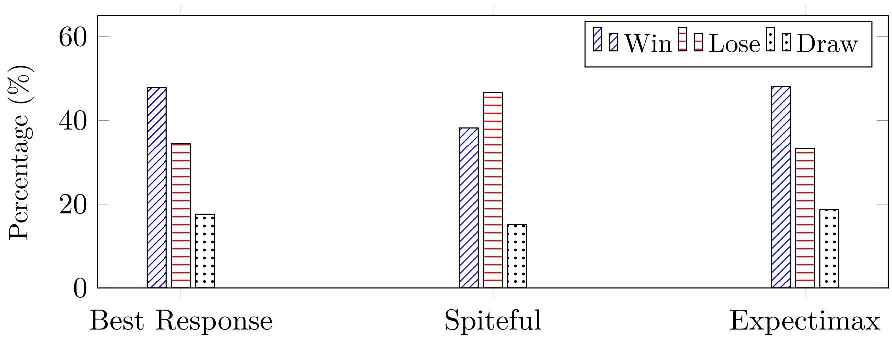


FIGURE 3.10: Percentage of Real-World Results with Close Tactic Selection.

probability of winning a game, although the expectimax approach achieves similar results while also reducing the chances of losing the game.

3.6.4 Experiment 4: Predicting State Transitions

To test the accuracy of the state transition models (one for each game-state) discussed in Section 3.3.2, we compare the model output (home goal, away goal or no goals) to the real-world outcome. We use a train-test split of 70% to 30% with a cross-validation approach for 10 folds. We assess each of the models separately using this approach and on average achieve an accuracy of 87.5% (with a standard deviation of 4.8%), the detailed results for each of the different states are shown in Figure 3.11.

This shows how our models effectively learn the state transition probabilities to high accuracy at each state. The lower scoreline states have more data points over the last two EPL seasons which we use to train and test the models. Therefore, we have a higher certainty over these state transition models in comparison to the ones trained for the higher scorelines that rarely occur in the real world. Hence, we do not show beyond 4-4 in Figure 3.11, but are available to use in our next experiment.

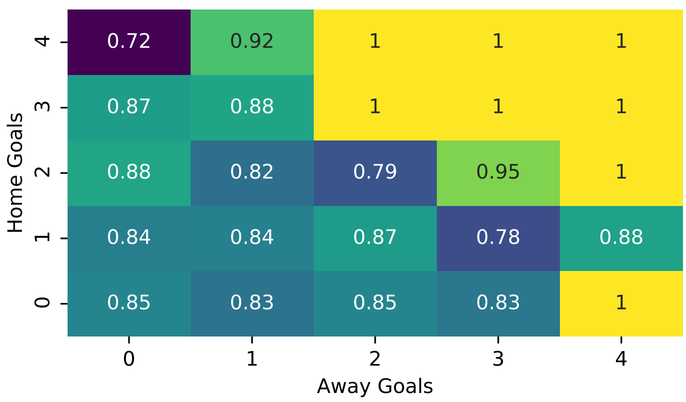


FIGURE 3.11: Heatmap of State-Transition Model Accuracy.

3.6.5 Experiment 5: In-Game Optimisation

When testing the decisions made using the methods, we iterate through all games in our dataset across the two seasons, calculating the payoffs of the actions that both teams can take at each game-state. We compare how often teams took our optimised action in the real world (based on the two different approaches suggested) and if not, evaluate how much our action suggestion would have boosted the team's in-game chances of moving to a more positive state and winning the game.[11]

We first test the action payoff model discussed in Section 3.5.2.2 which uses the state transition probability, substitution and the time of the game to calculate the payoff of the given substitute. By so doing, our model predicts the next state with an average accuracy is 95.5% (standard deviation of 4.5%), tested using a train-test split of 70% to 30% 10-fold cross-validation.

When comparing our action recommendations to those that were taken by the managers in the real world, we find that the aggressive approach makes the same decision 14.75% of the time and the reserved approach makes the same decision 14.11% of the time. If we look at teams making similar player substitutions to our recommendation (selecting a player who plays in the same position as our recommended substitution) then these increase to 40.10% and 39.75% respectively. In Figure 3.16, we show the average payoff for each substitute comparing the real world and our two approaches.

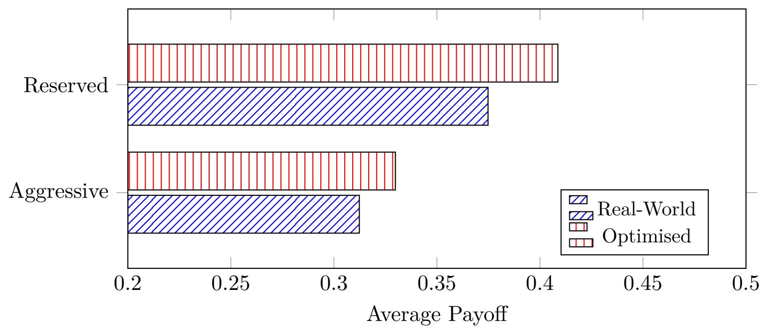


FIGURE 3.12: Payoffs of Real-World vs. Optimised Decisions

On average our more aggressive approach boosts winning payoffs by 2.0% and the more reserved approach reduces the opponents' winning payoff by 3.4%. This shows that the changes in tactics that are made in a game can have an impact on the overall outcome and help teams to move into more positive states or stay in the current state if a team is winning a game. By using the stochastic game we optimise the efficiency of these decisions by 3.4% which could have a significant difference to a team across a season in a game such as football, where every marginal gain can have a large impact on long-term results.

[11] We do not have data for the players that are included as substitutes so we consider all squad players (instead of just the 7 substitutes) which impacts our accuracy.

3.6.6 Experiment 6: Learning the Fluent Objective

Here, we test our fluent objective model in each game week. We evaluate our season simulation prediction model using a Markov-chain Monte-Carlo (MCMC) simulation with respect to its accuracy as the season progresses. To predict the outcome probabilities of individual games in our simulation we use the model defined in Section 3.5.1.2.

In our MCMC simulation, We predict all remaining games 100,000 times and find the most likely league standings after 38 game-weeks. We can compare this to the final league ranks and compare the absolute difference in their actual finishing position and their predicted finishing position. In Figure 3.13, we show an average of all clubs' difference between actual and predicted finishing position (with error shaded area representing 95% confidence intervals). This is run after each game week so we have more information about the games that have already been completed. Week 0 is the prediction before any games have been played and week 37 is the final prediction after 37 out of 38 games have been played.

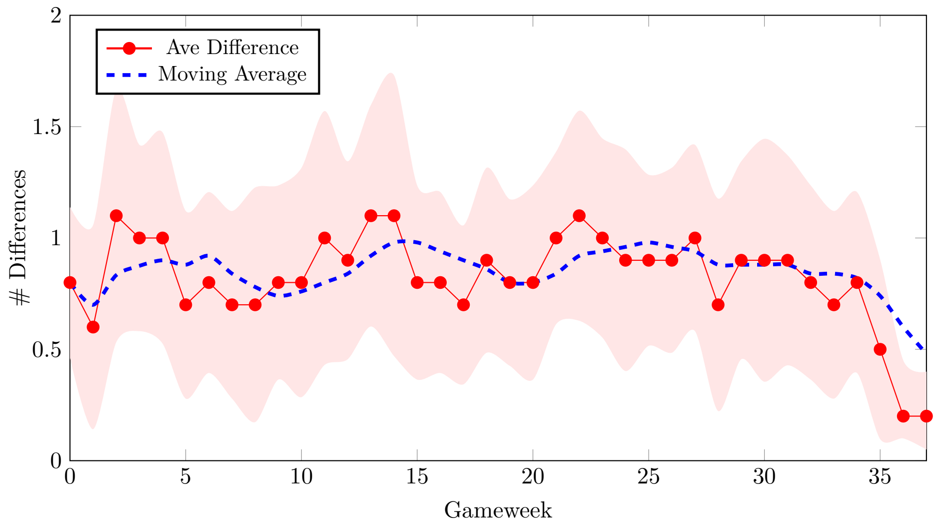


FIGURE 3.13: 2018/19 EPL Average Difference Between Actual and MCMC Predicted Finishing Position.

As shown in Figure 3.13, we can see how in the first half of the season the league standings remain fairly unpredictable due to the number of different possible combinations that we are attempting to predict — there are a total of $2.43e^{18}$ different combinations of team order that the league could finish in.[12] We do see however that as the season unfolds and we have a better idea of team performance the simulation accuracy improves. This is also to be expected as we are simulating fewer games later into the season and we have more evidence from those having taken place in the real world. This shows that

[12]The vast number of possible combination is why we use position differences rather than the overall accuracy of the entire standings after each game-week.

we have a suitable method to extract a distribution of where we expect a team to finish and can derive the fluent objective using a MAP estimation.

3.6.7 Experiment 7: Setting the Fluent Objective

Here we test how effectively we set the fluent objective using our method defined in Section 3.5.3.2. After each game-week simulation we set the fluent objective for all 20 EPL teams and then assess if the objective that is set for the team was achieved at the end of the season or not. In this experiment we show the percentage of teams that met their predicted objective in each game-week. This is shown in Figure 3.14 where the error bounds shaded area recognises the 95% confidence interval, week 0 is the prediction before any games and week 37 is the final prediction.

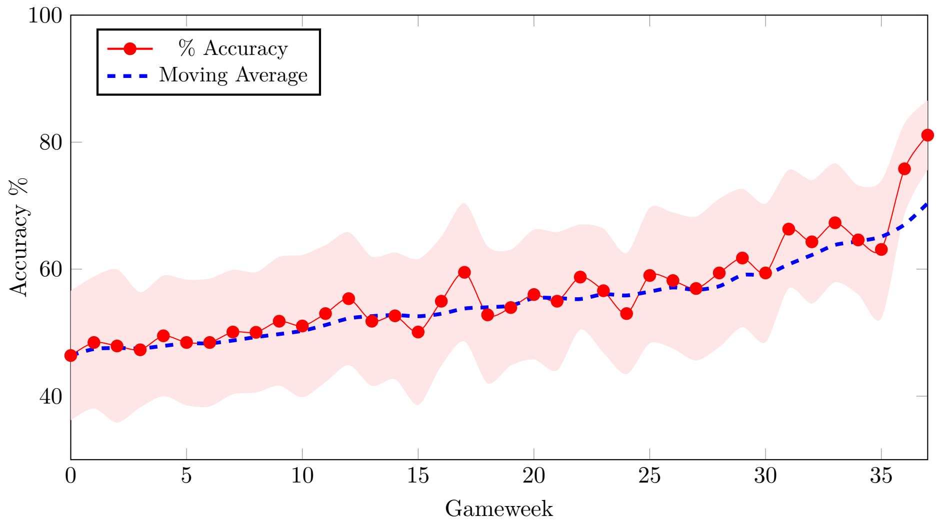


FIGURE 3.14: Week by Week Fluent Objective Prediction Accuracy (2018/19 EPL Season).

As we can see in Figure 3.14 the fluent objective accuracy rises as the season progresses and from week 15 onwards we see the accuracy of the fluent objective setting rise more clearly. This shows that we can set realistic to aim for as the season progresses in relation to the actual league outcomes and what was achieved by the teams. Not every team in the league can meet their objective as there may be more teams aiming for something than can achieve it (e.g., 3 teams aiming to win the league). Also, 3 teams are always, relegated meaning that even in the best case only 85% of teams will achieve their objective. This means in weeks 36 and 37, we reach the maximum teams meeting their objectives.

3.6.8 Experiment 8: Learning from Observing Games

To test the impact of the addition of the weights w that we estimate in P, we evaluate how the weights can boost our ability to predict the outcomes of games based on the tactical decisions and therefore improve our payoff model. To evaluate our P weights, we compare the accuracy of the predictions of the model presented in (Dixon and Coles 1997) both with and without P (this model makes up part of the feature set that is used for calculating the payoffs). We then assess the differences in terms of the models' ability to be able to accurately predict the outcome of the game and show the results in Figure 3.15 (error bars showing 95% confidence intervals).[13]

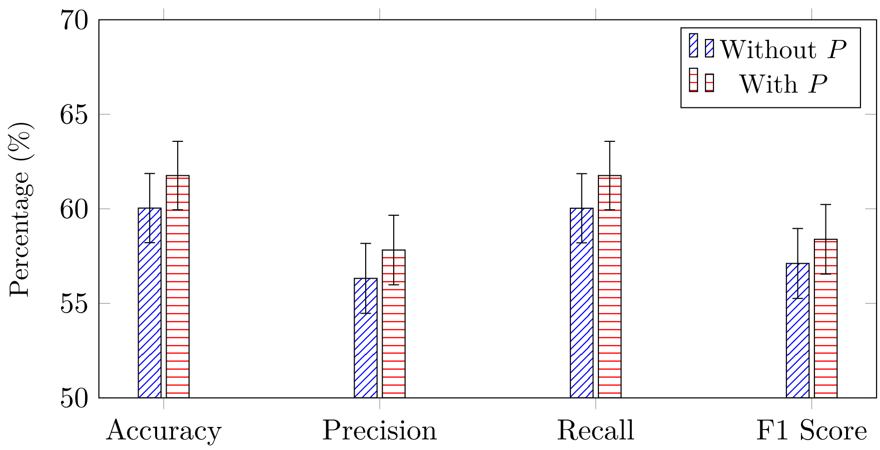


FIGURE 3.15: Payoff Model Performance Comparison.

As we can see in Figure 3.15 by using the weights in P we can boost the accuracy of the model, and therefore the accuracy of our payoffs, achieving a boost of 1.76%. We also see that there is an increase in the precision, recall and F1 of our model by 1.50%, 1.72% and 1.27% respectively. Even though this represents a fairly small increase to the results of the model in Dixon and Coles, it shows that by learning from what tactics have worked (both for your team and others), we can boost our ability to calculate the tactical decision pay-off and therefore our ability to optimise decisions made. Over a large scale of time such as a 38 game-week season, a 1.76% boost in performance could be the difference between finishing a place higher in the league which can have huge financial gain and help to achieve the fluent objective.

3.6.9 Experiment 9: Optimising Team Long-Term Performance

Here, we test how we incorporate the fluent objective O and weights in P into the tactical decision-making optimisation model and evaluate how this improves team performance

[13] The precision, recall and F1 score are computed as a weighted average of the ability to predict each outcome using SciKit Learns' multi-class support.

to help them meet their objective. To test this we simulate an entire season week by week and apply our model to a single team in the simulation. After each game week, we simulate the remaining games and recalculate O and P as outlined in Figure 3.4. We then compare our results using the new model across a simulated season (100,000 times) against a simulation where we do not use the O and P. We show the results as an average from each separate simulation we run for each team, with that team being the only one using the optimisation model. We show the average difference in the mean-expected finishing position from the distribution of each team that we run our season simulation for, both using the new model and without.

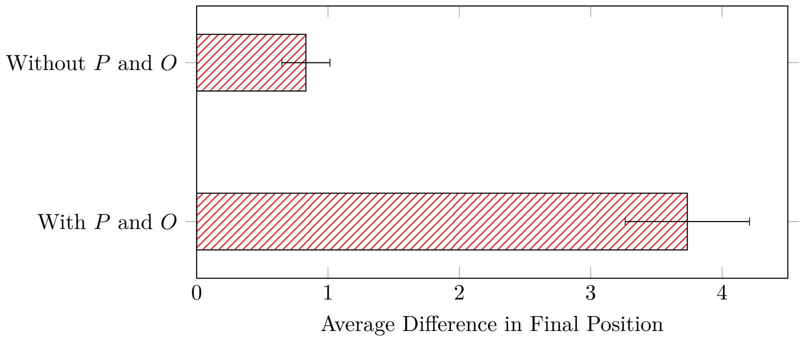


FIGURE 3.16: Payoffs of Real-World vs. Optimised Decisions

This shows how our model can improve the probability of teams' finishing positions and see that on average there is a 2.90 position improvement when using O and P compared to without for our test set of teams. This is achieved as by using O and P teams can add more context to their decisions, also by selecting the optimal tactics each week in the simulation using the pre-game Bayesian model we would also expect to see a boost to the performance. Below, we highlight an example of the distribution improvement of the simulation when aiming to optimise the performance of Southampton FC.[14] Figure 3.17 shows the distribution with O and P applied and not applied.

Figure 3.17 highlights how we can use the fluent objectives to boost their expected finishing position, increasing the mean of the expected finishing distribution by up to 35.6%. We see similar improvements to this across our test set of teams.

3.6.10 Experiment 10: Evaluating Team Monetary Objectives

Up to this point in the chapter, we have focused on using goal-oriented objectives to optimise our long-term tactical decision making regarding winning particular matches or improving the team's finishing position. However, professional football teams are businesses, and one key goal for them is maximising their prize money each season.

[14] Selected as a random example and is the only team using the optimisation model in the simulation.

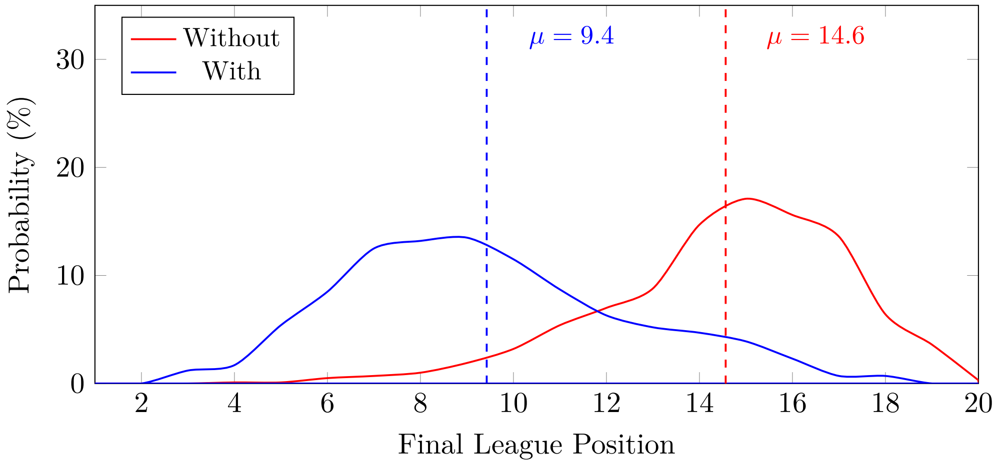


FIGURE 3.17: Example League Outcome Probability Distribution for Southampton FC in 2018/19.

Therefore, in this experiment we use monetary targets for teams to optimise their tactics to increase the chances of maximising the return each season.[15]

By using the prize money data from the English Premier League (focusing on competition prize money and not TV rights money), we can assign a value to a finishing position so that after each simulation of a season we can give a prediction of the money won by a team. To set the fluent objective in this experiment we start by using an increase in the teams' prize money won in the previous season; this is then updated as the season goes on to set higher targets if the team is likely to over-perform. The experiment is run using the same process as the previous experiment, with 100,000 simulations run to test all of the models presented in this chapter across an EPL season. The results for our example team used in the previous experiment are shown in Figure 3.18.

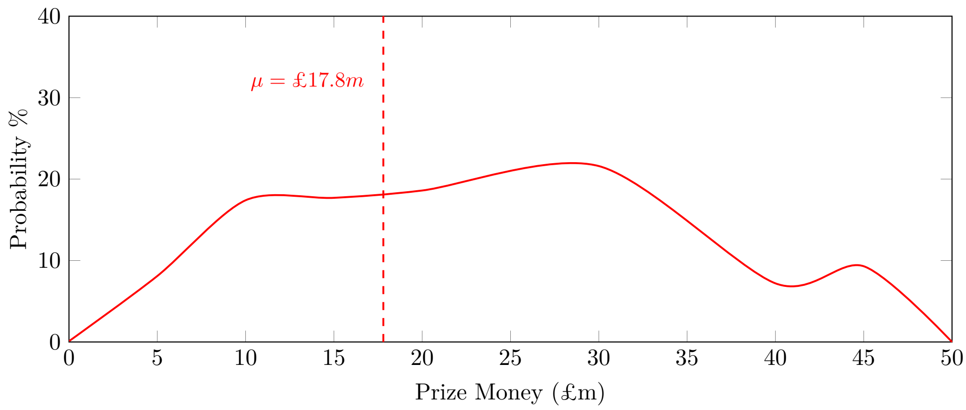


FIGURE 3.18: Prize money distribution for Southampton FC in 2018/19 EPL Season.

This simulation shows a proof-of-concept where various objectives can be set by a team and that by doing so a team would be able to make use of our models to be able to

[15]Monetary data sourced from football business expert: http://swissramble.blogspot.com.

predict and maximise their potential prize money from their league competitions. As Figure 3.18 demonstrates, we can see where the likely prize amounts would be and can use this information to better set the fluent objective and select match tactics. When we compare optimising using the prize money rather than set goal objectives as in previous experiments we see that the expected finishing position in our simulation drops to 10.1 from an average of 9.4 shown for Southampton FC in Figure 3.17 using standard optimisation. However, this is still an increase on the 14.6 average with no optimisation. The slight drop in finishing position when optimising using the monetary values could be due to smaller jumps in prize money in the finishing positions in mid-table. Therefore, when using prize money, it may make more sense to ensure a mid-table finish with reserved tactical approaches rather than using riskier tactics to push higher up the table where if this fails the team would lose more than they may gain.

3.7 Future Work and Wider Applications

Due to the success we have shown when using fluent objectives for an application to football, we intend to test our approach in other domains. For example, they could be used in security games and UAV swarms as the objective also often change over a given time frame. This style of the model could also be applied in security games or for emergency response where we aim to optimise the performance of teams of agents in evolving environments with ever-changing objectives (Ramchurn, Huynh, et al. 2016; Shieh et al. 2012). Further work into wider domains will help to further verify how the modelling of objectives can aid long-term performance.

As well as across new domains, all of these models may be applicable across other team sports games where tactical decisions and team selections are made. The modelling framework we have outlined in this chapter could be refocused, it would be interesting to see the effects of the optimisation to football/soccer in comparison to when applied to basketball, hockey, American football or rugby.

Given better live data streams of games, we could further improve our state-based in-match Stochastic model. This would further help serve coaches as much information in real-time as possible so that they can focus on performance and making optimal substitutions and tactical changes. There is ever-increasing research into the prediction of games and new models (such as (Beal, Middleton, et al. 2021) that uses NLP techniques) which could be used to improve our payoff model. We could explore similar approaches to the reinforcement learning methods used in AlphaGo (D. Silver et al. 2016) to gain a deeper understanding of the in-game dynamics.

We would like to explore how the team make decisions at the board level in terms of investment into the team and staff could help see the simulation of a season change. For example, if a team is likely to have high squad turnover (buying and selling of players) we could use this information to assess each decision a club will make and how this will optimise their performance.

3.8 Chapter Summary

This chapter presents a novel model for making both pre-match and in-match decisions for football and shows how this modelling can be used for multi-step game optimisation as well as optimisations across many multi-step games over a season. For the pre-match Bayesian game setting, we find that we can effectively predict the style and actions of the opposition and then present three models for selecting optimised actions in response. We find that an expectimax approach is the best to take, however in the real world teams tend to go for a spiteful approach. Overall, the Bayesian game model can help real-world teams make effective decisions to win a game and the stochastic game can help coaches/managers make optimised changes during a match.

We also have introduced the concept of a *fluent objective* that allows us to re-evaluate team performance and base decisions based on a wider environment. We find that we can build models that are able to predict the final outcome of the table on a regular basis, and then using a MAP estimation to effectively set the fluent objective each week. We also learn from other games that happen in the overall environment and find that this can boost the performance of pay-off models in our multi-step games. We find that our model can be used for football teams who are looking to improve their overall expected league position (on average improves teams by 2.90 positions) and, show that the concept of a fluent objective can help to optimise long-term performance in a competitive league setting. Overall, this chapter presents novel models that can help teams to optimise their decision-making by using data-driven techniques focusing on the long-term outcomes of performance.

In the next chapter, we begin to explore how the players in the team can work together to contribute to the overall outcome of the games. This builds on the tactical analysis that we have discussed in this chapter by using the extracted teamwork values to help managers/coaches to form teams by maximising teamwork.

Chapter 4

Learning The Value of Teamwork to Form Efficient Teams

In this chapter, we describe a novel approach to team formation based on the value of inter-agent interactions. Specifically, we propose a model of teamwork that considers outcomes from chains of interactions between agents. Based on our model, we devise a number of network metrics to capture the contribution of interactions between agents. This is then used to learn the value of teamwork from historical team performance data. We apply our model to predict team performance and validate our approach using real-world team performance data from team sports games. We test our models for evaluating teamwork in football (soccer) from the 2018 FIFA World Cup and English Premier League, as well as in Basketball using data from two seasons of the National Basketball Association (NBA) in the USA. Team sports present the perfect testing ground for our models of teamwork for both valuing how players work together and then forming teams to optimise this teamwork. Our model is shown to better predict the real-world performance of teams by up to 46% compared to models that ignore inter-agent interactions.

4.1 Introduction

Team Formation (TF) is a fundamental concept that underpins many multi-agent systems where heterogeneous agents with individual properties (e.g., roles, capabilities, costs) come together to undertake tasks. TF involves the evaluation of different sets of agents in order to determine how well they will, individually or collectively, perform their tasks. By so doing, it is then possible to pick sets of agents that form the most effective teams. For example, teams of emergency responders are formed based on individual agents' abilities to navigate a difficult environment or address threats (Chalkiadakis and

Boutilier 2012). Similarly, in ride-sharing settings, groups of riders can be efficiently formed to minimise travel time and costs (Bistaffa, Farinelli, Chalkiadakis, et al. 2017).[1]

Existing TF algorithms, such as (Fitzpatrick and Askin 2005), create effective human teams in the workplace using a mathematical programming formulation and a heuristic solution. Also, in (Scerri et al. 2005) a task allocation algorithm is discussed for extreme teams in disaster response. The models and algorithms in these papers have been shown to be successful in their domains. However, these models typically ignore the fact that sets of agents interact in very specific ways. For example, agents in a team may transfer partly finished products to each other along a production line or a firefighter may secure a building first before sending in a medic. In this chapter, we consider how such *directed interactions* between agents can be valued and considered in the prediction of team performance.

Against this background, in this chapter, we expand on our earlier work (Beal, Changder, et al. 2020) where we propose a novel approach to forming teams using patterns that appear in a network of interactions between agents. We expand on this with further empirical evaluation to validate our approach by applying our models and algorithms to the real-world team formation problems presented by the domain of team sports and in particular football and basketball. Team sports presents us with the perfect live-test bed for our modelling of teamwork. In team sports teams of players work together to win games across long periods of time. Data is collected about how successful teams are and how players are involved which allows us to properly test our machine learning models and optimisations in real-world datasets to solve a real-world problem of evaluating how subsets of players within a team contribute to the overall performance.

We show that our teamwork-focused model outperforms other player-focused approaches at predicting the teams that would be chosen by human expert managers across 64 games from the 2018 FIFA World Cup and 760 games from 2017-2019 in the English Premier League. We also evaluate our models across 2460 basketball games from two seasons of the National Basketball Association (NBA) in the USA (2017/18 and 2018/19). Thus, this chapter advances the state of the art[2] in the following ways:

1. We propose a novel approach to team formation based on the value of inter-agent interactions. Specifically, we propose a model of teamwork that considers the outcomes of the chains of such interactions.

2. Based on our model, we propose a number of network metrics to capture the contributions of individuals and sets of agents.

[1] Note that TF is different from coalition formation in terms of its focus on inter-agent interactions and non-selfish agents.

[2] This chapter further validates the work presented in (Beal, Changder, et al. 2020) by widening the sports used to model teamwork, extracting the chains of interaction and assessing the real-world impact of humans to teamwork that we extract.

3. We show that by using machine learning models we can extract the value of teamwork which can be learnt from data and then applied to the prediction of team performance.

When taken together, our results establish the first benchmarks for team formation based on the learnt value of teamwork. Furthermore, our work opens up a new area of research into the use of teamwork-based models to understand how human teams work.

The rest of this chapter is organised as follows. In Section 4.2 we outline the background to teamwork, while Section 4.3 defines the network model and the optimisation problem. Section 4.4 discusses the application of our model to the sports we use to test and validate, and then in Section 4.5 provides the detail of methods that we use to apply the model to the problems posed by football and basketball. We perform several Experiments on our model in Section 4.6 and discuss our findings in Section 4.7. Finally, Section 4.8 concludes.

4.2 Background

In this section, we explore the related work focused on how agents work together and models of teamwork. We also explore the differences between football and basketball to add further context to the findings in our chapter.

4.2.1 Related Work

There are examples in multi-agent systems literature where teams are formed using analysis of agents within a network. Gatson and DesJardins (Gaston and DesJardins 2005), propose several strategies for agent-organised networks and evaluate their effectiveness for increasing organisational performance. They also present an agent-based computational model of team formation and analyse the theoretical performance in two simple classes of networks: ring and star topologies (Gaston and DesJardins 2008).

Recently, (Bistaffa, Farinelli, Chalkiadakis, et al. 2017) proposed a cooperative game-theoretic approach to deal with the problem of social ridesharing. They first formed a social network representation of a set of commuters, then proposed a model to form the coalition and arrange one-time rides at short notice. The authors model their problem as a Graph-Constrained Coalition Formation (GCCF) (Bistaffa, Farinelli, Cerquides, et al. 2017). Their model is based on two principles, first, they solve the optimisation problem for making coalitions while minimising the cost of the overall system. The set of feasible coalitions in their model is restricted by a graph (i.e., the social network). Secondly, they address the payment allocation aspect of ridesharing.

There are also examples in the literature looking at how humans interact within social networks. The work in (Y. Jiang, Y. Zhou, and W. Wang 2012), explores task and resource allocation in social networks. In their model, the agent that contributes most is rewarded and receives a preferential allocation of new tasks. There has also been work to further explore social networks from a multi-agent perspective (Franchi and Poggi 2012; Y. Jiang and J. Jiang 2013).

Boon and Sierksma (Boon and Sierksma 2003) discuss the design of optimal teams and calculate the value-added from new team members. Following on from this, (Vilar et al. 2013) looks to understand how players' and teams' strategies result in relationships between teammates and opponents in the area of play. There have also been applications to form optimal teams for fantasy football using a MIP and performance predictions in (Matthews, Ramchurn, and Chalkiadakis 2012). Our models differ from the previous work as we model the team as a weighted-directed network of agents and value players based on their influence on the team and the interactions that happen between players. We then form teams using a novel algorithm with MIP techniques.

To our knowledge, none of the discussed approaches has looked at directed interactions between team members (such as passes) and how chains of interactions lead to different team outcome events. More importantly, these approaches have only been validated on synthetic data. Instead, our work is validated on granular data about team performance in real-world games involving teams of humans presented by team sports games.

4.2.2 Team Sports Background

Here, we give a background on the sports we focus on in this chapter. More specifically, we focus our attention on the team sports of football and basketball. These games were selected because of their similar nature, where teams of agents work together with a common aim of outscoring their opponent whether this is goals in football or points scored through baskets in basketball. The teams also work together to prevent their opponent team from scoring. In these games, the way that the players/agents combine and link up is of vital importance as they pass the ball between one another as they progress up the pitch or court. Therefore, this chapter aims to value the importance of this teamwork and compare how it differs between the 2 sports. We also show how this can be used to select the best team of players.

As well as the similarities between football and basketball there are also some key differences in the way the games work that affects the teamwork between the players. In football 11 players are on each team and the game lasts 90 minutes in two half's, during the game only 3 "substitutes" are made. In basketball teams of 5 play against each other in a game lasting 48 minutes across four quarters. There is no limit to the number of substitutions a team can make during a game so the turnover of players is much higher than in football.

In Chapter 2, we present a full review of the key AI work in football and basketball. Some notable examples in football, include the "VAEP" model to value the contribution of individual players in a team (Decroos et al. 2019), a possession value framework is presented in (Fernández, Bornn, and Cervone 2019) and in (Beal, Chalkiadakis, et al. 2020; Beal, Chalkiadakis, et al. 2021) the authors use game theory to optimise football tactics. In basketball, work includes deep imitation learning to "ghost" the movement of players (Seidl et al. 2018), a possession value framework (Cervone et al. 2014) and player decision making is assessed in (Zheng, Yue, and Hobbs 2016b).

4.3 A Formal Model for Valuing Teamwork

Our TF model is based on our observations of many real-world team-based systems, such as sports teams or teams of emergency responders, as follows:

- Many teams operate through directed (one-to-one, one-to-many, or many-to-many) interactions. For example, in a sports team, a player would pass a ball to another. However, team members will not always interact equally with every other team member. In this chapter, as a first step, we will focus on one-to-one interactions. Indeed, we show that such a setup gives rise to complex interactions that pose difficult computational challenges.[3]
- Team members may have different roles and abilities to perform tasks. For example, emergency response teams will have members with different skill sets, equipment, and training. In a football team, each player will have a position on the pitch and specific abilities.
- Team actions can have multiple consequences. In the simplest case, they may have binary outcomes (succeed in achieving a mission or failing to do so). In many cases, however, team success is more nuanced (e.g. the achievement of a sub-ideal goal).
- Team formation typically involves picking a subset of agents that work well together, using some metric of efficiency. For example, emergency responders will choose a subset of available partners that are most fitting to the task or have the right skills. Similarly, a football manager will pick the best team (measured by their likelihood to win a match) of 11 players out of a squad of 23.

[3] We will consider more complex forms of interactions in future work such as out of possession defensive teamwork when the ball is not involved.

As can be seen, choosing the optimal team can be a difficult task given the complexity of the roles and relationships among team members as well as the environment they evolve in. In what follows, we formally define the key constructs of our teamwork model and devise multiple network-based metrics over which the value of teamwork can be learnt and used to predict the performance of different teams.

4.3.1 Basic Definitions

We define the set of n agents as $\mathcal{A} = \{a_1, a_2, \ldots, a_n\}$. Agents can interact with each other to achieve some overarching goals. We consider such interactions to be directed (e.g., a UAV allocating a task to another UAV or a player passing a ball to another player in a football team). We define the set of interactions I as a set of ordered pairs (a_y, a_z), where a_y initiates the interaction towards a_z. The same pair of agents may engage in such interactions multiple times, and hence I is effectively a multi-set of interactions.

Agents' interactions are constrained by a directed graph representing potential roles and relationships between the agents. In the general case, all nodes in the graph will have at least two directed edges (one outgoing, and one incoming) between them and another agent. We denote this graph as $G = (\mathcal{A}, \mathcal{I}, w(.))$ with agents $\mathcal{A}$ as vertices and edges $\mathcal{I}$ representing relationships between pairs of agents. The weight on each edge is the number of directed interactions from the set I between pairs of agents. For example, for each edge $i \in \mathcal{I}$, the weight of i, $w(i)$ is defined as the number of times the edge $i = (a_x, a_y)$ appears in the multi-set I (the multiplicity of i in I). Formally, $w(i) = \sum_{s \in I} \mathbf{1}_{\{i\}}(s)$. This is an iteration over all elements of I and if $i = s$ a value of 1 is added (equivalently, if $i \in I$) and 0 otherwise. We build on these definitions to model how interactions between agents result in specific events.

4.3.2 Modelling Chains of Interactions

In many situations, agents will interact sequentially with each other (i.e., agent a_x interacts with a_y who in turn interacts with a_z). In this chapter, we only consider the cases where an agent interacts with one other agent at a time.[4] To this end, we define a walk in the graph G as a sequence of interactions over the edges of the graph. Formally, a walk $\mathcal{P}$ of length l in the graph G from vertex a_y to vertex a_z ($a_y \rightarrow a_z$) is a sequence $[a_x, a_y, \ldots, a_z]$ where $l = |[a_x, a_y, \ldots, a_z]| - 1$. An example of a walk $\mathcal{P}$ is described for football in Section 4.4.1 and this is shown in Figure 4.2. Another example of a walk in a real-world application would be the movement of a data packet through a mesh network where the packet moves from router to router until it reaches the destination.

[4]We will consider cases where this assumption does not apply in future work. For example, a ship may deploy many sub-ships that have all been allocated the same task to complete.

A walk leads to an event of a specific type. For example, a data packet being used to complete a file download, or a football player scoring a goal at the end of a sequence of passes. There may be many different event types. Formally, the set of possible events $\mathcal{E}$ is defined as $\mathcal{E} = \{e_1, e_2, \dots, e_k\}$ where e_κ is the event and k is the number of possible events from the walk.

Each of the possible events e_κ may have a different impact on the environment, therefore affecting the overall performance of the team. Thus, for each $e \in \mathcal{E}$, the value function $\alpha : \mathcal{E} \rightarrow [0, 1]$ determines an associated value. For example, in a game of football, if the event e_κ is a "goal" event, this has a bigger impact on the overall outcome of the game and team performance in comparison to if e_κ is a "loss of possession" event. Another example could be walks leading to a "person saved" event, in an emergency response setting, having a greater impact than walks not leading to such events.

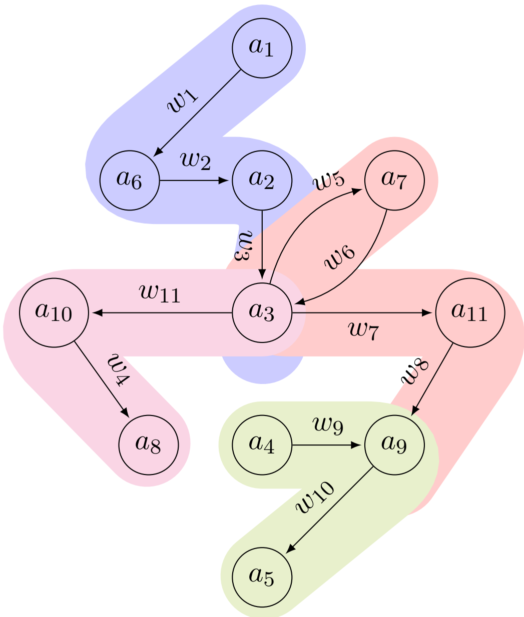


FIGURE 4.1: An example of 4 walks through a sample graph of 11 agents for an event e_k. The directed edge between two vertexes represents the interaction between them and each highlighted colour represents a walk.

Note that each walk originates from one agent and involves chains of directed interactions between pairs of agents, resulting in an event. Hence, we next propose methods to extract the contribution of each agent as well as sets of agents to these individual events.

4.3.3 Extracting the Value of Teamwork

Walks $\mathcal{P}$ and associated events $\mathcal{E}$ can be used to infer the value of agents and sub-sets of agents within the team. We propose three metrics to value the contribution of individual agents and sets of agents to the given outcome event as follows:

- Centrality: $v_{cent} : \mathcal{A} \rightarrow \mathcal{R}$ refers to the sum of the weight of edges incident (incoming and outgoing) to a_i. This measures the influence of agent a_i in the network.
- Distance from event: $v_{dist} : \mathcal{A} \rightarrow \mathcal{R}$ defines the average distance of agent a_i for each event. This represents the influence of agent a_i on an event.
- Walk frequency: $v_{freq} : 2^{\mathcal{A}} \rightarrow \mathcal{R}$ refers to the number of times an agent a_i or a subset of agents appears in all walks. This represents the influence of an agent in the team.

It is important to note that these metrics attempt to summarise team performance in different ways, each with a different degree of information loss. Using centrality results in the most loss of information as it ignores whom the interactions are made with. Using distance from the event (i.e., last node in a walk) better associates agents to events but also ignores the specific interactions that result in such events. Finally, walk frequency considers all pairwise interactions that lead to specific events, and as we show later, is more representative of teamwork and can be used to predict the performance of teams more effectively.

Now, for each event, we will have different values for each metric for each agent or set of agents (as for walk frequency). However, each event has a different impact on team performance (e.g., goals lead to a win, loss of possession likely to lead to a loss, person saved leads to a successful rescue mission), and to determine the contribution of an agent or subset of agents to team performance, we need to learn the impact each metric has on the team's performance. We assume that each event is independent[5] and therefore use a weighted sum of the values for each of the possible events. This is shown in Equation 4.1.

$$v_m(a_i) = \sum_{k=1}^{K} \alpha_k v_m(a_i|e_k) \tag{4.1}$$

where, $v_m(a_i)$ is the value of a_i using the metric m, K is the number of possible events, α_k is the weight of the event e_k (which is learned from the data) and $v_m(a_i|e_k)$ is the value of a_i given the event e_k. We next expand on the above metrics in the following sub-sections.

[5] In reality, some events may not be entirely independent and therefore, more complex summarisation functions would need to be used. But as we show in this chapter, the assumption of independence does hold when it comes to predicting team performance.

4.3.3.1 Network Centrality

Here we value an agent a_i based on it's centrality in the network. This value is equal to the sum of the weights of the edges incident to node a_i (both incoming and outgoing edges). For example, in the graph shown in Figure 4.1, $v_{cent}(a_9|e_k) = w_8 + w_9 + w_{10}$. Equation 4.2 shows the value calculation using the centrality metric for any agent a_i for the given event e_κ:

$$v_{cent}(a_i|e_\kappa) = \sum_{a_j \in \text{Adj } (a_i)} w(a_i, a_j) + w(a_j, a_i) \tag{4.2}$$

4.3.3.2 Distance From Event

Given a set of events $\mathcal{E} = \{e_1, e_2, \ldots, e_k\}$ and all the possible walks in graph $g \in G$. The value for an agent a_i for any event $e_\kappa \in \mathcal{E}$ is defined as the average of the shortest path length of agent a_i from the event e_i for each walk where the agent a_i is present. The distance from the event of an agent a_i is the number of agents following a_i in a walk. We can define this as $[a_i, g_l]$ where a_i is the occurrence of the agent in walk g and g_l is the final agent in the walk closest to the event. For example, in Figure 4.1 in the walk $[a_4, a_9, a_5]$ the distance from event for a_9 is 1 and in the walk $[a_7, a_3, a_7, a_9, a_{11}, a_3]$ this distance for a_9 is 2. Hence, the average distance of agent a_9 from the event is $(1+2)/2 = 1.5$. This is formalised in Equation 4.3 where we find all the lengths from events for a_i in G and the occurrences of a_i in G $(sum_{g \in G} 1_{\{g=a_i\}}$.

$$v_{dist}(a_i|e_\kappa) = \sum_{g \in G} \left(\frac{[a_i, g_l]}{1_{\{g=a_i\}}} \right) \tag{4.3}$$

4.3.3.3 Walk Frequency

The walk frequency of a set of agents (singleton, pairs, triplets etc.) $A' = [a_x, a_y, ..., a_z]$ is the number of times the agent A' appears in all the walks in G. This is formalised in Equation 4.4. For example, in Figure 4.1, suppose the walk $[a_4, a_9, a_5]$ appears three times, the walk $[a_7, a_3, a_7, a_3, a_{11}, a_9]$ appears four times, the walk $[a_3, a_{10}, a_8]$ appears twice and the walk $[a_1, a_6, a_2, a_3]$ appears once. In this case the value of $A' = [a_9]$ will be $v_{freq}(A'|e_k) = 3 + 4 = 7$. The same reasoning can be applied to subsets of agents.

$$v_{freq}(a_i|e_\kappa) = \sum_{g \in G} 1_{\{g=a_i\}} \tag{4.4}$$

We can compute such a metric for all individual agents as well as different sub-groups of agents from pairs to above formed as chains of interactions (i.e., subsets of a walk). Given a walk $\mathcal{P}$ of length l, the sub-groups of length j that we are able to extracted from the walk $\mathcal{P}$ is calculated by picking the consecutive $j+1$ vertices in the walk $\mathcal{P}$. The total number of sub-groups of agents in a walk of length l is $\sum_{j=1}^{l-1}(l-j+1)$. In this chapter, we mainly focus on sub-groups involving pairs of agents as combining such pairs in a combinatorial optimisation algorithm to consider chains of interactions. We next describe how we will learn the weights of events α_k to compute Equation 4.1.

4.3.4 Learning Event Weights From Data

To learn the set of weights $\mathcal{D}$, which correspond to the impact of the possible walk events $\mathcal{E}$, we use a Logistic Regression algorithm (Hosmer Jr, Lemeshow, and Sturdivant 2013). This allows us to extract the coefficient weights of each of the input features i.e., the weight α_κ (which corresponds to an event e_i) which is used to calculate the final value $v_m(a_i)$ for each agent or sub-team of agents.

Hence, for an outcome y (e.g., a team wins a match, a political party wins an election), the probability that an agent a_i contributes to this outcome is dependent on the individual events (e_κ) to which an agent contributes, as captured by the metrics computed in the previous section.

The result of running the logistic regression algorithm to extract the event impact is the set of weights $\alpha_\kappa \forall e_\kappa \in \mathcal{E}$. Given this, we can now compute efficient teams according to the learnt measures.

4.3.5 Forming Efficient Teams

We use two methods to form efficient teams using values calculated in the previous section. Firstly, we form teams based on the values of singleton agents. Secondly, we form teams based on the value of agent pairs p, so that teams are formed between agents who communicate and work well together.

4.3.5.1 Agent-Centred Approach

To form the efficient team based on singleton agents, we use the values $v(a_i)$ for each agent a_i. Given constraints on the number of agents to be picked overall and the number of agents per role allowed in the team (see Section 4.5), this results in a combinatorial optimisation problem that is solved using standard mixed-integer programming (MIP) techniques. Similar methods are also used in (Pochet and Wolsey 2006; Fitzpatrick and

Askin 2005; Matthews, Ramchurn, and Chalkiadakis 2012). Here we can use all the above metrics m (i.e., centrality, distance from the event, and walk frequency).

4.3.5.2 Team-Centred Approach

Here we consider how the team works effectively and hence only consider the walk frequency metric. Specifically, we reconstruct the value of teamwork based on two core concepts which we call the *strength of teamwork* and *interactional alignment* which we describe as follows.

- **Strength of Teamwork:** This is based on the contribution of the pairwise interactions, which in this case is shown by a high frequency of directed successful interactions between the agents. This can be calculated using the 3 methods (centrality, distance and frequency) that are discussed in Section 4.3.3.
- **Interactional Alignment:** This is the measure of the strength of teamwork between overlapping pairs within the selected team. This values the strength of teamwork that the selected agents in a pair will bring when paired with other selected agents in the team. This helps us avoid selecting pairs of agents that have a strong value between themselves but are weak when combined with the rest of the team. We calculate this using Equation 4.6.

We combine these two measures to maximise the values of the selected pairs (p_i) while also maximising the value of the pairs that they overlap within the selected team as a whole. Specifically, we propose a MIP defined by Equation 4.5. In more detail, the output of the MIP is a team of N agents formed by evaluating pairs of agents from the set $\mathcal{O} \subset 2^{\mathcal{A}}$ where for each $p \in \mathcal{O}, p \subset \mathcal{A}, |p| = 2$. We use two types of binary decision variables $x_i, z_j \in \{0, 1\}$ for pair i and agent j respectively. Variable x_i denotes whether a pair of agents is selected and z_j whether an agent is selected. The objective function maximises the sum of $V(p_i)$ (the value for p_i using the agent pair values we have calculated) and $V'(p_i)$ which represents the *interactional alignment* (the value of pair p_i calculated by Equation 4.6). This value is weighted by β which can be learnt from the data. The first constraint ensures that individual agents can be selected, even if they are in pairs that are not selected. The agent decision variables z_μ and z_λ represent the two agents in a given pair $p_i = \{a_\mu, a_\lambda\}$. The second constraint ensures that only N

agents are selected from all agents $\mathcal{A}$ available.

$$\begin{aligned} \text{maximise} \quad & \sum_{i=1}^{|\mathcal{O}|} \left(V(p_i) \cdot x_i + \beta V'(p_i) \right) \\ \text{subject to} \quad & z_\mu \geq x_i, z_\lambda \geq x_i, \forall p_i = \{a_\mu, a_\lambda\} \\ & \sum_{j=1}^{|\mathcal{A}|} z_j = N \end{aligned} \tag{4.5}$$

$V'(.)$ is defined by Equation 4.6 as the sum of all pair values where there is an overlap with pair p_i. By maximising the *interactional alignment*, this allows us to increase the strong links between pairs while decreasing the weak links.

$$V'(p_i) = \sum_{k=1}^{|\mathcal{O}|} \left(V(p_k) \cdot x_i \right)_{\{p_i \cap p_k, k \neq i\}} \tag{4.6}$$

The generic solution presented in Equation 4.5 could also be expanded to consider the notion of roles within the team structure. An example of this would include the formation of a team in football where only one goalkeeper can be selected and the rest of the players are selected with different tactical roles. Similarly, within the emergency response domain we may need to form a team made up of specialists in different areas (e.g., paramedic and fire service). If roles are an important element to the team formation problem, we add extra constraints to our MIP formulation this is shown in Section 4.5.2.

4.4 Model Application to Team Sports

To validate the models defined in Section 4.3 we apply our techniques to the problem of team formation in football and basketball. In this section, we highlight how team sports relate to our model and how they can be applied.

4.4.1 Football Application

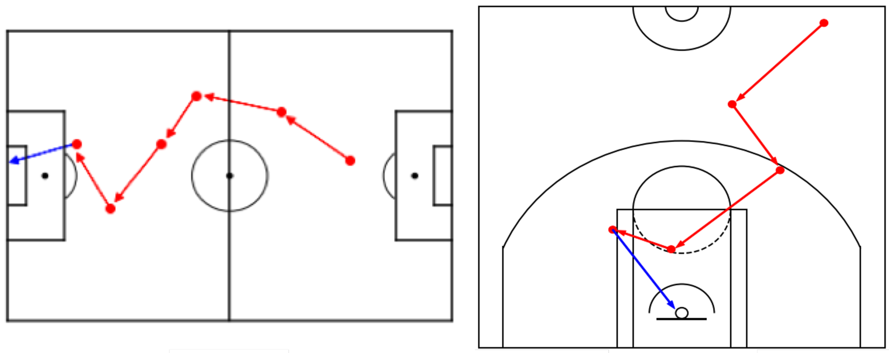
In football, a manager/coach selects a team of 11 players from a squad of 23 (sometimes more depending on the competition rules). The objective is to select a team with the highest chance of winning a game. Against this background, we define the squad of players as our set of agents $\mathcal{A}$, the interactions $\mathcal{I}$ are the passes between the players in earlier games, and the graph G represents the network of passes between all the players in the squad. The walk $\mathcal{P}$ is a passage of play for the team which is made up of several passes. In football, a passage of play is ended by some event (e.g., tackle, shot, goal, miss, and ball out of play). We characterise events into 4 possible outcomes, $\mathcal{E} = \{e_1, e_2, e_3, e_4\}$, where e_1 is a Goal, e_2 a shot on-target, e_3 a shot off-target and e_4 is a loss of possession. We then learn the weights α_i for each outcome. In this case we assume $\alpha_1 > \alpha_2 > \alpha_3 > \alpha_4$. Using the model discussed in the last section we calculate the value of each player $v(a_i)$ and form an optimal team based on the values considering the specific positional constraints of a football team. An example of a walk is shown in Figure 4.2 where the red arrows represent the passes between players and the blue arrow represents the outcome of the walk which in this case was a shot on target.

There are positional constraints that are specific to football, making it more complex than the model we defined in Section 4.3. Each team in a game of football must have 1 goalkeeper and 10 outfield players which are formed from defenders, midfielders, and strikers. In most positional formations in football, there are between 3-5 defenders, 3-5 midfielders, and 1-3 strikers. An example formation is 4-4-2 which is 4 defenders, 4 midfielders, and 2 strikers.

4.4.2 Basketball Application

In basketball, a coach selects a team of 5 players from a squad of 12 and again the objective is to select a team with the highest chance of winning a game. Similarly to the previous subsection for football, we define the squad of players as our set of agents $\mathcal{A}$, the interactions $\mathcal{I}$ are the passes between the players, and the graph G represents the network of passes between all the players in the squad. The walk $\mathcal{P}$ is an attack of 5 players who are currently on the court which is made up of several passes. This play is ended by some event ($\mathcal{E} = \{e_1, e_2, e_3, e_4\}$) which is made up from a turnover, a 3 point shot, a 2 point shot or a foul. We then learn the weights α_i for each outcome and again using the model discussed in the last section we calculate the value of each player $v(a_i)$ and form an optimal team. An example of a walk in basketball is shown in Figure 4.2 where the red arrows represent the passes between players and the blue arrow represents the outcome of the walk which in this case was a 3-pointer.

In basketball, each of the 5 players has a different position. These are shooting guard, point guard, centre, power forward, and small forward. These present our positional constraints when aiming to form the optimal team as we want to aim we select a team made up of the players that can fill these positions while maximising the teamwork between the players.



(a) Football (b) Basketball

FIGURE 4.2: Example of a Walk in Football and Basketball

4.5 Forming Efficient Sports Teams

In this section, we describe the techniques that we use to solve our model and form efficient teams.

4.5.1 Calculating Player Values

To calculate the values of the players in the network, we first create the weighted graph that we need for our model. We do this using the walks (patterns of play) which happen in a game. We can then calculate their values for each of the possible walk events using each of the metrics defined in Section 4.3. The possible walk events we use are:

- **Football**: a goal, a shot on target, a shot off target and a lost possession.
- **Basketball**: 3 point basket, 2 point basket, foul and loss of possession/turnover.

We first do this for singleton players so that we have values based on their centrality, walk frequency and distance from the outcome. We then value the player pairs based on their frequency in the network. This gives us the values for both players and pairs from each match which we can then use to learn the impact weights of the outcomes. We can also extend this to look at longer chains of agents and evaluate their contributions as a three or a four etc.

4.5.2 Learning The Outcome Weights

To calculate the weights of the walk events we use logistic regression as discussed in Section 4.3.4. Using the values for the players/pairs for each walk event in each game we use the match outcome (team win, loss or draw) as the y value in our logistic regression formula. This means that we train the model to calculate the weights based on what impact it will have on the match outcome and, therefore, the overall team performance. The final value for the players/pairs will then be a weighted sum (defined in Equation 4.1) which uses these learned weights and will inform the team formation process.

4.5.3 Team Formation

We describe the two methods we take to form teams using both the singleton player values and the pair values.

4.5.3.1 Singleton Agents

The first method uses the values of singleton players calculated using the centrality, walk frequency and distance from the outcome (as discussed in Section 4.3). We use these values alongside constraints over players' positions to form the optimal team. The approach we use to solve this is an edited version of the MIP approach shown in Equation 4.7. Where we maximise $\Sigma_{i=1}^{\mathcal{O}}(V(a_i) \cdot z_n)$ and do not consider the pair decision variable x_i. The other constraints remain the same. This approach will help us to form a team of players who all contribute but may not link up well together as a team. We, therefore, expand on this method in the next subsection to form teams using player pairs.

4.5.3.2 Agent Pairs

Using the values of the player pairs we form teams using the MIP formula presented in Equation 4.7 (this is a refinement of Equation 4.5). When forming teams we ensure that all the pairs of players are part of the same squad and can be selected together. We also consider the positions of the players so that we pick a team in a reasonable positional formation. This is represented by position range constraints.

$$
\begin{aligned}
\text{maximise} \quad & \Sigma_{i=1}^{|\mathcal{O}|}(V(p_i) \cdot x_i + \beta V'(p_i) \cdot x_i) \\
\text{subject to} \quad & \Sigma_{n=1}^{|\mathcal{A}|}(z_n) = 11 \\
& z_\mu \geq x_i, z_\lambda \geq x_i, \forall p_i = \{a_\mu, a_\lambda\} \\
& \Sigma_{n=1}^{|\mathcal{A}|}(gk_n \cdot z_n) = 1 \\
& 3 \leq \Sigma_{n=1}^{|\mathcal{A}|}(def_n \cdot z_n) \leq 5 \\
& 3 \leq \Sigma_{n=1}^{|\mathcal{A}|}(mid_n \cdot z_n) \leq 5 \\
& 1 \leq \Sigma_{n=1}^{|\mathcal{A}|}(str_n \cdot z_n) \leq 3
\end{aligned}
\tag{4.7}
$$

where a binary decision variable x_i represents the selected pairs ($p_i = \{a_\mu, a_\lambda\}$) and z_n represents whether a player is picked or not. There is then a number of binary sets for each position (*gk*, *def*, *mid* and *str*) containing if a player plays in the corresponding position or not and we aim to maximise the pair values V and V' in the selected team.

For basketball, this same method can be tweaked slightly. Instead of filling the constraints shown in Equation 4.7, in basketball, there would be custom constraints to fill the positions: shooting guard, point guard, centre, power forward and small forward. Other than this the technique to form the optimal team would remain the same.

4.6 Empirical Evaluation

In this section, we discuss the experiments that have been used to test and evaluate our models. We discuss the data that we use for football and basketball which are industry-leading datasets used by top professional teams in the games. Thus, these rich real-world datasets allow us to rigorously assess the value of our model. It is worth noting that the initial testing for these models can be found in (Beal, Changder, et al. 2020).

4.6.1 Data

To evaluate the models that we have discussed in this chapter, we use a dataset for football and a dataset for basketball. The football dataset was collected from two seasons (2017/18 and 2018/19) from the English Premier League (EPL) as well as data from the 2018 FIFA World Cup.[6] The dataset contains 784 games that we can evaluate our football model on. The dataset breaks down each of the games from the tournament into an event-by-event analysis where each event gives different metrics including event type (e.g., pass, shot, tackle etc.), the pitch coordinates of the event and the event outcome. To learn the model weights, we use a 10-fold cross-validation approach, splitting the dataset randomly into 70% training and 30% test.

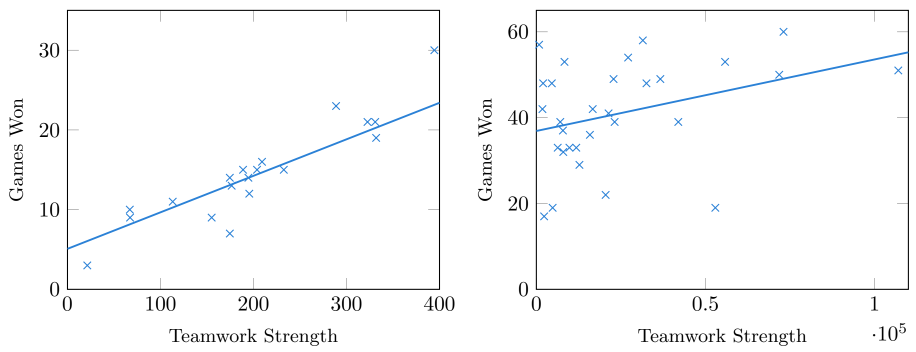
The basketball dataset is extended play by play data from two seasons of the National Basketball Association (NBA) in the USA (2017/18 and 2018/19) containing 2460.[7] There are more games here due to NBA teams playing 82 games in a regular season compared to 38 played by EPL teams. Again, this dataset allows us to break down basketball games by each play in the game and see what players have been involved in.

The experiments[8] performed are outlined in the sub-sections to follow.

[6]Football data provided by StatsBomb - www.statsbomb.com.
[7]Basketball data provided by EightThirtyFour - https://eightthirtyfour.com/data.
[8]Tests have been run using Scikit-Learn and TensorFlow.

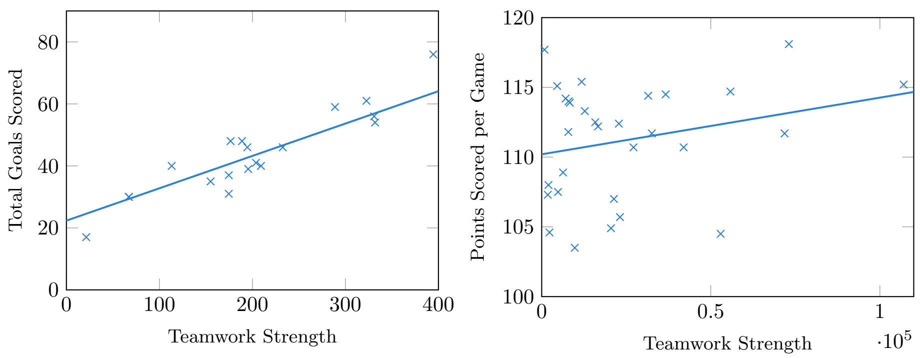
4.6.2 Experiment 1: Comparing Teamwork Values Across Sports

In this experiment, we compare how teamwork values in the different sports correlate to real-world performance metrics. We first look at how the sum of teamwork values correlate to the number of games won by teams, the results from this are shown in Figure 4.3. We then look to compare the teamwork value to the number of goals scored and for basketball comparing the values to the number of points scored. The results from this are shown in Figure 4.4. This experiment allows us to see how the teamwork values that we extract from the data correlate to how we expect the teams to perform in the real world.



(a) Football (b) Basketball

FIGURE 4.3: A comparison of teamwork values and the number of games won in a season.



(a) Football (b) Basketball

FIGURE 4.4: A comparison of teamwork values and the number of goals/points obtained in a season.

As we can see from the figures, there is a positive correlation between the number of wins and the number of points/goals scored in both football and basketball. The Spearman's rank correlation coefficient for football are 0.92 and 0.89 for games won and goals scored respectively. For basketball, the coefficients are 0.35 and 0.15 for games won and points scored respectively. We also see there is a stronger fit in football with r-squared values of 0.84 and 0.87, whereas basketball is -0.12 and -0.06.

We believe that we see a better fit and a more positive correlation in football than basketball for these metrics due to the more consistent state of the team and players on the pitch in football. As only 3 substitutions can be made in a game in football, players play together for longer and build better relationships. This is especially the case due to the formation of the 11 players in football meaning pairs of players must learn how to play together in the team (e.g., a pair of strikers trying to score goals, a pair of centre-backs defending or a full-back and winger dominating one side of the pitch). It could also be to do with the high scoring nature of a game of basketball, meaning that the teamwork values are less important for the many points scored and instead it is caused by individuals efforts in attack. Whereas in football a goal is rare and the build-up to a goal relies on a cohesive team working together.

Due to this, we focus our more extensive evaluation of the teamwork model on the data presented by football in the rest of this section.

4.6.3 Experiment 2: Performance Comparison to Teams Formed by Human-Experts

We evaluate our model to compare both the singleton approach and the pair's approach with the teams selected by the human-expert manager (focusing on both the starting 11 players and the 11 players who finish the game after substitutes). We form teams based on the method in Equation 4.7 to maximise teamwork and then compare this to the team selected in the real-world. The results are presented in Figure 4.5 (where error bars represent a 95% confidence interval).

- **Model 1:** Centrality Value.
- **Model 2:** Walk Frequency.
- **Model 3:** Distance from Event.
- **Model 4:** Pair Values (Equation 4.7).

This shows that the pair values optimisation method gives the closest teams to the human experts on average with a difference of 2.3 per game for the starting team. This suggests that the human managers (either consciously or subconsciously) consider the ability of players in the team to work together as the other methods only consider individual player values. At an average of 2.3, this could give managers suggestions of how changes could be made to the team that may give a better chance of winning the game.

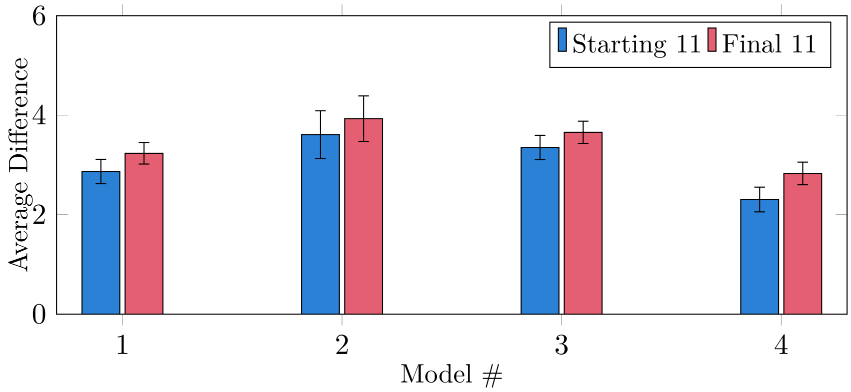


FIGURE 4.5: Average Difference Between Model and Real-World Human Manager Selections (where lower is better).

4.6.4 Experiment 3: Match Outcome Prediction

The results shown in Experiment 1 suggest that in football there is a strong positive correlation between team performance our the teamwork metrics we present in this chapter. Therefore, leading on from this experiment we further explore the predictive ability of our models by using their outputs to predict the likelihood of winning games and several given performance metrics. In Figure 4.6 we show the percentage accuracy when we predict the winner of the game based on which team has the highest teamwork sum. In Table 4.1 we show the RMSE results from using teamwork to train predictive model for given performance metrics.

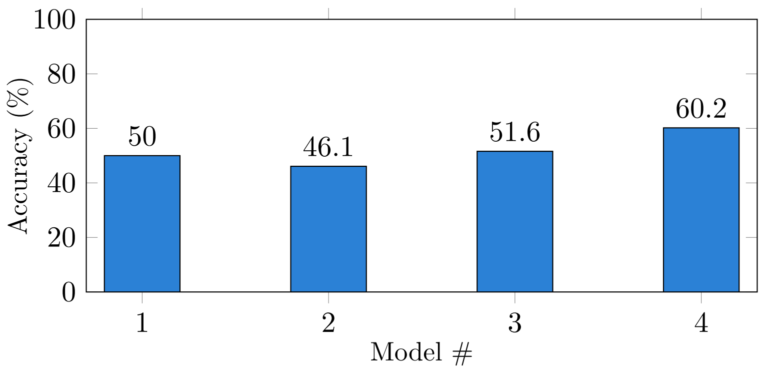


FIGURE 4.6: Accuracy of Valuation Methods For Outcome Predictions.

Model #	Individuals	Pairs
Shots	4.33	**3.74**
Goals For	1.00	**0.87**
Goals Against	1.14	**1.06**
Passes	105.69	**57.07**

TABLE 4.1: Valuation Methods Root Mean Squared Errors for Performance Metrics (where lower is better).

The results in Figure 4.6 and Table 4.1 suggest that the teamwork metric is a more accurate predictor of performance than individual player values, meaning that the teams with higher valued pairs are more likely to win the game and have better performance indicators. This is especially true when we predict the number of passes that a team will make in a game as this metric shows the strongest predictor when using the teamwork values and is a 46% increase on any other approach.

We also see that in basketball teams with higher teamwork value win in 52.9% of games, compared to 60.2% in football.

4.6.5 Experiment 4: Expanding the Chains of Interaction

When teamwork valuation has been tested in prior work (Beal, Changder, et al. 2020), the paper only assesses the value of single agents and pairs of agents and does not consider chains of interactions longer than 2 agents when valuing the contribution of the agents and when forming a team. Therefore, in this experiment, we assess the effects of expanding the chains of interactions and looking at the value of triplets of players and beyond. We aim to see how the longer chains will correlate to the real-world metrics that are discussed in the previous experiment. We found that these values have a stronger correlation in football (over basketball), therefore in this experiment we choose to just focus on football. We compare how the differing lengths of chains of interactions impact the real-world correlations, time to form teams (Section 4.5.3.2) and the closeness to real-world selections by human expert managers/coaches.

We first focus on how the different lengths of chains affect the correlation to the number of goals and wins a team obtains over a season. We show how the Pearson Correlation Coefficient (PCC) changes for these metrics at different chain lengths. The results from this are shown in Figure 4.7.

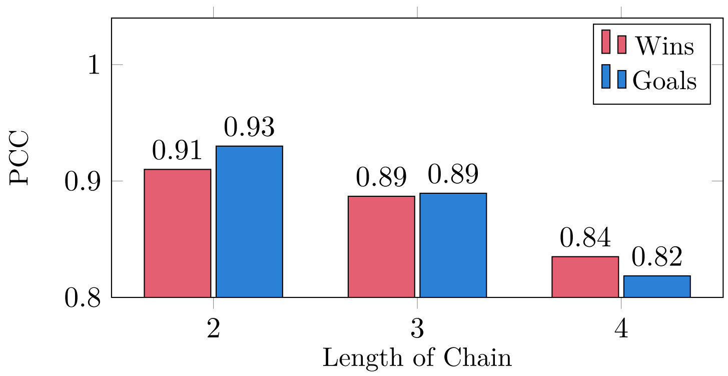


FIGURE 4.7: Effects of Chain Length on Pearson Correlation Coefficient.

As we expand the chains of interactions we see the strength of correlation between the values of the chains and the performance of the teams tail off. We see there is a 4% drop between a chain length of 2 to 3 and then a drop of 7% between a chain of 3 and 4 for the numbers of goals scored. In chains of 5 players, we see the PCC drop to 0.383 and 0.353 for wins and goals respectively. This shows that after a pair of players, the more players you include in the chain the less relevant to real-world metrics of teamwork that our computed teamwork values represent. This could be due to there being fewer examples of larger chains linking up in games for us to extract value from. This is also useful for professional teams to be able to identify pairs of players that have the largest impact on the team and then use our "Interactional Alignment" value (shown in Equation 4.6) to find the best-overlapping pairs when forming a team to play in a game.

The teamwork between the pairs of players is the best way to represent the teamwork in the squad which can then allow us to use the team formation techniques presented in this chapter. As we expand the length of the chains we see the runtime of the team formation algorithm grow due to the growing number of possible combinations of players and overlapping chains when forming the team using the model in Section 4.5.3.2. This runtime does drop again after chains of more than half the team at 6 players but as we have shown this is not useful due to the drop in PCC as we add more players. We also see that as discussed by Beal, Changder, et al. (2020), pairs of players used in the formation algorithm will provide the closest team to what is selected by a human expert team manager/coach.

4.6.6 Experiment 5: Coach Impact on Teamwork

Many different factors can affect how players work together on the pitch. These include but are not limited to: the tactics used, the languages spoken by the players, the coach/-manager and the form of the team. Therefore, in this experiment, we aim to evaluate the long term impacts on player teamwork and how this can be affected by *change events* in the wider environment. One key change event that happens over time in football teams is the change in managers/head coaches. In the English Premier League, the average time a manager is in charge is only 69 games (under 2 seasons) so we can model the impact of these changes in management and leadership in our data. This means that there are often chances to change tactics and leadership styles which in turn can affect the teamwork on the pitch.

Below in Figures 4.8 (Chelsea) and 4.9 (Arsenal), we show real-world examples of how these changes in manager affects the teamwork between players. We use an extended dataset of player teamwork back to 2017. We show a 10 match rolling average of the sum of teamwork values between the starting 11 players in each game. In these examples, the orange dotted lines represent the manager change event and these are labelled by the manager who took over.

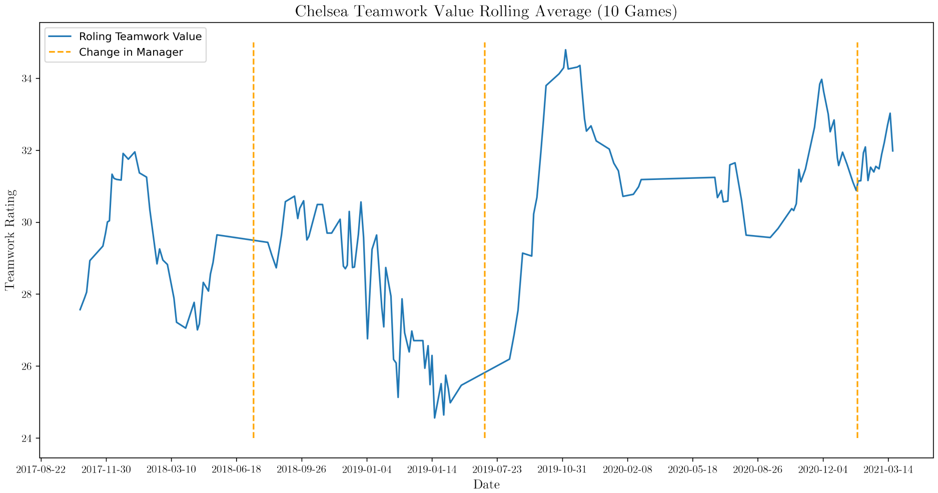


FIGURE 4.8: Chelsea Teamwork Rolling Average

Our first example shows Chelsea, we can see there were 3 manager change events (dashed orange line) over the period we focus on (2017-2021). Prior to each manager being changed we can observe a drop-off in the team chemistry that would also be reflected in a drop-off in results that lead to the laying-off of the previous manager. We can see that the first change appointment of Maurizio Sarri failed to improve the teamwork between the players in the team and saw a drop-off during his time in charge. The next change event after Sarri, was when Frank Lampard took over.[9] Here, we see a very large spike in teamwork showing that his changes in style and leadership were able to help improve the ways that players link up in the game. After the initial spike in chemistry, this eventually plateaus. After some small drops in teamwork and poor results on the pitch, we see the third change event where Lampard was replaced by Thomas Tuchel at the start of 2021. At this point, it is too soon to see what his long-term effect will be. It is worth noting that he took charge when the team was in a much better position in terms of teamwork than any manager in the past.

In our next example we show Arsenal over the same period, we see similar results. We see that the first change event represents Unai Emery taking over from Arsène Wenger, where prior to the change there was a small fall in teamwork. Once Emery took over there was a sharp rise before plateauing. Again, before our next change event which was when Mikel Arteta took over there was a dip in teamwork - we see that he was unable to bring a spike in teamwork but has fallen away further before starting to rise before the end of the 2020/21 season. This suggests he took over a team he could not bring together or the players were unable to adapt to his leadership style which meant it took the players longer to adapt.

[9] An ex-Chelsea player with a very high reputation at the club and in English football.

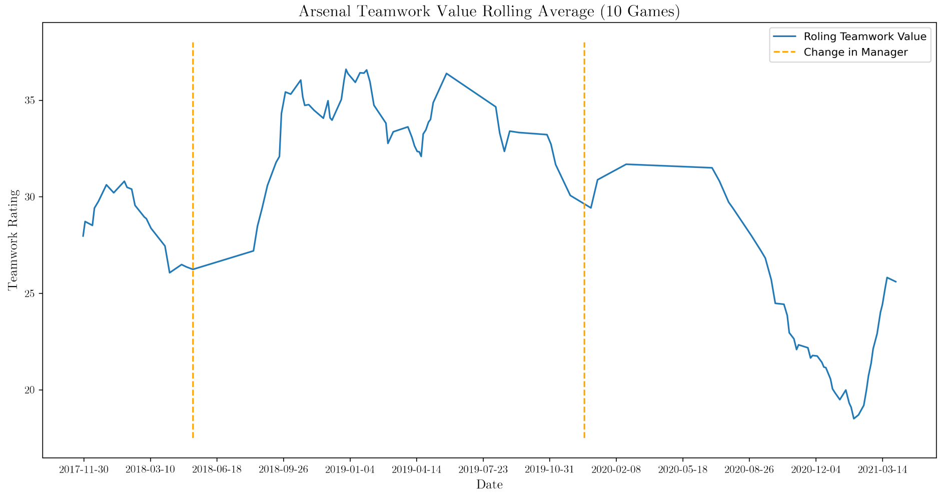


FIGURE 4.9: Arsenal Teamwork Rolling Average

These examples show how the changes in managers can have a huge impact on the teamwork between the players on the pitch. This is likely down to their impact on the leadership of the players and the changes in team player styles that may suit the players more given them more enjoyment and freedom while playing. We also see how the drop off in teamwork can often lead to the replacement of a manager and trigger the decision-makers at clubs to want to make a change to help turn around their fortunes.

4.7 Future Work and Wider Applications

Our results also suggest that this model could be applicable across many domains and, given a high-quality dataset, we could further validate the model performance to see if similar results are found (e.g., in emergency response or data transfers). The notion of teamwork would also be valuable to model and predict in many different industries and businesses with the recruitment of new staff. Being able to assess how well staff members in any domain work together would be extremely valuable. This would allow the most efficient teams to be formed in any domain. For our teamwork models to apply to new domains, we must be able to value the output of team members and their joint contributions.

As well as expanding the applications of our models into new domains, we could also expand our models in team sports. Building on this chapter, we could begin to predict how well players who do not currently play together would play given a player makes a transfer and moves team. This would allow teams to evaluate new players and transfers to ensure that they are likely to “gel” with their new teammates. We could also further evaluate the predictions of match-outcomes, based on our team valuations of a starting 11 team, against other match-outcome prediction approaches such as (Dixon and Coles

1997; Constantinou, Fenton, and Neil 2012). We would also extend the models to address how the team formation could be improved by factoring in an opposition team (in games such as football this can have a significant difference to how a team is formed).

4.8 Chapter Summary

In this chapter, we describe a novel approach to team formation based on directed interactions between agents. Our model of teamwork considers event outcomes of the chains of interactions shown as walks within graphs. We defined and tested multiple network metrics to value the contribution of agents and sets of agents and show how the value of teamwork (including *interactional alignment*) can be learnt from data and then applied to predict the performance of teams. We tested and validated our models of valuing agents and forming teams by applying our models to problems posed by football and basketball. In football, we showed that our model using teamwork pairs between players can produce similar team selections to an international level human-expert manager while also being suggesting changes to the team.

In the next chapter, we move on to look at match outcome prediction and the challenges that it presents. Specifically, we explore the use of Natural Language Processing techniques to improve on previous models by learning from the human experts.

Chapter 5

Combining Machine Learning and Human Experts to Predict Match Outcomes in Football

In this chapter, we present a novel application of a combination of Natural Language Processing and Machine Learning models which we use to predict the match outcomes of games of football. We use articles written by domain expert human journalists from the media and validate our approaches by applying the discussed models to predict the outcome of games in the English Premier League. We focus on a time period over 6 seasons, using a dataset based on newspaper match previews from *The Guardian*. When compared to the best statistical approaches and bookmakers' odds, the models presented in this chapter boost the traditional statistical methods by 6.9% in terms of match-outcome accuracy and are able to identify a greater number of rare outcomes.

5.1 Introduction

Real-world events such as sports games or elections involve competing teams, each with capabilities and tactics, aiming to win (e.g., seats during an election, or scoring more goals in a football match). The performance of such teams is typically not only dependent on the teams' abilities but also on the environment within which they operate. For example, a political party may have the best orators and policies but their opponents may be better at getting votes in key areas. Similarly, a top football team may be playing the worst team in a league but the fact that the latter may be facing relegation (to a lower league) may provide them with extra motivation to win the game. Given these circumstances, in many cases, the performance of such teams may not be easily predictable.

Traditional AI and machine learning techniques to predict the outcome of real-world events tend to focus on the use of statistical machine learning using historical data about the individual teams (N. Silver 2012; J. Y. Campbell and Shiller 1988; Dixon and Coles 1997; Matthews, Ramchurn, and Chalkiadakis 2012). However, as per the examples above, historical performance may not be useful when team performance may be dependent on dynamic factors such as human performance (morale, injuries, strategies) or environmental variables (weather, competition context, public mood). In turn, humans can be better judges than algorithms when faced with previously unseen situations. Journalists, online communities, and experienced analysts may be better at evaluating human and environmental elements to forecast an outcome. For example, companies are increasingly relying on sentiment analysis from live Twitter data and news reports to forecast stock prices or outcomes of football matches (Schumaker, Jarmoszko, and Labedz Jr 2016; Bollen, Mao, and Zeng 2011) (see Section 5.2 for more details). However, such approaches focus on opinion aggregation rather than trying to extract the potential indicators of performance for individual human teams.

In particular, in sporting events, many human factors impact how a team performs in given games. There are often situations that would be very hard to represent in numbers and statistics alone. For example, sporting rivalries often affect human emotions and team performance and teams fighting to avoid relegation from a league often obtain unexpected results.

Against this background, we propose a new approach to predict real-world sporting events involving humans based on the combination of Natural Language Processing (NLP) and statistical machine learning techniques. In more detail, we focus specifically on football games in the English Premier League (EPL) using match previews from the media alongside statistical machine learning (ML) techniques. The prediction of football match outcomes is a challenging computational problem due to the range of parameters that can influence match results. To date, probabilistic methods devised since the seminal work of Maher (Maher 1982) has generated fairly limited results and appear to have reached a glass ceiling in terms of accuracy. By improving on the current approaches for football match outcome prediction using our new methods we show that there is more to team games that involve humans than just raw statistics and incorporating human factors into a prediction model can improve accuracy.

Thus, the models in this chapter advances the previous state of the art for both NLP/ML prediction and sports match outcome prediction in the following ways:

1. New dataset for testing NLP/ML algorithms for sports match outcome prediction for football (soccer). Our dataset includes a previously unexplored feature set in terms of football match outcome predictions, including human knowledge that is

overlooked in traditional statistics. The dataset includes match data and previews for 1770 games football games over 6 seasons.

2. We propose a novel combination of Open Information Extraction (OpenIE), Sentiment analysis and supervised ML methods for predicting the outcome of games of football using human opinions from domain experts in the media.

3. We test and validate our approach by predicting the outcomes of 1770 football games over 6 seasons and compare our performance to: sentiment analysis of the articles, more traditional statistical approaches as well as bookmakers' odds (reflecting human bets).

4. We show that we can boost the accuracy of statistical approaches by 6.9% when predicting the outcome of events and that our models perform better when predicting draws and longshot results in football which are both harder to predict when using statistical methods.

The rest of this chapter is organised as follows. Section 5.2 covers background literature for NLP, OpenIE and sports outcome prediction. Section 5.3 models the problem of predicting real-world events. Section 5.4 provides the detail of how we model human opinions and 5.5 discusses the prediction methods that we use. We perform multiple experiments on our model in Section 5.6 and discuss our findings in Section 5.7. Finally, Section 5.8 concludes.

5.2 Background and Related Work

In this section, we first provide a brief description of the work in NLP and explain why these techniques are relevant. We also describe previous work in football match outcome prediction and OpenIE methods.

5.2.1 NLP for Prediction

There are many examples in past work that have used NLP techniques to make predictions on specific real-world events. Here, we define a real-world sporting event as a game involving two agents or teams of agents where the game ends in some given class (e.g., a winner, tie or loser). A sentiment analysis approach has been used in (Schumaker, Jarmoszko, and Labedz Jr 2016) to predict EPL results and turn a profit in the betting markets. They achieved an accuracy of 50% and they found that they generate more profits than a crowd-sourced odds approach when used with a betting strategy (we compare our results to this in Section 6.3). Also, there is an example of a similar analysis being performed for American Football results in the National Football League (NFL) shown in (S. Sinha et al. 2013). Over a 12 week period (177 games) in the 2012 season, they were able to correctly predict the winner 63.8% of the time.

There are further examples of work focusing on analysing and predicting the outcomes of elections in the US. The first of these (H. Wang et al. 2012) uses sentiment analysis on Twitter feeds to build a system that can gauge public opinion of the candidates in an election. Their system can analyse sentiment in the entire Twitter traffic about the election, delivering results in real-time. Furthermore, in (Tumasjan et al. 2010) the authors again use sentiment analysis and text analysis software to conduct a content analysis of over 100,000 tweets. Work in (Radinsky and Horvitz 2013) predicts a number of real-world events from new articles such as identifying increases in the likelihood of disease outbreaks, deaths, and riots. Following on from this, in (Chakraborty et al. 2016) socio-economic indicators (such as food prices) are predicted using news events, they achieve good results reducing the root mean square error of prediction by 22% in comparison to standard models. Also, work in (Xie et al. 2013) uses financial news articles and semantic frames to predict stock price movements and other recent work such as (Verma et al. 2011) has shown the success of NLP when extracting situational awareness to aid emergency response teams.

The real-world sports event prediction model we outline in this chapter differs from current work as we do not need to use sentiment analysis or social network data to make our predictions. We use text extracts from human experts and journalists to find patterns that relate to the real-world event outcomes which are more complex than standard tweets from the general public.

5.2.2 Open Information Extraction for Sports Prediction

Open information extraction (OpenIE) approaches extract propositional tuples from free text sentences, typically focused on verb or noun mediated phrases. Key characteristics of OpenIE (Banko et al. 2007; Cui, Wei, and M. Zhou 2018; Z. Jiang, Yin, and Neubig 2019) are domain independence, unsupervised extraction, and scalability to large amounts of text. For example the sentence “Cristiano Ronaldo was born in Portugal” might generate a verb-mediated propositional tuple (Cristiano Ronaldo;was born in;Portugal), and the sentence “Manchester United ex manager, Josè Mourinho” a noun-mediated propositional tuple (Manchester United, ex manager, Josè Mourinho). Being unsupervised it avoids the need to compile a large training corpus for each domain.

OpenIE has been used for knowledge-based population (KBP) which includes sub-tasks of slot filling and entity linking. Slot filling is where all known information is added to target entities, and entity linking is where references to entities are disambiguated.

Entity types are usually focused on people, locations and organisations, with knowledge-base properties associated as attributes of entities or relations between entities.

Downstream applications of OpenIE (Mausam 2016) focus on web and news applications such as identifying co-occurring news articles (Balasubramanian, Soderland, and Etzioni 2012). The focus of our work is OpenIE applied to sports prediction, which is a new downstream application type. Our context allocation method can be seen as a type of entity linking approach, where entity mentions in sports previews are semantically grounded to specific entities in a sports result statistics dataset.

5.2.3 Sports Outcome Prediction

In Section 2.1 we give a full overview of work in sports outcome prediction. In this subsection, we highlight some of the more relevant works to the models in this chapter. The model discussed by Dixon and Coles (Dixon and Coles 1997) is based on the different abilities of both teams and the Poisson distribution is used to model the number of goals scored by each team and therefore predict the outcome. Their model was found to have a prediction accuracy of 56.65% over the past 5 seasons. This model is still one of the most accurate and widely used.[1] Following on from this, (Constantinou, Fenton, and Neil 2012) apply Bayesian Models to football match outcomes and the website *www.fivethirtyeight.com* predict football games using a power index metric that they calculate for each team using a Poisson process.[2]

Many other statistical and machine learning based models have been tested for match outcome predictions in many sports such as (Miljković et al. 2010; Haghighat, Rastegari, and Nourafza 2013; Leung and K. W. Joseph 2014). However, these approaches have been used for many years without any significant improvement. We believe (as highlighted in Section 2.5) that part of the reason for the lack of improvement to models is that only the traditional statistics are used. In sport, many uncertainties affect the outcome of a game that can be very hard to express through statistics and numbers. These include: the importance of a game, team rivalries, managerial changes, new signings and rotations to the team. A journalist can interpret this information in a better way and make assumptions that machines are unable to. By analysing a series of pre-match previews from the media we extrapolate features from human opinions that may affect the outcome of a game. In our model, we apply NLP and sentiment analysis techniques alongside machine learning to create a human-expert and machine model to make predictions that have not been applied before in the sports domain.

[1] Discussed in a 2013 book "Analysis of Football Prediction Methods" by William Brojanigo.
[2] https://fivethirtyeight.com/methodology/how-our-club-soccer-predictions-work.

5.3 The Model

The model outlined in this chapter provides an approach to interpret a complex text from newspapers and online articles to predict a given match outcome in a game of football. In this section, we describe the steps of the prediction model in a generalisable way so that these techniques could be reapplied in domains (e.g., elections and other sports events). We use a combination of NLP and ML techniques, which in the past have been used for knowledge-based population (Ji and Grishman 2011) and document classification (Manevitz and Yousef 2001).

5.3.1 Model Definitions

We define the set of upcoming events as E with a set of i possible outcomes, $\mathcal{O} = \{o_1, o_2, \ldots, o_i\}$ where $i > 1$. Each $e \in E$ has a set of texts written about it $T = \{t_1, t_2, \ldots, t_j\}$ where $j \geq 1$. For example, in an election, the possible outcomes would be the different political parties that could win and in a sporting event, the outcome would be the two teams that could win (or draw). The text written about the events in these examples are preview articles in the press discussing information regarding the event.[3]

Each text $t \in T$ is built up from multiple sentences denoted by $t = \{s_1, s_2, \ldots, s_k\}$ where k is the number of sentences in the text article. We transform each sentence s into a form that we can use as a feature to give $f(s)$, where f represents the transformation function to output a numeric vector representation of the sentence. Using a sentiment analysis approach, the function $f(s)$ would represent the sentiment value of the sentence, where $-1 \leq f(s) \leq 1$.

Each sentence may be related to zero or more of the outcomes in $\mathcal{O}$ (e.g., a sentence may be discussing a political party or the home team in a football game). We calculate the probability $p(f(s)|o_i)$ that each sentence relates to a team competing in the game and then allocate that sentence to the most likely team. If there is uncertainty in the allocation then the allocation is set to no team (equal probability for both teams in Equation 5.1). An allocation is defined as a pair of the sentence vector and its allocated outcome (home/away team win), $a = (f(s), o_i)$ where $a \in \mathcal{A}$ and $\mathcal{A}$ is the set of allocations $|\mathcal{A}| = k$.

With the allocations, we calculate the final features (X) for the outcome y where $y \in \mathcal{O}$ so that we can use these features to train a machine learning model. When using a text vectorization approach, the features are created by the addition of all the sentence vectors that are allocated to a given event outcome (e.g., this would be an addition of

[3] For football examples see: https://www.theguardian.com/football/series/match-previews and for election example see: https://www.bbc.co.uk/news/uk-politics-49826655.

the vectors that relate to a given political party) this gives the final vector for each outcome $v(o_i)$. If we are using a sentiment approach the average sentiment from all allocated sentences will be used. We now define our feature set X for each event E as $X = [v(o_1), v(o_2), \ldots, v(o_i)]$ and the target y is the actual outcome $o \in \mathcal{O}$ of E which we aim to predict using a function ϕ so that $\phi(X) = y$.

5.3.2 Model Process

To solve our model we start with some set of texts T from a given source and we are aiming to produce a prediction of the outcome from a real-world event, y. Specifically, we apply the steps which are outlined in Figure 5.1. Here, we discuss each stage and the methods that we use. We will discuss these in further detail in Section 5.4 and 5.5.

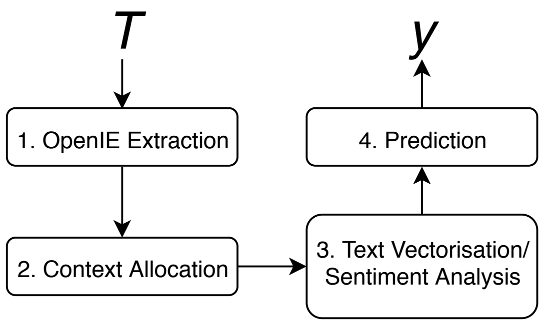


FIGURE 5.1: NLP Real-World Prediction Model Process Diagram

1. **OpenIE Extraction:** Relation tuples are extracted in the form {argument, relation, argument} for each sentence in the articles text, for example {Machester United, ex-manager, Mourinho}.

2. **Allocation of Text Context:** We allocate each sentence to an outcome ($a = (f(s), o)$). For example, in the case of football, each sentence must be allocated to one of the teams that are playing in the match that the article is discussing. This is expanded on in Section 5.4.2.

3. **Text Vectorisation/Sentiment Analysis:** We convert the sentences into vectors using a Count Vectorizer technique so we have a numerical representation of the words in a sentence. We can use sentiment analysis (instead of vectorisation) to give the sentiment for each of the sentences. These give the $f(s)$ value and the final features are computed to form X. These are discussed in more detail on Sections 5.4.3 and 5.4.4

4. **Prediction:** Once we have formed our feature set (X) for each event that we are aiming to predict, we train a machine learning model (ϕ) using historic data (with outcomes) and the features X. This is used to make the final predictions of events based on the original articles. The outcome that we aim to predict (y) corresponds to one of the possible outcomes in $\mathcal{O}$ (e.g., for football predictions $y = home, draw, away$).

5.4 Modelling Human Opinion

In this section, we discuss in detail the NLP methods that we have used to formulate our features and the methods used to calculate the sentiment. We then use the outputs from these with machine learning algorithms in Section 5.5.

5.4.1 OpenIE Extraction

To create relation tuples from each match preview report we convert the document corpus to a set of sentences and apply an existing OpenIE algorithm. The result of OpenIE is a set of verb-mediated relational proposition annotations for each sentence in the form argument, relation, argument. Each argument or relation phrase is a n-ary phrase and is not semantically grounded. We use the next step, context allocation, to disambiguate mentions of sports-related entities and behaviours to entities within the sports dataset and achieve entity linking. We use OpenIE as we aim to capture predicates as well as noun-phrase named entities. Predicates (e.g. verb action words) provide markers for features around what players, managers and teams are doing of feeling.

5.4.2 Context Allocation

Using the extracted text sentences from the OpenIE extraction, we allocate each sentence to a team involved in that match that we are aiming to predict. This gives the sentence context to what we are aiming to predict in a similar way to which word embedding is used in an information retrieval context (D. Ganguly et al. 2015). This is an important part of our model process as it means that we get the correct features allocated to the corresponding teams. Our approach to the allocation problem uses key term dictionaries for each team so that for each sentence we can calculate the probability of it being related to one of the individual teams involved in the game. The probability that a sentence belongs to a team is calculated using Equation 5.1.

$$p(s|t) = \frac{\sum_{n=1}^{N} 1_{\{s_n \in \mathcal{D}_t\}}}{N} \tag{5.1}$$

Where $p(s|t)$ is the probability that sentence s belongs to team t and the sentence is allocated to the team with the highest probability. Sentence s represents a list of words where N is the number of words and $\mathcal{D}_t$ is the dictionary of words belonging to team t. The dictionary that we use is human-generated and is built up from key terms from each team (e.g., team manager, stadium, nicknames, list of players). A shortened example dictionary for Southampton in the 2018/19 season would be:

Southampton = {*Hassenhutl, Saints, St.Marys, Bertrand, Ward-Prowse, Ings, Redmond, Hojbjerg, Romeu, Yoshida*}

5.4.3 Text Vectorisation

To create a vector from our text that we can use as features in the prediction models we use a CountVectorizer approach.[4] This allows us to tokenize our collection of articles and build a vocabulary of known words. It also allows us to encode new articles when predicting unseen data. This returns an encoded vector for each sentence with a length of the entire vocabulary and an integer count for the number of times each word appeared in the sentence. The returned vectors may contain a lot of zeros, therefore, we perform further analysis on the vectors to extract the key features that have the highest impact on the outcome of the event that we are aiming to predict.

5.4.3.1 Vector Output

Using each of the vectors created from the sentences (which have been allocated to a team) we produce our final vector output which can be used with machine learning models for prediction in Section 5.5. To formulate our final vector we combine the sentences that are allocated to a team t for an individual game by summing the respective elements into a single vector for each team, for each game - this is shown in Equation 5.2.

$$v(t) = \sum_{k=1}^{K} v(s_{k,t}) \tag{5.2}$$

Where, $v(s_{k,t})$ is the vector of sentence s_k that is allocated to a team t and K is the number of sentences. We produce our final vector for game g by using both the home

[4]A number of techniques were tested but it is found that CountVectorizer gives the best results.

team vector $v(h)$ with a weighting α to represent the home team advantage (S. Clarke and Norman 1995) and away team vector $v(a)$ together to give the final vector $v(g) = [\alpha \cdot v(h), v(a)]$. Each event/game will have the corresponding vector which will be used to make the prediction. For each game g we use the vector $v(g)$ to predict the outcome of the game. This gives us $\phi(v(g)) = y$ where $y \in \{home, draw, away\}$ and ϕ is some machine learning algorithm applied to predict the outcome.

5.4.4 Sentiment Analysis

We also perform sentiment analysis for each team in each game which we will compare against the text vector approach (shown in Section 5.6.4). A sentiment approach gives us a value for the probability that each sentence in an article belongs to the positive, neutral and negative topics. Then as we allocate each sentence to a team playing in the game we can calculate if the sentiment for that team in the article is positive, neutral or negative. One approach that could be used for this is training a machine learning model to classify sentences into the sentiment. However, we do not have a labelled dataset to be able to do this, therefore it would need to be manually labelled. Due to this, to calculate the sentiment of our sentences we use approaches that are discussed in (Lin and He 2009) where a Latent Dirichlet Allocation (LDA) approach is used to add sentiment to movie reviews without labels. In what follows, the methods are outlined in more detail.

5.4.4.1 Prior Information

We use a paradigm word list that consists of a set of positive and negative words (e.g., fantastic and terrible). These words can be treated to define the positive and negative semantics of a sentence. The majority of the words were derived from the word lists used in (B. Pang, L. Lee, and Vaithyanathan 2002).

5.4.4.2 Latent Dirichlet Allocation (LDA)

We use an LDA approach (Blei, Ng, and M. Jordan 2003) to calculate the probability that each sentence belongs to either a positive, negative or neutral “topic”. Each sentence is viewed as a mixture of various “topics” where each sentence is considered to have a set of topics that are assigned to it via LDA. By using the techniques discussed by Blei, Ng and Jordan (2003), we calculate the probability that each sentence belongs to the positive, negative and neutral classes. In our model, the corpus D is the set of all sentences in the media articles, the document w is the sentences from the text, we iterate go through all words w_n and calculate a multinomial probability $p(w_n|z_n)$ conditioned on the topic z_n. In this case, our topics are $z = \{positive, neutral, negative\}$.

We calculate an average probability (from all the allocated sentences) for each of the possible outcomes from the event that we are aiming to predict. For the football match outcome problem we create a feature set formed using an average of the positive, neutral and negative probabilities for each team and use these as features in a machine learning model to predict the final output.

5.5 Match Outcome Prediction

In this section, we discuss the methods that we use to make predictions using the text vector features and sentiment analysis that are discussed in Section 5.4.

5.5.1 Feature Importance

Due to the number of features that are created when transforming text to vectors, we perform analysis to select the most important features and pick out the words which have the highest impact on the outcome of the event. This will help to reduce noise in our model. To do this we apply a multi-class Logistic Regression approach (Hosmer Jr, Lemeshow, and Sturdivant 2013) to calculate the weight of each of the features in the model.

This could also be done by using PCA methods (Malhi and Gao 2004) or several other feature selection approaches. We show the important feature words that we identify and discuss the findings in Section 5.6.3.

5.5.2 Machine Learning Methods

Using the selected features discussed in the previous sub-section, we tested nine machine learning methods and found that the Random Forest was the optimal method to use for our dataset.[5]. In football matches there are three possible outcomes that we must consider in our models, these are: a home win, a draw and an away win. These outcomes are determined by the number of goals scored by each team.

The feature set X that is used for our model is the set of the selected features discussed in the previous section (either using text vectorisation or sentiment analysis) and the

[5]It is worth noting that this may not be the case for all applications of this model. Classification methods tested: nearest neighbours, linear SVM, RBF SVM, decision tree, random forest, neural network, naive Bayes, QDA, logistic regression

target set y is the corresponding outcome from the game that the feature text refers to. This means that we use multi-class Random Forest classification method as we have 3 possible classes as the targeted outputs.

5.5.2.1 Random Forest

A random forest model (Breiman 2001) is formed with a collection of different tree predictors where X is the feature set, $h(X, \Theta)$ is the individual tree's output and Θ is a random vector generated, independent of the past random vectors but with the same distribution.

$$p(y|X) = \frac{\sum_{k=1}^{K} h(X, \Theta_k)}{F} \tag{5.3}$$

The outcome prediction is given by taking an average of the collection of tree predictor outputs which gives the probability that the features X belong to the outcome y and F is the number of trees in the forest. The model hyper-parameters are fine-tuned using a GridSearch method.

5.6 Experiments and Evaluation

This section outlines the experiments which we perform to test the NLP models discussed in this chapter.[6] Experiments are run using historic data taken from the Guardian match previews as well as other statistics from the English Premier League (EPL) and historical odds were taken from OddsPortal.[7] All experiments are run using match previews written before the game took place and with pre-match bookmaker odds to ensure that each test is fair and there is no contamination of data in our experiments.

5.6.1 Text Dataset from the Media

As well as the model outlined in this chapter, we provide a novel dataset for researchers to use and implement their own models for the challenge of predicting sports events using human experts from the media. This dataset can be found at *https://github.com/RyanBeal7/GuardianPreviewData* with all data sourced from the Guardians English Premier League match previews over the past 6 seasons.[8] The dataset does not contain

[6] All experiments run using the SciKit Learn Toolkit.

[7] Pre-match historic odds: https://www.oddsportal.com/results/soccer (taken at kick-off).

[8] Guardian articles are open source and can be found here: https://www.theguardian.com/football/series/match-previews.

an exhaustive list of all games during the seasons that we focus on, although we are able to source data from 1770 games across 6 seasons from 2013/14 to 2018/19. The models in this chapter shows the first analysis of a dataset of this type, combining text and statistics to predict football matches. In this domain, 1770 games is a large dataset for predictions of football games, other examples of papers for this problem usually only tests on 1 or 2 seasons of data (380 games per EPL season). An example snippet from a match preview regarding a game between Southampton and Tottenham in 2019 is as follows:

"Which Tottenham team will show up at St Mary's? While others were blowing first-leg leads, Tottenham competed a clinical Champions League victory over Dortmund – but they are winless in their last three league games. Mauricio Pochettino begins his touchline ban and will be without Kieran Trippier, although Dele Alli and Harry Winks could feature."

This type of dataset can be used to analyse how the wording and sentiment regarding teams in similar reports (over the 1770 games) correlate to the match outcome. We expect that human-related factors (e.g., team rivalries and behind the scene changes in staff/players) will be brought through in our predictions and improve on the traditional statistical approaches. The data for this is outlined in the next subsection.

5.6.2 Statistical Data

As will be discussed in Experiment 2, we also use statistical-based methods to compare against our prediction models that use only the text-based features. We sourced data for basic statistics for the 1770 games that our experiments focus on from FB-Ref.com which provides a number of useful statistics in football, the link for the EPL data can be found at *https://fbref.com/en/comps/ 9/Premier-League-Stats*. We use features outlined in (Dixon and Coles 1997) which includes the attacking and defensive efficiencies of the teams in the game which are calculated using goals scored and conceded before a given game.

5.6.3 Experiment 1: Important Feature Words

Using the feature importance methods which we discussed in Section 5.3, we evaluate and pick out the words that have been shown to have the highest impact on the outcome of the games. This allows us to evaluate what our model is learning in terms of the words that the human-expert journalists write. The more weight added to a word the more impact that this would have within the model and is more likely to correlate to a match outcome (win, draw, loss). The top 10 features and their impact values are shown in Figure 5.2.

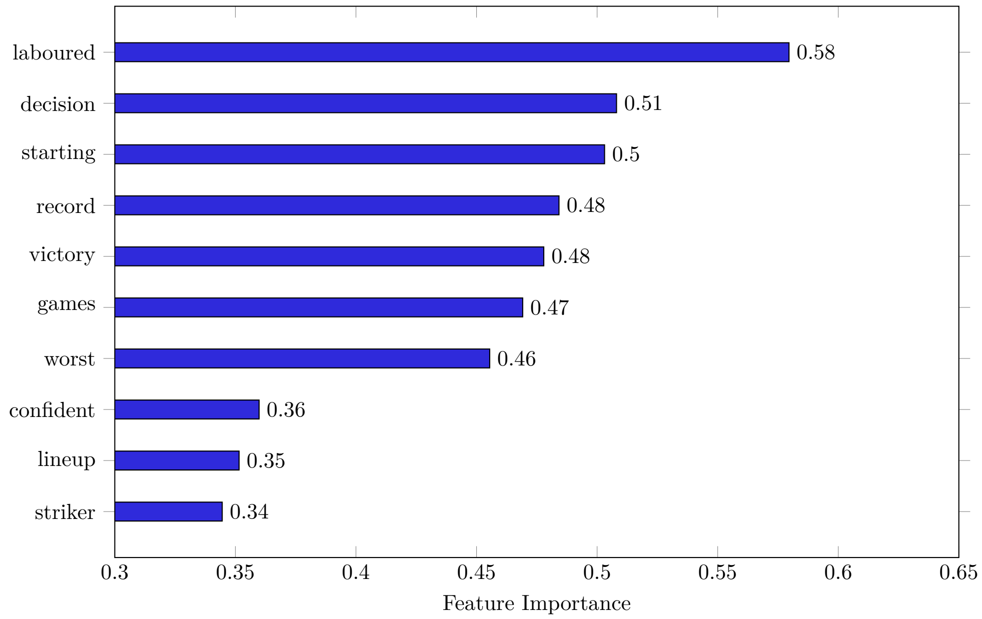


FIGURE 5.2: Top 10 Words with Highest Ranked Feature Importance.

The words "decision", "starting", "lineup" and "striker" suggest that discussions in the article text regarding teams tactical decisions and starting lineup have an impact on the prediction of the match outcome. The other words that are shown here seem to be related to team form and performances both in a positive way (e.g. "victory") and a negative way (e.g. "laboured"). The words regarding possible tactical decisions are particularly interesting as it shows that the model can pick out some of the words relating to human decisions that may be missed by traditional statistics.

5.6.4 Experiment 2: Accuracy of NLP Outcome Prediction

Using the methods that we have discussed throughout this chapter we test and compare the accuracy, precision and recall against alternative methods which are outlined below. We decided to compare our results to that of a well-known football prediction model by Dixon and Coles (Dixon and Coles 1997). This model is still seen as one of the leading examples of statistical modelling for football and has not been greatly improved through the use of newer machine learning techniques (this is discussed in Section 2.1). Therefore, we believe for our models to be successful we aim to improve on the Dixon and Coles model, as well as the bookmakers' accuracy.

The results from this test are shown in Figure 5.3 and Table 5.1. The test was run taking an average of the performance across 3 seasons (2016-2019), using a training set[9] of all games prior to that season and a test set of 300 games in the season.[10]

[9]Size of training set: 16/17 = 516, 17/18 = 787, 18/19 = 1033.

[10]It is worth noting that the class distribution of EPL games from across 25 seasons of EPL football is 46.2% home wins, draws 27.52% and away wins 26.32%.

- **Model 1 (Text Vectors):** Uses the process shown in Figure 5.1 with features formed using text vectorisation methods.
- **Model 2 (Sentiment Analysis):** Uses the process shown in Figure 5.1 with features formed using sentiment analysis methods.
- **Model 3 (Dixon and Coles):** Represents the outputs from the model described by Dixon and Coles (1997).
- **Model 4 (Bookmakers):** Uses the pre-match bookmakers' favourite as the predicted winner of the game.
- **Model 5 (Text Vector Combination):** Uses features calculated from text vectorisation (Model 1), outputs from Dixon and Coles (Model 3) and the pre-match bookmakers' odds (Model 4). The features are the probability of a home win, away win and draw from each of the three models. Therefore, we have 9 features per game and apply a Random Forest classifier.
- **Model 6 (Sentiment Combination):** Here we combine Model 2, 3 and 4 in a similar fashion as above but by using sentiment features rather than text vectorisation. We then use a Random Forest classifier with a feature set made up of from 9 probabilities.

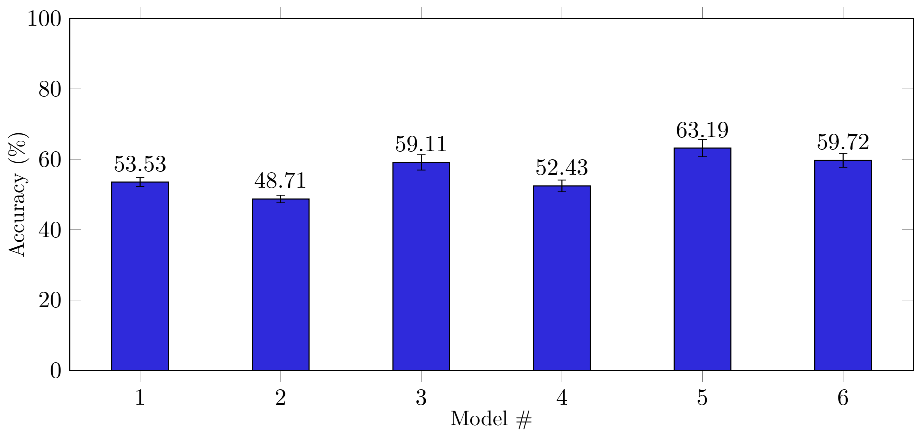


FIGURE 5.3: Comparison of Model Accuracies.

Model #	Precision	Recall	F1 Score
1	**0.649**	0.413	0.388
2	0.163	0.333	0.218
3	0.503	0.491	0.456
4	0.451	0.452	0.445
5	0.612	**0.563**	**0.569**
6	0.515	0.497	0.464

TABLE 5.1: Precision/Recall Results.

These results show that using NLP methods, on their own, do not produce remarkable results. Model 1 (using just text vector methods) produced an accuracy of 53.5% and model 2 (using sentiment analysis) produced an accuracy of 48.7%. Neither of these can better the Dixon and Coles predictions (6.9% boost). However, we found that when we use the prediction probabilities output from Model 1 with the bookmakers' probabilities and Dixon and Coles probabilities we can improve these methods. Model 5 (using text vector, bookmakers and Dixon and Coles probabilities) achieves an accuracy of 63.2% which is a 10.8% increase on the bookmakers' accuracy and 4.1% more than Dixon and Coles (6.9% boost). We show that Model 5 has the highest F1 score of all the models, which is 0.113 more than Dixon and Coles and 0.124 more than the bookmakers. This shows that the text features boost traditional statistical methods to produce a higher accuracy and F1 score.

We also test how Model 5 performs without the use of the text vector features. This test is to show that it is the new features that cause the boost in accuracy. We find that without the text features, the F1 score is 10% less and the accuracy 7% less than the Model 5 results and therefore can claim that the boost is due to the text vector features.

5.6.5 Experiment 3: Longshots and Draws

The traditional statistical models and bookmakers approach to predicting football match outcomes are typically poor at predicting draws and longshot results. A longshot result is when the winning team has a bookmaker probability of less than 20%. Therefore, we test the ability of our new approach to predict these events by using the same models defined and trained in Experiment 2. To do this we split all 1770 games into training and test sets (random 80/20 split). In the test set, there are 75 draws and 47 longshot outcomes. The results are shown in Table 5.2.

Model #	Draw (%)	Longshot (%)
1	0 (±0)	**38.9** (±1.91)
2	0 (±0)	25.9 (±1.34)
3	0 (±0)	0 (±0)
4	0 (±0)	0 (±0)
5	**26.5** (±1.26)	22.2 (±1.09)
6	1.33 (±0.07)	3.70 (±0.19)

Table 5.2: Longshot and Draw Comparison.

These results show that by using our new models we can more accurately identify when draws and longshot results are likely to happen. This may be because there are some more subjective factors and knowledge that affects games that cannot be considered in more statistical approaches. Some examples of this may be when the text articles discuss the possible line-up of the team and if a manager may rotate. Another may be that if a team signs a new manager or player, articles discuss the possible impacts of this, which would be hard to quantify by just using stats. As we will see in later experiments, these human factors play a bigger part later in the season. For example, a team that has been poor all year but is fighting relegation (and therefore is a longshot) may have a better chance than the stats suggest when games start to make a big difference. This is especially the case in the EPL where relegation can cost a team over £50,000,000 in lost revenues.[11]

5.6.6 Experiment 4: Season Performance

In this experiment, we assess how the top model from the past experiments (Model 5) performs over an entire EPL season in comparison to Dixon and Coles and the bookmakers. To do this, we train our model using all data from the seasons before the 18/19 season (all articles and statistics) and then run our model for each game-week to show how many matches would be predicted correctly across the season. This is shown below in Figure 5.4 where we show the accumulation of correct results across the season.[12]

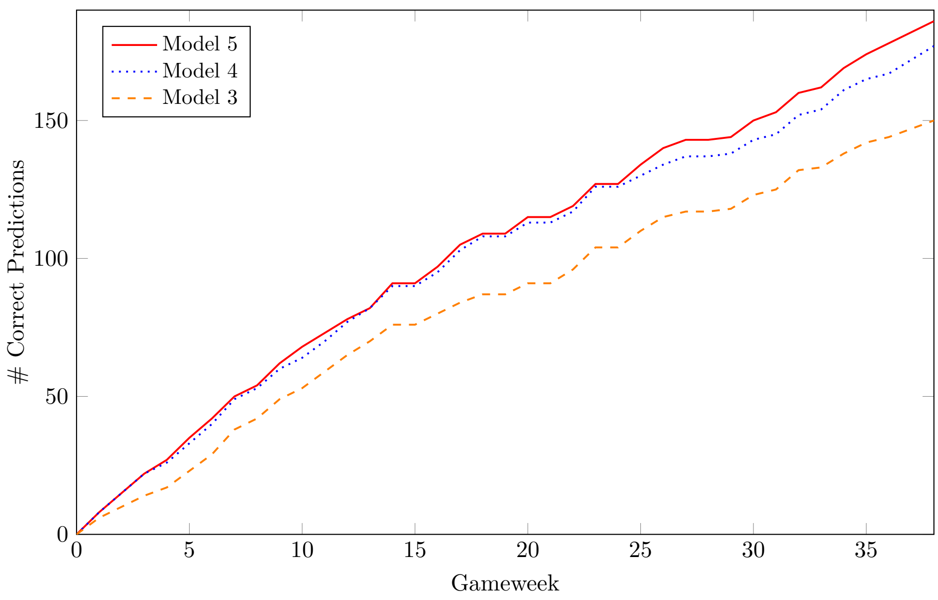


FIGURE 5.4: 2018/19 EPL Week by Week Analysis

This shows that Model 5 can continually perform well and predict correct results across a given EPL season when set up in a real-world scenario. We also find that there is a 2.23% increase in accuracy between week 1 and week 38 for model 5, showing that as the season progresses our model improves. It also shows how the articles can better model

[11] https://bleacherreport.com/articles/1591059-the-cost-of-relegation-5-reasons-to-stay-in-the-eplslide0

[12] In this test there are a total of 280 games which we could obtain data for out of the total 380 possible games.

the different scenarios that teams may be in later on in the season which numbers do not represent as well. For example, if a team has played poorly all season but is now in a relegation battle they may have more to play for than a mid-table team with no chances of winning the league or relegation. The same goes when teams are fighting to win the league or qualify for European competitions. Injuries and rotation also play a big part later in the season as teams who have progressed into the later rounds of the FA Cup and European competitions have many more games.[13] This shows the key contribution of the media preview analysis and how by taking into the human factors that are written about by the domain-expert journalists we are able to better predict the outcome of football matches.

5.7 Future Work and Wider Applications

In this chapter, we have focused on using football outcome prediction to validate the models that we discuss as there is a high volume of data available and there are several other approaches in the literature that we are able to compare our results to. This allows us to see the improvements that are made by our new models. Our models could also be applied across a number of sports and it would be interesting to see how these methods improve the predictions in different sports. This could be used as a comparison to which sports have the most human factors that impact the match outcomes and allow us to see which sports are the most predictable using statistics alone.

We also believe that our models can be generalised further and if they were to be applied to other domains we would see similar results. For example, our model could be used to predict which political party would win an election, as similarly there are many articles available which preview the election and historic statistics that can be used to make predictions. This could have been used recently to help predict the outcome of the Brexit referendum. Another possible application for our model could be the prediction of which new policies and laws will be passed by Parliament/Congress as there would be prior data and media discussion.

We could also build on the ensemble learning that we have shown would work in this chapter (as the text-based methods boost the performance of statistical-based methods). Further testing could be performed looking at a number of existing football match outcome prediction approaches (many of these are discussed in 2.5) using statistical approaches and find which methods see the largest boost in performance when using our text-based methods. We could also devise and test our own feature set with the statistic and text features combined.

[13] Discussed in injury reports by JLT and in Section 2.4 of this thesis.

Finally, we would like to explore a wider variety of data sources for the preview articles. By doing this another application for the models in this chapter would be to assess the reliability of different newspapers, news outlets and journalists. This would help to analyse who is writing the best material which best correlates to the outcome of real-world events. This could again be explored using datasets in football as there is clearly defined event outcomes and many different data sources that discuss the same event.

5.8 Chapter Summary

In conclusion, this chapter has presented a new model for interpreting articles from the media in order to make predictions of match outcomes in football. We believe these models are generalizable and could be used across a number of domains in sport and beyond. We have explored how our models can improve on the leading traditional statistical approaches for football match outcome prediction. We show that we boost these methods by 6.9% in terms of outcome accuracy and we also show that our new text-based models identify rarer events such as draws and longshot results. We find that the model accuracy increases as the season progresses and human factors/emotions begin to play a bigger part in the game. Overall, our results suggest that our models have been successful and could be applied to the prediction of other real-world events outside of sports prediction.

In the next chapter, we discuss the results from the previous three chapters and discuss the future uses for AI in team sports and how this can change the industry.

Chapter 6

Discussion

In this chapter, we summarise the key contributions and we discuss the impact the research is expected to have in the AI and sports analytics communities. We will also discuss avenues for future work.

6.1 Game Theory for Football Tactics

In Chapter 3 we presented models for optimising both the short-term (single game) tactics for football teams as well as longer-term tactical optimisation across an entire season to help teams reach their long-term objectives. One key observation from our testing of our Bayesian/stochastic models for individual matches is how attitude differs between home and away teams. In terms of "closeness" (fully defined in Section 3.6), which identifies the model which produces the most similar actions to those selected in the real-world. We find that for away team tactics, those are output by the spiteful approach are "close" in 69% of cases in comparison to 33% and 32% for the best response and expectimax respectively. This shows that away teams are more likely to select tactics that minimise the chances of the opposition winning rather than trying to maximise their chances of winning. By comparison, when we assess the "closeness" of our model outputs for home teams, we find results of 38% for best response, 50% for spiteful and 53% expectimax, showing an increase in the number of teams that aim to win a game. This is expected because there is an advantage of playing at home.

Turning to our model for optimising the long-term performance of teams, in our experiments, we focused on the outcome where our models are used only by an individual team. However, when we run our experiments with all teams optimising their tactics using our model, we find that the models are not effective and the final standings are very similar to what we see when we simulate without the new fluent objective and prior game weights. This is to be expected due to the optimisations being cancelled out by opposing teams using the same approach. We see that there is a boost of under 1 position on average per team when every team uses the model in the same season. This shows that teams can boost their performance over the season but only if they utilise the game-theoretic approaches while others do not.

When taken together our models can provide several benefits to football teams and organisations. It is unlikely that any football manager/coach will blindly follow the recommendations from our model, but they do allow coaches to gain assurance/extra information against their own assumptions and expert knowledge of the game. Our pre-game Bayesian model can also help game analysts employed by leading clubs to learn more about their oppositions using AI to identify where an opponent may be weak and what tactics to use to exploit this. Again, the in-match model may help coaches identify the right times to make changes in a game to boost their chances of winning or drawing, particularly at times of high stress/pressure during a game.

Across longer periods, by allowing teams to simulate seasons and extrapolate the distribution of the likely outcomes of decisions, they can gain more information about how the decision will affect performance and financial targets. In turn, this can improve their decision-making and identify if wholesale changes need to be made. For example, if a team is under-performing and it seems likely that objectives will be missed then the directors at the club may want to make changes to coaches or staff. On the other hand, if a team is over-performing and are likely to exceed their goals then the directors may want to invest more money into player transfers to ensure that the team keeps performing well as they know that they have assurances against future income. These types of decisions could be added into the model to help decision-makers subjectively decide when to invest or make changes.

6.2 Sports Teamwork

In Chapter 4, we propose a model to extract and optimise the value of teamwork between players within team sports. Although football and basketball may seem to be fairly similar (two teams of players competing against each other to score goals/points), we find that there are key differences in the way that teamwork affects on-field performance. In football, there is a stronger correlation between teamwork value to both goals scored and games won across a season. This may be due to the time that players play together in football where (a 90-minute game) only 3 substitutions can be made, whereas in basketball unlimited changes can be made. There may also be a higher correlation to success metrics in football due to the lower scoring nature of the game, meaning that we are able to more effectively learn the impact of the possession outcome events in football.

We also explored the optimal number of agents to use when valuing teamwork and forming efficient teams. We found that the model was most accurate (in terms of correlating to real-world performance metrics of teams) when using a pair of players in our chains and assessing their *interactional alignment* when forming the teams. As we add more agents into the chains, we find that the correlation to the success metrics drops. This suggests that the way agents work together as pairs is more important than identifying longer chains of 2+ players within the wider overall team that work well together.

This also helps to validate our use of *interactional alignment* where we form teams using overlapping pairs rather than expanding the length of chains and find this correlates most to how human managers make decisions. We tested our team formation methods by comparing the outputs to that of a human expert team manager. Our results show that our model can form teams that are similar to the selections of human experts and that we can suggest a small number of changes that could improve the team. This comparison also suggests that human experts consider the teamwork between players during team selection.

As well as assessing how human managers select their teams, we also tested how they can affect teamwork. We saw that often a change in the manager who changes leadership/playing styles can help improve the teamwork and on-pitch performance. This type of analysis could help team decision-makers (e.g., team owner or chairman) decide when they need to make changes.

6.3 NLP to Predict Football Games

The results of our experiments presented in Chapter 5 highlight many interesting points. Firstly, we analysed the words that have the highest correlation to the outcome of the event. As we discussed in Experiment 1 in Section 5.6, we found that words that related to human decision making (e.g. a manager picking players) had a higher weighting. This may be because these are the types of variables that are missed by statistics and it can be hard to use numbers to predict how a human will make decisions. On the other hand, a human expert writing a match preview can infer more about how managers and players may act and would know more about the individual's personalities and moods within the team. Therefore, this gives us a method to encode human expert knowledge from their writing. We also found that by using count vectorization approaches from the main articles also significantly outperformed those using the sentiment analysis on the same text corpus.

By using the media previews we identify more longshot results and more draws which are traditionally hard to predict in statistical models due to them being anomalies in the dataset. This may be because these types of results rely on small changes which again traditional statistical methods overlook. For example, if a manager is likely to make a lot of changes due to upcoming important fixtures then it may make it more likely for the underdog team to get a draw or beat the better team. As shown in the experiment, the current statistical methods rarely predict a draw even though they occur in 27.52% of games.

The results we show in Figure 5.4 highlight how our models perform better when compared to traditional statistical models later on in a football season. We believe that is due to the greater number of human factors that play an increasing role in the latter part of an EPL football season. For context, in the EPL the top 4 teams qualify for the UEFA Champions League and the next two teams qualify for the UEFA Europa League; competitions that are worth a lot of money and prestige. At the other end, the bottom 3 teams are relegated to the Championship which is a large drop in income. Therefore, the closer we get to the end of the season the closer these outcomes become and teams start to fight harder to get into European competitions and avoid relegation. In particular, a team that has played poorly all season may be more likely to beat a team with nothing to play for because they are motivated to avoid relegation.

Overall, our experiments found that we achieve the most accurate results when we do not just use the predictions from text or just from the statistics, but when we combine the two as features in an ensemble ML model. We believe this to be the case as the model then incorporates all the important stats (e.g., form, league position, goals scored etc.) as well as the new features from the media previews that incorporate more of the more subjective features which the statistics are unable to consider. A 6.9% boost in accuracy shows that this has a significant impact on the results.

6.4 The Future of AI in Team Sports

Our research has begun to explore how AI can effectively be used in team sports. Although we have made several contributions in this area, open challenges remain. The first of these could be to further explore the defensive side of team sports. The majority of the work presented in this thesis has focused on the attacking side of the games when the teams are in possession of the ball. When a team is attacking there are opposition defenders aiming to prevent the attacking from team scoring points, work focused on the defending team is yet to be fully explored. Defensive contributions are often hard to quantify due to them preventing things from happening rather than creating goals or points. Work presented in (Merhej et al. 2021) begins to explore this problem in football by predicting what would have happened if not for a defensive action.

With the models we have presented, the next step would be to ensure that these could be applied and used in the real-world by teams, leagues and governing bodies in the sporting world. A key area where these could be used would be to add assurance to subjective decision-making that sports teams make. The decisions that are made in sports are highlighted in Figure 1.1. By using AI we can help sports teams to predict what the impact of certain decisions will be. For example, using our teamwork model from Chapter 4, we could use this to predict how well a player is likely to work with players in a new team if he/she was to make a transfer. We could also use the long-term decision making processes in Chapter 3 to help teams predict the long-term impact of the decisions they make on the league and financial performances.

A key barrier to entry for the models described in this thesis being used in the real world is the trust in the models. Therefore, a future research area for sports analytics should be Explainable AI (XAI). This would help to serve more information to users so they can see how the AI is generating predictions and therefore build trust in its outputs rather than just serving a black-box model. Another challenge is how to present the uncertainty of a prediction because the prediction of human performance cannot be exact.

The type of data that is being collected in sports is also changing drastically. Instead of the "event-based" data that we have used for a lot of the research in this thesis, there has now been the introduction of tracking data. Tracking data collects the location of all players and the ball up to 20 times a second throughout the game. For example in a game of football with 22 players and a ball, this would be 2,484,000 data points per game. This presents an extremely rich dataset of human teams working together over long periods, that can be learnt from for performance purposes. This new data also allows more physical metrics to be collected on players around their sprint speeds and work rate throughout the game. Another example of how this can be used is shown in Figure 6.1 which shows a pitch control model[1] showing the areas of the pitch a team is deemed to be in control of at any given moment which can help coaches to improve the movement of players when not in possession of the ball.

There is scope for future work by applying the research in sports analytics in other domains which share similar problems to sports (e.g., security, politics and emergency response). Sports offer rich datasets over long periods of time which can be utilised to test many different theoretical AI models on real-world data. For example, the models we have presented in each chapter of this thesis could have wider applications in new domains. The tactical optimisation models in Chapter 3 could be used as a framework to evaluate decisions in the business world as a way of evaluating the objectives and decisions made and their potential impact on given performance indicators in that company. The teamwork model discussed in Chapter 4 could be applied to teams in other settings such as emergency response, both to value the contribution of agents as well as forming effective teams. The NLP model in Chapter 5 could be applied to predict

[1] Quantifying pitch control by William Spearman in 2016: https://www.researchgate.net/publication/334849056$_{Quanti}$

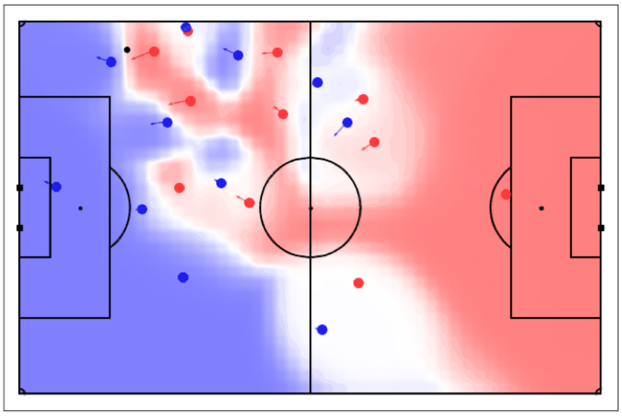


FIGURE 6.1: Example of Pitch Control in Football.

elections based on sentiment in the media and social media channels combined with statistical models such as those used on *fivethirtyeight.com*.[2]

Our work also exposes a number of wider questions that could be explored. Many of the model outputs in this space to help aid humans make decisions rather than fully optimise decision making. Therefore, this poses the question of what is the best way to combine AI and human decision making? The AI models can learn from the way humans make decisions and sports so a feedback loop for the modelling would be important. Also, how would the landscape of team sports change if every team deployed AI systems; would this improve the standard of the players/teams? Finally, much of the work discussed in this thesis is mainly focused on the top-end of professional team sports, but how can new AI and datasets be used to help at the grassroots level? It would be beneficial to the sports to help bring in more players and to promote healthier lifestyles.

[2]https://fivethirtyeight.com/features/how-fivethirtyeights-2020-presidential-forecast-works-and-whats-different-because-of-covid-19.

Chapter 7

Conclusions

In this thesis, we have explored how AI can be used in team sports and how the wealth of data in sports can benefit AI by providing a testbed for new models and algorithms. Specifically, we have covered three main topics. We have shown how game theoretic techniques can be used to improve tactical decision making in football both for individual games and in the long term over a whole season. We have shown how we can extract the value of teamwork and form efficient teams to maximise the teamwork between agents based on interactions they make in a game. Finally, we showed how NLP can be used to improve the accuracy of traditional statistical methods for match prediction.

Our research has shown a number of key findings. In Chapter 3, we have shown that by using a Bayesian game to model pre-game tactics in football and a stochastic game to model in-game tactics, we can increase the chances of a team winning an individual game by up to 16.1%. When we begin to consider the long-term impacts of decisions by simulating full seasons to add more context, we can increase a team's mean expected finishing distribution in the league by up to 35.6%. Next in Chapter 4, we have described a novel approach to team formation based on the value of inter-agent interactions and the notion of teamwork between agents in a team. When evaluated with real-world data from teams of football players, our model is shown to predict the performance of teams by up to 46% more accurately than models that ignore inter-agent interactions. Finally, in Chapter 5, we have shown that when we learn from human experts in the media we are able to boost the traditional statistical methods by 6.9% in terms of match-outcome accuracy and can identify a greater number of rare outcomes.

This thesis, as well as similar work in (Decroos 2020), has begun to fully explore the ways that AI can be used to fully utilise the data that is now being collected across team sports. Practically, we have made contributions by showing that by using event-based data in team sports we can create new insights and learn more about the ways players and tactics contribute to the overall match outcomes. We can also use AI to use new datasets that previously could not add value such as text data in the media and on social network sites. Theoretically, we have made contributions by presenting new methods to form teams based on inter-agent interactions. We have also shown how real-world games can be modelled using multi-step game theoretic techniques by feeding the outputs from a Bayesian game into a stochastic game. The growing availability and innovations in data collection in team sports will continue to generate interesting problems that will benefit both the AI and sports analytics communities.

In conclusion, the work presented in this thesis has exposed a number of novel models to benefit the AI and sports analytics communities. We have shown how we can support and optimise the tactical decision making processes made in team sports by predicting both short and long-term impacts. We have shown how we can extract the value of teamwork and how well agents work together. By so doing, we present a novel team formation method that considers *interactional alignment* to maximise the teamwork value of agent pairs. Also, we have shown that by using contextual text information from the media (both traditional and social) we can boost the prediction accuracy of standard match outcome prediction models in team sport. Finally, we have exposed future research areas for the continued study of applied AI in team sports.

Bibliography

Alcorn, M.A. (2018). “(batter—pitcher)2vec: Statistic-Free Talent Modeling With Neural Player Embeddings”. In: *MIT Sloan Sports Analytics Conference.*

Anderson, C. and D. Sally (2014). *The Numbers Game: Why Everything You Know About Football is Wrong*. Penguin.

Andrienko, G., N. Andrienko, G Anzer, P. Bauer, G. Budziak, G. Fuchs, D. Hecker, H. Weber, and S. Wrobel (2019). “Constructing spaces and times for tactical analysis in football”. In: *IEEE transactions on visualization and computer graphics.*

Avsar, Z.M. and M. Baykal-Gürsoy (2002). “Inventory control under substitutable demand: A stochastic game application”. In: *Naval Research Logistics (NRL)* 49, pp. 359–375.

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

Baio, G. and M. Blangiardo (2008). “Bayesian hierarchical model for the prediction of football results”. In: *Journal of Applied Statistics* 37.2, pp. 253–264.

Baker, R. and I. McHale (2013). “Forecasting exact scores in National Football League games”. In: *International Journal of Forecasting* 29.1, pp. 112–130.

Balasubramanian, N., S. Soderland, and O. Etzioni (2012). “Rel-grams: a probabilistic model of relations in text”. In: *Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction*, pp. 101–105.

Bandura, A. (2008). “Observational learning”. In: *The international encyclopedia of communication.*

Banko, M., M.J. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni (2007). “Open information extraction from the web.” In: *IJCAI-07*. Vol. 7, pp. 2670–2676.

Bate, R. (1988). “Football chance: tactics and strategy”. In: *Science and Football: Proceedings of the First World Congress of Science and Football.*

Beal, R., G. Chalkiadakis, T.J. Norman, and S.D. Ramchurn (2020). “Optimising Game Tactics for Football”. In: *19th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 141–149.

Beal, R., G. Chalkiadakis, T.J. Norman, and S.D. Ramchurn (2021). “Optimising Long-Term Outcomes using Real-World Fluent Objectives: An Application to Football”. In: *Proceedings of the 20th International Conference on Autonomous Agents and Multi Agent Systems*, pp. 196–204.

Beal, R., N. Changder, T.J. Norman, and S.D. Ramchurn (2020). “Learning the Value of Teamwork to Form Efficient Teams”. In: *Thirty-Fourth AAAI Conference on Artificial Intelligence*, pp. 7063–7070.

Beal, R., S.E. Middleton, T.J. Norman, and S.D. Ramchurn (2021). “Combining Machine Learning and Human Experts to Predict Match Outcomes in Football: A Baseline Model”. In: *Thirty-Third Conference on Innovative Applications of Artificial Intelligence*, pp. 15447–15451.

Beal, R., T.J. Norman, and S.D. Ramchurn (2019). “Artificial intelligence for team sports: a survey”. In: *The Knowledge Engineering Review*. Vol. 34.

– (2020a). “A critical comparison of machine learning classifiers to predict match outcomes in the NFL”. In: *International Journal of Computer Science in Sport* 19.2.

– (2020b). “Optimising daily fantasy sports teams with artificial intelligence”. In: *International Journal of Computer Science in Sport* 19.2.

Beilock, S. and R. Gray (2007). “Why do athletes choke under pressure?” In: *G. Tenenbaum R. C. Eklund (Eds.), Handbook of sport psychology*, pp. 425–444.

Bistaffa, F., A. Farinelli, J. Cerquides, J.A. Rodrıguez-Aguilar, and S.D. Ramchurn (2017). “Algorithms for graph-constrained coalition formation in the real world”. In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 8.4, p. 60.

Bistaffa, F., A. Farinelli, G. Chalkiadakis, and S.D. Ramchurn (2017). “A cooperative game-theoretic approach to the social ridesharing problem”. In: *Artificial Intelligence* 246, pp. 86–117.

Blei, D.M., A.Y. Ng, and M.I. Jordan (2003). “Latent Dirichlet Allocation”. In: *Journal of Machine Learning Research* 3, pp. 993–1022.

Bojinov, I. and L. Bornn (2016). “The Pressing Game: Optimal Defensive Disruption in Soccer”. In: *MIT Sloan Sports Analytics Conference*.

Bollen, J., H. Mao, and X. Zeng (2011). “Twitter mood predicts the stock market”. In: *Journal of Computational Science* 2.1, pp. 1–8.

Boon, B. H. and G Sierksma (2003). “Team formation: Matching quality supply and quality demand”. In: *European Journal of Operational Research* 148, pp. 277–292.

Bornn, L., P. Ward, and D. Norman (2019). “Training Schedule Confounds the Relationship between Acute:Chronic Workload Ratio and Injury”. In: *MIT Sloan Sports Analytics Conference*.

Borsa, D., N. Heess, B. Piot, S. Liu, L. Hasenclever, R. Munos, and O. Pietquin (2019). “Observational Learning by Reinforcement Learning”. In: *Proceedings of the 18th International Conference on Autonomous Agents and Multi Agent Systems*, pp. 1117–1124.

Boswell, J. (2008). “Fantasy Sports: A Game of Skill That Is Implicitly Legal under State Law, and Now Explicitly Legal under Federal Law”. In: *Cardozo Arts Entertainment Law Journal* 25, pp. 1257–1278.

Boulier, B. L. and H.O. Stekler (2003). “Predicting the outcomes of National Football League games”. In: *International Journal of Forecasting* 19.2, pp. 257–270.

Bouzarth, E., B. Grannan, J. Harris, A. Hartley, K. Hutson, and E. Morton (2021). “Swing shift: a mathematical approach to defensive positioning in baseball”. In: *Journal of Quantitative Analysis in Sports* 17.1, pp. 47–55.

Bowen, L., A.S. Gross, M. Gimpel, and F.X. Li (2016). “Accumulated workloads and the acute:chronic workload ratio relate to injury risk in elite youth football players”. In: *British Journal of Sports Medicine* 51.5, pp. 452–459.

Bransen, L. and J. Van Haaren (2020). “Player Chemistry: Striving for a Perfectly Balanced Soccer Team”. In: *MIT Sloan Sports Analytics Conference.*

Breiman, Leo (2001). “Random Forests”. In: *Machine Learning.* 45, pp. 5–32.

Burke, B. (2019). “DeepQB: Deep Learning with Player Tracking to Quantify Quarterback Decision-Making & Performance”. In: *MIT Sloan Sports Analytics Conference.*

Campbell, John Y and Robert J Shiller (1988). “Stock prices, earnings, and expected dividends”. In: *The Journal of Finance* 43.3, pp. 661–676.

Cappellari, L. and S.P. Jenkins (2003). “Multivariate probit regression using simulated maximum likelihood”. In: *The Stata Journal* 3.3, pp. 278–294.

Cervone, D., A. D’Amour, L. Bornn, and K. Goldsberry (2014). “POINTWISE: Predicting Points and Valuing Decisions in Real Time with NBA Optical Tracking Data”. In: *MIT Sloan Sports Analytics Conference.*

Chakraborty, S., A. Venkataraman, S. Jagabathula, and L. Subramanian (2016). “Predicting socio-economic indicators using news events”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, pp. 1455–1464.

Chalkiadakis, G. and C. Boutilier (2012). “Sequentially optimal repeated coalition formation under uncertainty”. In: *Autonomous Agents and Multi-Agent Systems* 24.3, pp. 441–484.

Chen, T., M. Kwiatkowska, A. Simaitis, and C. Wiltsche (2013). “Synthesis for multi-objective stochastic games: An application to autonomous urban driving”. In: *International Conference on Quantitative Evaluation of Systems*, pp. 322–337.

Clarke, S. and Norman (1995). “Home ground advantage of individual clubs in English soccer”. In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 44.4, pp. 509–521.

Clarke, S.R. and J.M. Norman (1995). “Home ground advantage of individual clubs in English soccer”. In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 44.4, pp. 509–521.

Collobert, R. and J. Weston (2008). “A unified architecture for natural language processing: Deep neural networks with multitask learning”. In: *Proceedings of the 25th international conference on Machine learning*. ACM, pp. 160–167.

Constantinou, A. C., N. Fenton, and M. Neil (2012). “pi-football: a Bayesian network model for forecasting Association Football match outcomes”. In: *Knowledge-Based Systems* 36, pp. 322–339.

– (2013). “Profiting from an inefficient association football gambling market: Prediction, risk and uncertainty using Bayesian networks”. In: *Knowledge-Based Systems* 50, pp. 60–86.

Correia, V., D. Araújo, C. Craig, and P. Passos (2011). “Prospective information for pass decisional behavior in rugby union.” In: *Human movement science* 30.5, pp. 984–97.

Coulom, Rémi (2007). “Computing Elo Ratings of Move Patterns in the Game of Go”. In: *ICGA Journal* 30.4.

Crowder, M., M. Dixon, A. Ledford, and M. Robinson (2002). “Dynamic Modelling and Prediction of English Football League Matches for Betting”. In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 51.2, pp. 157–168.

Cui, L., F. Wei, and M. Zhou (2018). “Neural Open Information Extraction”. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 407–413.

Cust, E., A. Sweeting, K. Ball, and S. Robertson (2018). “Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance”. In: *Journal of Sports Sciences* 37.5, pp. 568–600.

Dang, V.D., R.K. Dash, A. Rogers, and N.R. Jennings (2006). “Overlapping coalition formation for efficient data fusion in multi-sensor networks”. In: *Proceedings of the 21st National Conference on Artificial Intelligence*. Vol. 1, pp. 635–640.

Decroos, T. (2020). “Soccer Analytics Meets Artificial Intelligence: Learning Value and Style from Soccer Event Stream Data”. In: *https://lirias.kuleuven.be/retrieve/587585/*.

Decroos, T., L. Bransen, J. Van Haaren, and J. Davis (2019). “Actions speak louder than goals: Valuing player actions in soccer”. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1851–1861.

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

Dixon, M. and P. Pope (2004). “The value of statistical forecasts in the UK association football betting market”. In: *International Journal of Forecasting* 20.4, pp. 697–711.

Dixon, M. and M. Robinson (1998). “A birth process model for association football matches”. In: *Journal of the Royal Statistical Society: Series D (The Statistician)*. Vol. 47, pp. 523–538.

Dobson, S. and B. Gerrard (1999). “The Determination of Player Transfer Fees in English Professional Soccer”. In: *Journal of Sports Management* 13.4, pp. 259–279.

Drawer, S. and C.W. Fuller (2002a). “An economic framework for assessing the impact of injuries in professional football.” In: *Safety Science* 40.6, pp. 537–556.

– (2002b). “Evaluating the level of injury in English professional football using a risk based assessment approach”. In: *British Journal of Sports Medicine* 36.6, pp. 446–451.

Dreiseitl, S. and L. Ohno-Machado (2002). “Logistic regression and artificial neural network classification models: a methodology review”. In: *Journal of biomedical informatics*. Vol. 35, pp. 352–359.

Dunis, C., P. Middleton, A. Karathanasopolous, and K. Theofilatos (2016). *Artificial intelligence in financial markets*. Springer.

Ermoliev, Y.M. and R.B. Wets (1988). *Numerical techniques for stochastic optimization*. Springer-Verlag.

Fang, F., P. Stone, and M. Tambe (2015). “When Security Games Go Green: Designing Defender Strategies to Prevent Poaching and Illegal Fishing”. In: *Proceedings of the 24th International Conference on Artificial Intelligence*, pp. 2589–2595.

Farzindar, A. and D. Inkpen (2015). “Natural language processing for social media”. In: *Synthesis Lectures on Human Language Technologies* 8.2, pp. 1–166.

Felsen, P. and P. Lucey (2017). “Body Shots: Analyzing Shooting Styles in the NBA using Body Pose”. In: *MIT Sloan Sports Analytics Conference*.

Fernández, J., L. Bornn, and D. Cervone (2019). “Decomposing the Immeasurable Sport: A deep learning expected possession value framework for soccer”. In: *MIT Sloan Sports Analytics Conference*.

Fister, I. Jr, K. Ljubic, P.N. Suganthan, M. Perc, and I. Fister (2015). “Computational intelligence in sports: Challenges and opportunities within a new research domain”. In: *Applied Mathematics and Computation* 262, pp. 178–186.

Fitzpatrick, Erin L and Ronald G Askin (2005). “Forming effective worker teams with multi-functional skill requirements”. In: *Computers & Industrial Engineering* 48.3, pp. 593–608.

Franchi, E. and A. Poggi (2012). “Multi-agent systems and social networks”. In: *Handbook of Research on Business Social Networking: Organizational, Managerial, and Technological Dimensions*, pp. 84–97.

Fuller, C.W. (2018). “Modeling the impact of players’ workload on the injury-burden of English Premier League football clubs.” In: *Scandinavian Journal of Medicine and Science in Sports* 28.6, pp. 1715–1721.

Gall, F. le, C. Carling, M. Williams, and T. Reilly (2010). “Anthropometric and fitness characteristics of international, professional and amateur male graduate soccer players from an elite youth academy.” In: *Journal of science and medicine in sport* 13.1, pp. 90–95.

Ganguly, D., D. Roy, M. Mitra, and G. Jones (2015). “Word embedding based generalized language model for information retrieval”. In: *Proceedings of the 38th international*

ACM SIGIR conference on research and development in information retrieval. ACM, pp. 795–798.

Ganguly, S. and N. Frank (2018). “The Problem with Win Probability”. In: *MIT Sloan Sports Analytics Conference.*

Gardiner, C. (2009). *Stochastic Methods.* Springer.

Gaston, M.E. and M. DesJardins (2005). “Agent-organized networks for dynamic team formation”. In: *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 230–237.

– (2008). “The effect of network structure on dynamic team formation in multi-agent systems”. In: *Computational Intelligence* 24.2, pp. 122–157.

Gauvain, J.L. and C.H. Lee (1994). “Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains”. In: *IEEE transactions on speech and audio processing.* Vol. 2, pp. 291–298.

Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin (2004). *Bayesian Data Analysis.* New York: Chapman and Hall.

Glickman, M.E. and H.S. Stern (1996). “A State-Space Model for National Football League Scores”. In: *Journal of the American Statistical Association* 93, pp. 25–35.

Goel, S., D.M. Pennock, D.M. Reeves, and C. Yu (2008). “Yoopick: A Combinatorial Sports Prediction Market”. In: *Twenty-Third AAAI Conference on Artificial Intelligence.*

Graham, I. and H. Stott (2008). “Predicting bookmaker odds and efficiency for UK football”. In: *Applied Economics* 40.1, pp. 99–109.

Gréhaigne, J.F., P. Godbout, and D. Bouthier (1999). “The foundations of tactics and strategy in team sports”. In: *Journal of teaching in physical education.* Vol. 18, pp. 159–174.

Gürpınar-Morgan, W., D. Dinsdale, J. Gallagher, A. Cherukumudi, and P. Lucey (2020). “You Cannot Do That Ben Stokes: Dynamically Predicting Shot Type in Cricket Using a Personalized Deep Neural Network”. In: *Proceedings of the 14th MIT Sloan Sports Analytics Conference.*

Hägglund, M, M Waldén, and J Ekstrand (2006). “Previous injury as a risk factor for injury in elite football: a prospective study over two consecutive seasons”. In: *British Journal of Sports Medicine* 40.9, pp. 767–772.

– (2009). “Injury incidence and injury patterns in professional football: the UEFA injury study”. In: *British Journal of Sports Medicine* 45.7, pp. 553–558.

– (2013). “Fewer ligament injuries but no preventive effect on muscle injuries and severe injuries: an 11-year follow-up of the UEFA Champions League injury study”. In: *British Journal of Sports Medicine* 47.12, pp. 732–737.

Haghighat, M., H. Rastegari, and N. Nourafza (2013). “A review of data mining techniques for result prediction in sports”. In: *Advances in Computer Science: An International Journal* 2.5, pp. 7–12.

Hanson, R. D. (2007). “Logarithmic Market Scoring Rules for Modular Combinatorial Information Aggregation”. In: *Journal of Prediction Markets* 1, pp. 1–15.

Harville, D. (1976). “The Use of Linear-Model Methodology to Rate High School or College Football Teams”. In: *Journal of the American Statistical Association* 72.358, pp. 278–289.

– (1980). “Predictions for National Football League Games Via Linear-Model Methodology”. In: *Journal of the American Statistical Association* 75.371, pp. 516–524.

Haugh, M. B. and R. Singal (2018). “How to Play Strategically in Fantasy Sports (and Win)”. In: *MIT Sloan Sports Analytics Conference*.

Hendricks, W., L. DeBrock, and R. Koenker (2003). “Uncertainty, Hiring, and Subsequent Performance: The NFL Draft”. In: *Journal of Labor Economics* 21.4, pp. 857–886.

Hobbs, J., P. Power, L. Sha, H. Ruiz, and P. Lucey (2018). “Quantifying the Value of Transitions in Soccer via Spatiotemporal Trajectory Clustering”. In: *MIT Sloan Sports Analytics Conference*.

Hochreiter, S. and J. Schmidhuber (1997). “Long Short-Term Memory”. In: *Neural Computation* 9.8, pp. 1735–1780.

Hosmer Jr, David W, Stanley Lemeshow, and Rodney X Sturdivant (2013). *Applied logistic regression*. Vol. 398.

Hucaljuk, J. and A. Rakipović (2011). “Predicting football scores using machine learning techniques”. In: *Proceedings of the 34th International Convention MIPRO*.

Hullin, B.T., T.J. Gabbett, D.W. Lawson, P. Caputi, and J.A. Sampson (2016). “The acute:chronic workload ratio predicts injury: high chronic workload may decrease injury risk in elite rugby league players”. In: *British Journal of Sports Medicine* 50.4, pp. 231–236.

Jang, M. and S. Cho (1999). “Ensemble learning using observational learning theory”. In: *International Joint Conference on Neural Networks*. Vol. 2, pp. 1287–1292.

Jayalath, K. P. (2018). “A machine learning approach to analyze ODI cricket predictors”. In: *Journal of Sports Analytics* 4.1, pp. 73–84.

Jayantha, S. B., A. Anthony, G. Abhilashab, N. Shaikb, and G. Srinivasaa (2018). “A team recommendation system and outcome prediction for the game of cricket”. In: *Journal of Sports Analytics* 4.1, pp. 263–273.

Ji, H. and R. Grishman (2011). “Knowledge base population: Successful approaches and challenges”. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 1148–1158.

Jiang, Y. and J.C. Jiang (2013). “Understanding social networks from a multiagent perspective”. In: *IEEE Transactions on Parallel and Distributed Systems*. Vol. 25. 10, pp. 2743–2759.

Jiang, Y., Y. Zhou, and W. Wang (2012). “Task allocation for undependable multiagent systems in social networks”. In: *IEEE Transactions on Parallel and Distributed Systems*. Vol. 24. 8, pp. 1671–1681.

Jiang, Z., P. Yin, and G. Neubig (2019). “Improving Open Information Extraction via Iterative Rank-Aware Learning”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 5295–5300.

Jordan, J.D., S.H. Melouk, and M.B. Perry (2009). “Optimizing football game play calling”. In: *Journal of Quantitative Analysis in Sports*. Vol. 5.

Joseph, A., N. Fenton, and M. Neil (2006). “Predicting football results using Bayesian nets and other machine learning techniques”. In: *Knowledge-Based Systems* 19.7, pp. 544–553.

Kaluarachchi, A. and S. V. Aparna (2011). “CricAI: A classification based tool to predict the outcome in ODI cricket”. In: *Fifth International Conference on Information and Automation for Sustainability*, pp. 250–255.

Kardeş, E., F. Ordóñez, and R.W. Hall (2011). “Discounted robust stochastic games and an application to queueing control”. In: *Operations research*. Vol. 59, pp. 365–382.

Karlis, D. and I. Ntzoufras (2008). “Bayesian modelling of football outcomes: using the Skellam’s distribution for the goal difference”. In: *IMA Journal of Management Mathematics* 20.2, pp. 133–145.

Kelly, D, G.F. Coughlan, B.S. Green, and B. Caulfield (2012). “Automatic detection of collisions in elite level rugby union using a wearable sensing device”. In: *Sports Engineering* 15.2, pp. 81–92.

Korte, B. and J. Vygen (2012). *The Knapsack Problem - Combinatorial Optimization*, pp. 459–470.

Landers, J.R. and B. Duperrouzel (2018). “Machine Learning Approaches to Competing in Fantasy Leagues for the NFL”. In: *IEEE Transactions on Games*.

Le, H.M., P. Carr, Y. Yue, and P. Lucey (2017). “Data-Driven Ghosting using Deep Imitation Learning”. In: *MIT Sloan Sports Analytics Conference*.

Lehtokangas, M., J. Saarinen, K. Kaski, and P. Huuhtanen (1995). “Initializing Weights of a Multilayer Perceptron Network by Using the Orthogonal Least Squares Algorithm”. In: *Neural Computation* 7.5, pp. 982–999.

Leung, C. K. and K. W. Joseph (2014). “Sports data mining: predicting results for the college football games”. In: *Procedia Computer Science* 35, pp. 710–719.

Lin, C. and Y. He (2009). “Joint sentiment/topic model for sentiment analysis”. In: *Proceedings of the 18th ACM conference on Information and knowledge management*, pp. 375–384.

Llana, S., P. Madrero, and J. Fernández (2020). “The right place at the right time: Advanced off-ball metrics for exploiting an opponent’s spatial weaknesses in soccer”. In: *Proceedings of the 14th MIT Sloan Sports Analytics Conference*.

Lucey, P., A. Bialkowski, P. Carr, E. Foote, and I. A. Matthews (2012). "Characterizing Multi-Agent Team Behavior from Partial Team Tracings: Evidence from the English Premier League". In: *Twenty-Sixth AAAI Conference on Artificial Intelligence*.

Lysens, R., A. Steverlynck, Y. van den Auweele, J. Lefevre, L. Renson, A. Claessens, and M. Ostyn (1984). "The Predictability of Sports Injuries". In: *Sports Medicine* 1.1, pp. 6–10.

Maher, M.J. (1982). "Modelling association football scores". In: *Statistica Neerlandica* 36.3, pp. 109–118.

Malhi, A. and R.X. Gao (2004). "PCA-based feature selection scheme for machine defect classification". In: *IEEE Transactions on Instrumentation and Measurement* 53.6, pp. 1517–1525.

Manevitz, L.M. and M. Yousef (2001). "One-class SVMs for document classification". In: *Journal of Machine Learning Research* 2, pp. 139–154.

Manning, C.D. and H. Schütze (1999). *Foundations of statistical natural language processing*. MIT press.

Massey, C. and R.H. Thaler (2013). "The loser's curse: Decision making and market efficiency in the National Football League draft". In: *Management Science* 59.7, pp. 1479–1495.

Matthews, T., S.D. Ramchurn, and G. Chalkiadakis (2012). "Competing with humans at fantasy football: Team formation in large partially-observable domains". In: *Twenty-Sixth AAAI Conference on Artificial Intelligence*, pp. 1394–1400.

Mausam, M. (2016). "Open information extraction systems and downstream applications". In: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 4074–4077.

McCabe, A. (2002). "An Artificially Intelligent Sports Tipper". In: *Australian Joint Conference on Artificial Intelligence* 2557, pp. 718–718.

McCabe, A. and J. Trevathan (2008). "Artificial Intelligence in Sports Prediction". In: *Fifth International Conference on Information Technology: New Generations (itng 2008)*.

McHale, I. and P. Scarf (2011). "Modelling the dependence of goals scored by opposing teams in international soccer matches". In: *Statistical Modelling* 11.3, pp. 219–236.

Merhej, C., R. Beal, S.D. Ramchurn, and T. Matthews (2021). "What Happened Next? Using Deep Learning to Value Defensive Actions in Football Event-Data". In: *KDD-21*.

Miljković, D., L. Gajić, A. Kovačević, and Z. Konjović (2010). "The use of data mining for basketball matches outcomes prediction". In: *IEEE 8th International Symposium on Intelligent Systems and Informatics*, pp. 309–312.

Mooney, C.Z. (1997). *Monte carlo simulation*. Sage.

Nelsen, R.B. (2006). *An Introduction to Copulas*. Springer.

Pak, A. and P. Paroubek (2010). “Twitter as a corpus for sentiment analysis and opinion mining.” In: *LREC*. Vol. 10, pp. 1320–1326.

Pang, B., L. Lee, and S. Vaithyanathan (2002). “Thumbs up?: sentiment classification using machine learning techniques”. In: *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language processing*, pp. 79–86.

Pang, Bo, Lillian Lee, et al. (2008). “Opinion mining and sentiment analysis”. In: *Foundations and Trends® in Information Retrieval* 2.1–2, pp. 1–135.

Paruchuri, Praveen, Jonathan P Pearce, Janusz Marecki, Milind Tambe, Fernando Ordonez, and Sarit Kraus (2008). “Playing games for security: An efficient exact algorithm for solving Bayesian Stackelberg games”. In: *Proceedings of the 7th International Conference on Autonomous Agents and Multi Agent Systems*, pp. 895–902.

Pochet, Y. and L.A. Wolsey (2006). *Production planning by mixed integer programming.*

Podlog, L., H. Pollack C.F. Buhler, F.N. Hopkins, and P.R. Burgess (2015). “Time trends for injuries and illness, and their relation to performance in the National Basketball Association.” In: *Journal of Science and Medicine in Sports* 18.3, pp. 278–282.

Power, P., A. Cherukumudi, S. Ganguly, F. Wei, L. Sha, J. Hobbs, H. Ruiz, and P. Lucey (2019). “Trading Places – Simulating Goalkeeper Performance using Spatial & Body-Pose Data”. In:

Power, P., J. Hobbs, H. Ruiz, Z. Wei, and P. Lucey (2018). “Mythbusting Set-Pieces in Soccer”. In: *MIT Sloan Sports Analytics Conference.*

Power, P., H. Ruiz, Z. Wei, and P. Lucey (2016). ““Not All Passes Are Created Equal:” Objectively Measuring the Risk and Reward of Passes in Soccer from Tracking Data”. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1605–1613.

Price, R.J., R.D. Hawkins, M.A. Hulse, and A. Hodson (2004). “The Football Association medical research programme: an audit of injuries in academy youth football”. In: *British journal of sports medicine* 38.4, pp. 466–471.

Pujol-Gonzalez, M., J. Cerquides, P. Meseguer, J.A. Rodrıguez-Aguilar, and M. Tambe (2013). “Engineering the decentralized coordination of UAVs with limited communication range”. In: *Conference of the Spanish Association for Artificial Intelligence.* Springer, pp. 199–208.

Radinsky, K. and E. Horvitz (2013). “Mining the web to predict future events”. In: *Proceedings of the sixth ACM international conference on Web search and data mining.* ACM, pp. 255–264.

Ramchurn, S.D., T.D. Huynh, F. Wu, Y. Ikuno, J. Flann, L. Moreau, J.E. Fischer, W. Jiang, T. Rodden, E. Simpson, et al. (2016). “A disaster response system based on human-agent collectives”. In: 57, pp. 661–708.

Ramchurn, S.D., M. Polukarov, A. Farinelli, C. Truong, and N.R. Jennings (2010). “Coalition formation with spatial and temporal constraints”. In: *Proceedings of the*

9th International Conference on Autonomous Agents and Multiagent Systems: volume 3-Volume 3, pp. 1181–1188.

Ramchurn, S.D., A. Rogers, K. Macarthur, A. Farinelli, P. Vytelingum, I. Vetsikas, and N.R. Jennings (2008). "Agent-based coordination technologies in disaster management". In: *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems: demo papers*. International Foundation for Autonomous Agents and Multiagent Systems, pp. 1651–1652.

Ranganathan, A. and R.H. Campbell (2003). "An infrastructure for context-awareness based on first order logic". In: *Personal and Ubiquitous Computing* 7.6, pp. 353–364.

Richmond, J (1974). "Estimating the Efficiency of Production". In: *International Economic Review* 15.2, pp. 515–21.

Robberechts, P., J. Van Haaren, and J. Davis (2021). "A Bayesian Approach to In-Game Win Probability in Soccer". In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 3512–3521.

Rue, H. and O. Salvesen (2000). "Prediction and retrospective analysis of soccer matches in a league". In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 49.3, pp. 399–418.

Scerri, P., A. Farinelli, S. Okamoto, and M. Tambe (2005). "Allocating tasks in extreme teams". In: *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 727–734.

Schapire, R. E. (2003). *The boosting approach to machine learning: An overview.* Springer New York, pp. 149–171.

Scholkopf, B., K.K. Sung, C.J.C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik (1997). "Comparing support vector machines with Gaussian kernels to radial basis function classifiers". In: *IEEE transactions on Signal Processing*. Vol. 45, pp. 2758–2765.

Schumaker, R.P., A.T. Jarmoszko, and C.S. Labedz Jr (2016). "Predicting wins and spread in the Premier League using a sentiment analysis of twitter". In: *Decision Support Systems* 88, pp. 76–84.

Sebastiani, F. (2002). "Machine learning in automated text categorization". In: *ACM computing surveys (CSUR)* 34.1, pp. 1–47.

Seidl, T., A. Cherukumudi, A. Hartnett, P. Carr, and P. Lucey (2018). "Bhostgusters: Realtime Interactive Play Sketching with Synthesized NBA Defenses". In: *MIT Sloan Sports Analytics Conference*.

Shapley, L.S. (1953). "Stochastic games". In: *Proceedings of the national academy of sciences* 39.10, pp. 1095–1100.

Shaw, L. and S. Gopaladesikan (2020). "Routine Inspection: A Playbook for Corner Kicks". In: *International Workshop on Machine Learning and Data Mining for Sports Analytics*. Springer, pp. 3–16.

Shi, Z., S. Moorthy, and A. Zimmermann (2013). “Predicting NCAAB match outcomes using ML techniques – some results and lessons learned”. In: *Sports Analytics Workshop MLSA PKDD/ECML*.

Shieh, E.A., D. An, R. Yang, M. Tambe, C. Baldwin, J. DiRenzo, B. Maule, and G. Meyer (2012). “PROTECT: An application of computational game theory for the security of the ports of the United States”. In: *Twenty-Seventh AAAI Conference on Artificial Intelligence*, pp. 2173–2179.

Shoham, Y. and K. Leyton-Brown (2008). *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.

Silva, V. De, M.P. Caine, J. Skinner, S. Dogan, A. Kondoz, T. Peter, E. Axtell, M. Birnie, and B. Smith (2018). “Player tracking data analytics as a tool for physical performance management in football: A case study from Chelsea Football Club Academy”. In: *Sports* 6.4, p. 130.

Silver, D., A. Huang, C.J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, and M. Lanctot (2016). “Mastering the game of Go with deep neural networks and tree search”. In: *Nature*. Vol. 529, pp. 484–489.

Silver, N. (2012). *The signal and the noise: the art and science of prediction*. Penguin UK.

Sim, K.M. and C.Y Choi (2003). “Agents that react to changing market situations”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. Vol. 33, pp. 188–201.

Simpson, E., S. Roberts, I. Psorakis, and A. Smith (2013). “Dynamic Bayesian Combination of Multiple Imperfect Classifiers”. In: *Decision Making and Imperfection*, pp. 1–35.

Sinha, A., F. Fang, B. An, C. Kiekintveld, and M. Tambe (2018). “Stackelberg Security Games: Looking Beyond a Decade of Success”. In: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pp. 5494–5501.

Sinha, S., C. Dyer, K. Gimpel, and N.A. Smith (2013). “Predicting the NFL using Twitter”. In: *ECML/PKDD 2013 Workshop on Machine Learning and Data Mining for Sports Analytics*.

Skellam, J. G. (1948). “A Probability Distribution Derived from the Binomial Distribution by Regarding the Probability of Success as Variable Between the Sets of Trials”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 10.2, pp. 257–261.

Snidal, D. (1985). “The game theory of international politics”. In: *World Politics*. Vol. 38, pp. 25–57.

Spearman, William (2018). “Beyond expected goals”. In: *Proceedings of the 12th MIT sloan sports analytics conference*, pp. 1–17.

Srinivas, K., B.K. Rani, and A. Govrdhan (2010). “Applications of data mining techniques in healthcare and prediction of heart attacks”. In: *International Journal on Computer Science and Engineering (IJCSE)* 2.02, pp. 250–255.

Stekler, H.O., D. Sendor, and R. Verlander (2010). “Issues in sports forecasting”. In: *International Journal of Forecasting* 26.3, pp. 606–621.

Stöckl, M., T. Seidl, D. Marley, and P. Power (2021). “Making Offensive Play Predictable-Using a Graph Convolutional Network to Understand Defensive Performance in Soccer”. In: *MIT Sloan Sports Analytics Conference*.

Stone, P. and R.S. Sutton (2001). “Scaling reinforcement learning toward RoboCup soccer”. In: *ICML*, pp. 537–544.

Stone, P., R.S. Sutton, and G. Kuhlmann (2005). “Reinforcement learning for robocup soccer keepaway”. In: *Adaptive Behavior*. Vol. 13, pp. 165–188.

Sugar, G. and T. Swenson (2015). “Predicting Optimal Game Day Fantasy Football Teams”. In: URL: `http://cs229.stanford.edu/proj2015/115%5C_report.pdf`.

Suziki, A. K., L. E. B. Salasar, J. G. Leite, and F. Louzada-Neto (2009). “A Bayesian approach for predicting match outcomes: The 2006 (Association) Football World Cup”. In: *Journal of the Operational Research Society* 61.10, pp. 1530–1539.

Synnaeve, G. and P. Bessiere (2011). “A Bayesian model for opening prediction in RTS games with application to StarCraft”. In: *IEEE Conference on Computational Intelligence and Games*, pp. 281–288.

Tumasjan, A., T.O. Sprenger, P.G. Sandner, and I.M. Welpe (2010). “Predicting elections with twitter: What 140 characters reveal about political sentiment”. In: *Fourth international AAAI Conference on Weblogs and Social media*.

Tuyls, K., S. Omidshafiei, P. Muller, Z. Wang, J. Connor, D. Hennes, I. Graham, W. Spearman, T. Waskett, and D. Steel (2021). “Game Plan: What AI can do for Football, and What Football can do for AI”. In: *Journal of Artificial Intelligence Research* 71, pp. 41–88.

Valckenaers, P., J. Sauter, C. Sierra, and J.A. Rodriguez-Aguilar (2007). “Applications and environments for multi-agent systems”. In: *Autonomous Agents and Multi-Agent Systems* 14.1, pp. 61–85.

Van Roy, M., P. Robberechts, W.C. Yang, L. De Raedt, and J. Davis (2021). “Leaving Goals on the Pitch: Evaluating Decision Making in Soccer”. In: *arXiv preprint arXiv:2104.03252*.

Verma, S., S. Vieweg, W.J. Corvey, L. Palen, J.H. Martin, M. Palmer, A. Schram, and K.M. Anderson (2011). “Natural language processing to the rescue? extracting situational awareness tweets during mass emergency”. In: *Fifth International AAAI Conference on Weblogs and Social Media*, pp. 385–392.

Vilar, Luıs, Duarte Araújo, Keith Davids, and Yaneer Bar-Yam (2013). “Science of winning soccer: Emergent pattern-forming dynamics in association football”. In: *Journal of Systems Science and Complexity* 26.1, pp. 73–84.

Wang, Hao, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan (2012). “A system for real-time twitter sentiment analysis of 2012 us presidential election cycle”. In: *Proceedings of the ACL 2012 system demonstrations*, pp. 115–120.

Wang, K.C. and R. Zemel (2016). “Classifying NBA Offensive Plays Using Neural Networks”. In: *MIT Sloan Sports Analytics Conference*.

Ward, P., M. Tankovich, J.S Ramsden, B. Drust, and L. Bornn (2018). “Volume and Intensity are Important Training Related Factors in Injury Incidence in American Football Athletes”. In: *MIT Sloan Sports Analytics Conference*.

Whitaker, G.A., R. Silva, and D. Edwards (2017). “A Bayesian inference approach for determining player abilities in soccer”. In: *Big Data* 6.4, pp. 271–290.

Wold, Svante, Kim Esbensen, and Paul Geladi (1987). “Principal component analysis”. In: *Chemometrics and intelligent laboratory systems* 2.1-3, pp. 37–52.

Xie, B., R. Passonneau, L. Wu, and G. Creamer (2013). “Semantic frames to predict stock price movement”. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 873–883.

Yang, T. Y. and T. Swartz (2004). “A Two-Stage Bayesian Model for Predicting Winners in Major League Baseball”. In: *Journal of Data Science* 2, pp. 61–73.

Yue, Y., P. Lucey, P. Carr, A. Bialkowski, and I. Matthews (2014). “Learning Fine-Grained Spatial Models for Dynamic Sports Play Prediction”. In: *IEEE International Conference on Data Mining*, pp. 670–679.

Zak, T., C. Huang, and J. Siegfried (1979). “Production Efficiency: The Case of Professional Basketball”. In: *The Journal of Business* 52.3, pp. 379–392.

Zheng, S., Y. Yue, and J. Hobbs (2016a). “Generating Long-term Trajectories Using Deep Hierarchical Networks”. In: *Advances in Neural Information Processing Systems 29*, pp. 1543–1551.

– (2016b). “Generating Long-term Trajectories Using Deep Hierarchical Networks”. In: *Advances in Neural Information Processing Systems 29*, pp. 1543–1551.

Appendix A

Appendix

A.1 Published Papers

Below is a list of the publications from AI in team sports research:

- Artificial Intelligence in Team Sports: A Survey; Knowledge Engineering Review 2019
- Learning the Value of Teamwork to Form Efficient Teams; AAAI-20
- Optimising Game Tactics for Football; AAMAS-20
- Optimising Daily Fantasy Sports Teams with AI; International Journal of Computer Science in Sport 2020
- A Critical Comparison of Machine Learning Classifiers to Predict Match Outcomes in the NFL; International Journal of Computer Science in Sport 2020
- Combining Machine Learning and Human Experts to Predict Match Outcomes in Football: A Baseline Model; IAAI-21
- Optimising Long-Term Outcomes using Real-World Fluent Objectives: An Application to Football; AAMAS-21
- What Happened Next? Using Deep Learning to Value Defensive Actions in Football Event-Data; KDD-21

A.2 Papers In Submission

Below is a list of publications currently awaiting review of the Artificial Intelligence Journal (AIJ):

- Optimising Short-Term and Long-Term Team Strategy in Football
- Learning The Value of Teamwork to Form Efficient Teams: An Application to Team Sports

A.3 Team Sports Background

In this section, we detail the key features of various team sports that present opportunities for AI research and impact. Table 1 shows the key aspects of the game that can be used for comparison.[1] In the sections that follow, we give a more detailed background of the six sports that we focus on and the different challenges that each of these presents.[2]

Table A.1: Team Sports Features.

Sport	Game Duration	#	Score Frequency
Association Football	90 minutes (2 halves)	11	69 minutes
American Football	60 minutes (4 quarters)	11	9 minutes
Rugby Union	80 minutes (2 halves)	15	12.5 minutes
Basketball	48 minutes (4 quarters)	5	30 seconds
ODI Cricket	50 overs per team (300 balls)	11	n/a
Baseball	9 innings (Each team bats and fields)	9	n/a

A.3.1 Association Football

In a game of Association Football, or football for short, each team aims to score goals (1 point) against the opposition (by getting the ball into an 24ft x 8ft goal) and the team with the most goals after the game duration wins. Football is the biggest sport in the world, making up 43% of the sports industry. There are hundreds of professional leagues across the world (e.g., the English Premier League (EPL) and Spanish La Liga are two of the worlds most popular leagues). In a classic football league each team plays every other team twice, once at home and once away. This means that the typical season consists of $(2N) - 2$ games, where N is the number of teams in the league (e.g., in the EPL there are 20 teams meaning each team plays 38 games). There are also a number of cup competitions that run alongside the main leagues (e.g., The Champions League and The FA Cup).

A number of factors can affect a game of football such as weather, the quality of pitch, and injuries. There are also a number of tactical decisions (e.g., team formation and style of play) that can increase a team's chances of winning a game. The 11 players are

[1] Score Frequency data sourced from (Anderson and Sally 2014).

[2] Market data sourced from - https://www.atkearney.com/ communications-media-technology/article?/a/the-sports-market.

set up in a formation with 1 goalkeeper and 10 outfield players. An example formation for the outfield players is 4-4-2 which commonly denotes 4 defenders, 4 midfielders and 2 strikers. The team formation is a key decision in football tactics to which effects team performance. Teams can also make in-game player substitutions (up to 3 in a game) which can help change the team's current in-game performance. Injuries happen across the football season, and this can have significant impact on the teams - in the 2016/17 EPL season there were a total of 735 injuries, which are often preventable muscular injuries.[3]

Increasingly, player recruitment plays a big part in modern day football. Players are bought and sold between teams across the world. Youth players are developed through clubs academies until they are ready to play in the first team. They can also be loaned out to other clubs to gain more experience. What makes football different in comparison to the other sports in this thesis is the rarity of goals. This is highlighted in Table 1 where (Anderson and Sally 2014) show that over the 2010/11 season there is a goal scored on average every 69 minutes. Due to this, a draw/tie is much more common in football than in other sports.

A.3.2 American Football

In a game of American Football[3] teams aim to score touchdowns while attacking (worth 6 points), which is followed by a kick (1 point if scored). Teams can also score field goals (3 points) or a safety (2 points). A game-day squad is made up of 45 players split into the offence, defence and special[4] teams. The coach makes a decision on how these players are positioned when on the field of play and usually also makes decisions on what plays to run during the game (where a play is a tactic used to move the ball down the field). Many factors affect teams' performances in American Football such as weather and even the air pressure of the ball.[5]

American Football makes up an estimated 13% of the global sports market. However, it is mainly played in North America where the main professional league is the National Football League (NFL). There are 32 teams that make up the NFL, each team plays 16 games in the regular-season. The teams that do well in the regular-season make it into the playoffs where teams play up to 4 more games to determine the winner of the league. In the NFL, players are traded rather than bought or sold as in football and, instead of having youth teams to develop younger players, players are drafted from the college leagues. Much of the team and player performances in American Football are easier to quantify than other sports in this thesis. This is due to the nature of the game as the

[3] Referred to as American Football throughout, not to be confused with Association Football.
[4] Special teams are units that are on the field during kicking plays.
[5] https://www.vox.com/2015/1/21/7866121/deflated-football-patriots-cheating.

yards that teams gain (which lead to points being scored) or prevent are measured and attributed to each player that contributes.

A.3.3 Rugby Union

In Rugby Union each team aims to score tries[6] half) against the opposition, these are worth 5 points and are followed by a conversion - a kick at between the posts, worth 2 points. Teams can also score points through penalties and drop-goals[7], both worth 3 points. The team with the most points after 80 minutes, wins. Teams are split into forwards and backs where the forwards are the 8 players that make up the scrum.[8] Unlike football, there is a standard way to set up players on a rugby field so there is not a formation decision for the coach to make.[9] There are still many other tactical decisions for the coach to make such as: (e.g., player selection, line-out formation, style of play). Usually club rugby is played in a league format similar to football where each team plays against every other teams both home and away (e.g., in the Aviva Premiership (England) there are 12 teams, each play 22 games in a season). Rugby Union has been the fastest growing sport since it became professional in 1995. It is popular in countries such as Britain, Australia, New Zealand, South Africa. Due to Rugby being a high impact sport, it presents many injury related challenges. In particular, how can the medical teams be assisted and how can players be further monitored.

A.3.4 Basketball

In Basketball, teams aim to score a point by getting the ball in the basket. When scored within a given zone it is worth 2 points, outside of this zone it is worth 3 points. A free throw is worth 1 point. The winning team is the team who accumulate the most points. The main league is the National Basketball Association (NBA) in the US and it makes up about 6% of the global sports market. In the NBA there are 30 teams, all teams play 82 games in the regular season and the top teams make the post-season playoffs (a knockout style competition to decide the overall NBA winner).

Basketball is the only team sport that we consider in this thesis, which is played indoors at the professional level. Thus, weather related factors do not have an affect on the game. Basketball is very high scoring in comparison to the others (as highlighted in Table 1). It is also much more fluid and faster flowing in comparison to the other American sports which similarly to football makes quantifying an individuals impact on a game outcome

[6] Try - placing the ball down in a given zone at the end of the oppositions.
[7] A drop-goal is scored when a player kicks the ball from hand through the opposition's posts.
[8] Scrum - a method of restarting play that involves players packing closely together with their heads down and attempting to gain possession of the ball.
[9] https://www.ruck.co.uk/rugby-positions-roles-beginners/.

more challenging. In the NBA, there are on average 296 passes per team per game, this compares to 453 per team per game in the EPL, although in football there are more players on the pitch over a bigger playing area. If we look at this per player each basketball player makes 59 passes per game, whereas each football player completes on average 41 passes per game.

A.3.5 Cricket

Cricket is played in a number of forms (e.g., Test and Twenty20), in this thesis we focus on One Day International games (ODIs) due to existing literature also being focused on ODI games. In an ODI, there is 1 innings per team made up of 50 overs each (1 over = 6 balls), which can end earlier if all batsmen are out. In each innings, the batting team aims to score runs and bowling team aims to take wickets and prevent runs being scored. The winning team is the team with the most runs scored in their innings. Cricket is hugely popular in countries such as India, England and Australia. The Indian market in particular makes up the majority of the market and is reportedly worth $5.3 billion.

Hitting runs and taking wickets are the main metrics used to measure player performances. Cricket, like Baseball, relies on a number of individual performances by players which make up the team performance whereas other sports rely more on the team performance as a whole. Due to both cricket and baseball being bat-and-ball games rather than invasion games like the rest of the sports in this thesis, means that they present different challenges and factors for us to consider. At the core of this is that, even though they are team games, the performance of players is mainly based on a 1v1 scenario (batsman vs bowler). This means that when we evaluate or predict performance we can focus on how an individual batsman performs against an individual bowler or vice-versa.

A.3.6 Baseball

Baseball is a game made up of 9 innings, where an innings is made up of both teams batting (while the other team fields) until they receive three outs. The batting team aims to score runs (a batsman gets round all bases), the fielding team aims to strike batsman out (3 swing and misses) and stop runs being scored. If the score remains tied at the end of the regulated number of innings, then an extra innings is played. The team with the most runs at the end of 9 innings is the winning team. Baseball makes up 12%

of the global sports market. The performances of the Baseball teams/players are often measured by key statistics based on their abilities to hit runs or get outs.[10]

Baseball is mainly an American sport and the main league is Major League Baseball (MLB). In the MLB there are 30 teams where every team plays a 162 games in the regular season with the best teams making the playoffs. The playoffs is a knockout style competition formed of 12 teams, where each round is a "best out of 7 games", to decide the "World Series" winner. Teams play games much more frequently in a Baseball season than in other sports' which may mean players have to be rotated more and monitored closely for injury.

A.4 Chapter 3 List of Symbols

Symbol	Description
T	Team competing in game (α = home, β = away).
A	Actions that a team can make (tactical decisions e.g., formation).
$\theta \in \Theta$	Set of styles that a team can be (e.g., Tiki-Taka).
u	Payoff (e.g., probability of team winning or moving to a more positive state).
p	Prior belief of the actions and styles used by the opposition team.
$x \in X$	Set of possible game states in football these are scorelines.
$\sigma(x)$	Team strategy at state x.
π	Model to calculate the probability of transition into a new state/scoreline.
t	time played in a game.
$o \in \mathcal{O}$	Set of possible fluent objective a team can have.
$p(o) \in \mathcal{P}$	probability of achieving objective o.
$\mathcal{D}$	Distribution of possible season outcomes.
P	Prior knowledge parameter learned from games.
G	Individual game.
N	Number of games played in a season.
$w \in \mathcal{W}$	Weight of how effective a style/formation is against a team.

[10] Baseball was the first sport to really see the power of data. In the 1970s, Bill James began writing an annual "Baseball Abstract", containing statistics he collected by hand. This inspired the Oakland A's and Billy Beane (their General Manager) to change the way they operate by using data to make key decisions. This is documented in the book "Moneyball" by Micheal Lewis. There are many statistics collected in Baseball and the professional teams are much more advanced at using data in comparison to other sports.

Research Thesis: Declaration of Authorship

Print name: Ryan Beal

Title of thesis: AI in Team Sports

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as listed in Appendix Section A.1 on page 147.

Signature:

Date: