



Norwegian University of
Science and Technology

Evaluating Association Football Player Performances Using Markov Models

Haakon Sveiverud Haave
Håkon Høiland

Industrial Economics and Technology Management

Submission date: June 2017

Supervisor: Magnus Stålhane, IØT

Co-supervisor: Lars Magnus Hvattum, Molde University College

Norwegian University of Science and Technology

Department of Industrial Economics and Technology Management

Evaluating Association Football Player Performances Using Markov Models

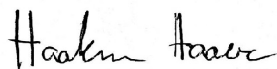
Association football player performance evaluation is subject to cognitive bias and subjective opinions. This may lead to undesired risk in the process of scouting and signing new players for clubs at the professional level. A likely result is inaccurate player valuations, and in turn financially expensive mistakes in the transfer market. This thesis seeks to develop data driven player performance evaluations, which objectively rates player performances in the Norwegian top division. The impact of each individual player involvement is estimated using Markov models. The performance evaluations are thought to be supportive in the process of scouting and evaluating players.

Preface

This Master's thesis is written for the Department of Industrial Economics and Technology Management at the Norwegian University of Science and Technology. It concludes the authors' Master of Science, specializing in Investment, Finance and Financial Management. Both authors specialized in Empirical and Quantitative Methods in Finance, and in Computer Science. The thesis was produced in the spring of 2017.

The presented work applies empirical and quantitative methods to association football. The thesis is a result of an initiative between Rosenborg Ballklub and the Department of Industrial Economics and Technology Management. Rosenborg Ballklub has stated that they want to explore the use of data driven tools and statistical analysis. This thesis seeks to bring such tools closer to the operations of the club, which can potentially yield a competitive advantage.

June 5, 2017



Haakon Haave



Håkon Høiland

Summary

Historically, association football player performance evaluation has been based on subjective opinions and intuition, which is subject to cognitive bias. This thesis develops data driven player ratings, argued to tend towards objectivity. Two Markov models are developed, and used to derive player ratings. Markov modelling of football requires discretization of a dynamic sport, but contextual features are chosen to minimize the loss of information.

In order to identify an accurate measure of player performance, different player ratings are proposed and validated. The validation consists of measuring the predictive power of the player ratings on match outcomes, the ratings' correlation with benchmark ratings and inter-season correlations. Based on the validation, one measure of player performance is chosen for further analysis.

Lists of the top performing players in different positions in Eliteserien are produced from the ratings. Furthermore, player performance profiles are constructed, which include players' measured performance across different types of involvements. In order to identify similar players, the distances between players' performance profiles are calculated, believed useful for scouting replacement players. The impact of team quality on player performance is identified and measured. This enables a player's estimated performance in a new club to be calculated. The combination of player ratings, identification of similar players and estimated player performances can potentially be used as a framework for scouting players in professional association football clubs.

Sammendrag

Evaluering av fotballspilleres prestasjoner har historisk sett vært basert på subjektive meninger og intuisjon. I denne masteroppgaven utvikles data-drevne spillerratinger som hevdes å tendere mot en objektiv evaluering av spillere. To Markovmodeller er utviklet og deretter brukt til å utlede spillerratinger. Markovmodellering av fotball krever diskretisering av en dynamisk sport, men kontekstuelle faktorer er valgt ut for å minimere tapet av informasjon.

Ulike spillerratinger er foreslått og validert for å identifisere et presist mål av spillerprestasjoner. Valideringsprosessen består av å måle spillerratingenes prediktive kraft på kamputfall, korrelasjon med referanseratinger og korrelasjon på tvers av sesonger. Ett spillerprestasjonsmål er valgt for videre analyse, basert på valideringen.

Ratingene brukes til å liste de beste spillerne i ulike posisjoner i Eliteserien. Videre konstrueres spillerprestasjonsprofiler som inneholder spilleres prestasjoner innenfor ulike typer involveringer. Avstanden mellom spilleres prestasjonsprofiler kalkuleres for å identifisere lignende spillere, noe som kan være nyttig når spillere skal erstattes. Lagnivåets påvirkning på spilleres prestasjoner er identifisert og målt. Dette muliggjør estimering av en spillers prestasjon i en ny klubb. Kombinasjonen av spillerratinger, identifisering av lignende spillere og estimering av spillerprestasjoner kan potensielt bli brukt som et rammeverk for å kartlegge spillere i profesjonelle fotballklubber.

Acknowledgments

First, we want to thank our supervisors, Magnus Stålhane and Lars Magnus Hvattum. They have been passionate about our work, and provided valuable feedback. Their guidance and commitment has been important for the progress and end result of this thesis.

We also want to thank Rosenborg Ballklub and Sports Director Stig Inge Bjørnebye for initiating this project. Special thanks to assistant coach Hugo Pereira for maintaining the collaboration, being passionate about the project and providing ideas. We are grateful for the opportunity to apply our field of study towards the sport we are so passionate about, and for the opportunity to collaborate with Rosenborg Ballklub. Finally, thanks to Opta Sports, Kenneth Wilsgård at Norsk Toppfotball and The Planning, Optimization and Decision Support Group at Molde University College for providing data.

Table of Contents

Preface	i
Summary	iii
Sammendrag	iv
Acknowledgments	v
Table of Contents	ix
List of Tables	xii
List of Figures	xiii
1 Introduction	1
1.1 Motivation	2
1.2 Research Questions	2
1.3 Report Structure	3
2 Basic Theory	5
2.1 Markov Chains	5
2.1.1 Definitions	5
2.1.2 State Valuation and Transition Probabilities	6
2.2 Markov Games	7
2.2.1 Definitions	7
2.2.2 State and Action Valuation	8
2.2.3 On Policy Learning and Value Iteration	8
2.3 Ordinal Logistic Regression: The Proportional Odds Model	10
2.4 k -fold Cross-Validation	11
2.5 Brier Score	11
2.6 k -Nearest Neighbours	12

3	Literature Review	15
3.1	Introduction to Sports Analytics	15
3.2	Data Driven Player Performance Evaluations	16
3.3	Markov Models in Football	17
3.4	Markov Models in Other Sports	18
4	Markov Models	21
4.1	Data	21
4.2	Model 1	22
4.2.1	State Definition	22
4.2.2	Constructing the Markov Chain	26
4.2.3	Value Iteration	26
4.3	Model 2	31
4.3.1	State Definition and Actions	31
4.3.2	Constructing the Markov Game	34
4.3.3	Value Iteration	35
4.4	Modelling Examples	37
5	Experimental Setup	41
5.1	Player Performance Evaluation	41
5.2	Validation Methods	46
5.2.1	k -fold Cross-Validation	46
5.2.2	Benchmark Correlation	48
5.2.3	Inter-Season Correlation	49
5.3	Identifying Similar Players	49
5.4	Team Impact on Player Performance	49
6	Evaluating Results	53
6.1	Validation	53
6.1.1	k -fold Cross-Validation	53
6.1.2	Benchmark Correlation	56
6.1.3	Inter-Season Correlation	57
6.1.4	Overall Validity Assessment	57
6.2	Player Ratings	58
6.3	Similar Players	61
6.4	Team Impact on Player Performance	65
6.5	Evaluation of Research Questions	68
7	Conclusion	71
8	Recommendations for Further Research	73
	Bibliography	74
	Appendix A Value Iteration Examples	81
	Appendix B k-fold Cross-Validation Regression Results	93

List of Tables

4.1	Model 1 context variables	23
4.2	Model 1 event types	24
4.3	Model 2 context variables	32
4.4	Model 2 actions	34
5.1	Player ratings example	43
5.2	Model 1 sequence example with impact values	45
5.3	Model 2 sequence example with impact values	46
5.4	Ordinal regression data structure	47
6.1	k -fold cross-validation coefficients	54
6.2	Brier scores	55
6.3	Match outcome distribution	55
6.4	Brier skill scores	55
6.5	2014 benchmark correlation	56
6.6	2015 benchmark correlation	56
6.7	2016 benchmark correlation	56
6.8	Inter-season correlations	57
6.9	Top 10 rated central defenders 2016	58
6.10	Top 10 rated full backs 2016	59
6.11	Top 10 rated central midfielders 2016	59
6.12	Top 10 rated wingers 2016	60
6.13	Top 10 rated strikers 2016	61
6.14	Explanation of position abbreviations	62
6.15	10 most similar players to RBK's central defenders	63
6.16	10 most similar players to RBK's full backs	63
6.17	10 most similar players to RBK's central midfielders	64
6.18	10 most similar players to RBK's forwards	65
6.19	Linear regression results	66
B.1	$I_{1,1}$ k -fold cross-validation regression results.	93

B.2	$I_{1,2}$ k -fold cross-validation regression results.	94
B.3	$I_{2,1}$ k -fold cross-validation regression results.	94
B.4	$I_{2,2}$ k -fold cross-validation regression results.	94
C.1	Position abbreviations.	95

List of Figures

4.1	Opta F24 data structure	22
4.2	Model 1 state definition	23
4.3	Zone illustration	24
4.4	Absorbing states	25
4.5	Model 1 value iteration example	28
4.6	Model 2 state definition	31
4.7	Tackle illustration of Model 1 compared to Model 2	32
4.8	Ball recovery illustration	33
4.9	Model 2 unsuccessful pass illustration	33
4.10	Model 2 value iteration example	36
4.11	Model 1 sequence modelling example	38
4.12	Model 2 sequence modelling example	39
5.1	Model 1 sequence example with state values	44
5.2	Model 2 sequence example with state and Q -values	45
6.1	Scatter plot of change in player performance against change in team rating	66
6.2	Scatter plot of player performance in successive seasons	67
A.1	Model 1 value iteration example	82
A.2	Model 2 value iteration example	88
C.1	Player positions in different formations.	96

Introduction

The prevalence of statistical analysis in sports has increased rapidly over recent decades. As abundant data has become available, and the computing power has improved, new possible applications for analytics in sports have emerged. The emerging field is known as sports analytics, and introduce the application of statistics, mathematics, operation research, economics and computer science to sports (Coleman, 2012). The purpose of sports analytics is to help decision makers make better assessments, hence better decisions. The objectives of sports clubs are twofold, the first relates to the sports performance on the field, and the second relates to management and financial issues. Although not mutually exclusive, these two areas represent the main areas to which sports analytics is applied (Maxcy and Drayer, 2014).

The competition on the professional level is tough in all sports. Association football, referred to as football in this thesis, is considered the biggest sport in the world, measured by media coverage, fan base and active players (Biggest Global Sports, 2017; Total Sportek, 2017). Rising ticket prices, high player salaries and gigantic transfer fees prove the sport's popularity and the enormous amounts of money circulating around the sport (The Football Forecast, 2016). In a highly competitive environment, where a small competitive advantage could lead to on-field as well as financial success, the importance of sports analytics is well established (Marr, 2015). All clubs in the English Premier League have data analysts with technical backgrounds in their staff. As of 2014, Manchester City employed 11 data analysts, while Brentford FC and the Danish club FC Midtjylland base their operations heavily on data and statistics (Lewis, 2014; Ingle, 2015).

In football, the on-field performance depends largely on how the individual players perform. Traditionally, the assessment of player performances have been qualitative and conducted by humans, which introduce cognitive bias. Simple statistics, such as number of goals scored, number of assists or key passes, and passing accuracy have been used to support the qualitative player performance evaluations. As sports analytics has emerged, it is now possible to evaluate player performances using comprehensive, quantitative models. Such models allow decision makers in sports clubs to monitor and assess the performances of own players, but the models also play a substantial role in scouting and player recruitment (Radicchi and Mozzachiodi, 2016).

Sports analytics has aroused academic interest. While the practice of sports analytics within sports clubs is kept unpublished and undisclosed, an increasing amount of academic work is being published. Originating from the statistics in newspapers and on the back of trading cards, sports analytics has grown as data and computing power have become available, first moving into the decision-making process in sports clubs, and now into universities and academic circles (Cochran, 2010).

1.1 Motivation

The Norwegian football club, Rosenborg Ballklub (RBK), wants to explore the sports analytics field. RBK is the most successful Norwegian club, having won the national top division, currently named Eliteserien, 24 times, and having qualified 11 times for the group stage of Europe's most prestigious football competition, the UEFA Champions League. The club has initiated cooperation with the Institute of Industrial Economics and Technology Management at the Norwegian University of Science and Technology to pursue data driven decision support. This master's thesis, and the work on which the thesis is based, results from this initiative. Few studies of Norwegian football and Eliteserien exist, and the integration of analytics in Norwegian clubs appear minimal. This thesis adds to the limited academic work on sports analytics in Norway. The integration of analytics in football clubs can reveal valuable competitive advantages, and an increasing amount of academic work on sport analytics might increase the focus on analytics in Norwegian clubs.

Player evaluation is an area where intricate analysis can provide a competitive advantage. Data driven player evaluation eliminates cognitive bias from the evaluation process. Decision makers in football clubs may use data driven performance measures to monitor and assess the performance of own players, and as a tool when scouting new players. Subjective player evaluations may lead to undesired risk in the process of signing new players in professional football clubs. Therefore, fair and precise player evaluations may help avoid expensive mistakes in the transfer market.

Finding suitable players for a specific role or position require a large and proficient scouting network. Data driven player evaluations may facilitate more effective disposal of the scouting resources in professional clubs. In the presence of valid data driven player performance evaluations, an initial screening of players can be done for several leagues without actually watching the players. In turn, a shortlist of players can be identified, and time and money can be saved in the process of further scouting the potential players.

1.2 Research Questions

This thesis considers data driven player performance evaluations in football. As such, a data driven player rating must be developed. As the roles of goalkeepers and outfield players are too different to be studied on the same premises, only outfield players are considered. Furthermore, the possible applications and limitations of data driven ratings need to be explored to determine the relevance for professional football clubs. To formalize the aims of the thesis, three research questions are presented below. The work in this thesis

seek to answer these questions, and the questions will be evaluated at the end of Chapter 6, based on the results.

RQ1: Can outfield association football player performances be objectively and accurately evaluated by data driven player ratings, derived from Markov models?

RQ2: Can data driven player evaluations be used to identify similar players?

RQ3: Is there a relationship between team quality and individual player performances?

1.3 Report Structure

Chapter 1 serves as an introduction, and presents the motivation and the research questions which lay the foundation of the thesis.

Chapter 2 presents the underlying theory of the statistical and technical methods applied in the thesis.

Chapter 3 gives an overview of the relevant literature, and an introduction to sports analytics.

Chapter 4 describes the definition and development of two different Markov models, used to evaluate player involvements later in the thesis, and describes the data used to build the models and evaluate players.

Chapter 5 describes the procedure for rating players, the validation tools, and the approaches for finding similar players and investigating team quality's impact on player performance.

Chapter 6 presents the results of the validation process and the highest rated players in each position. An analysis of identified similar players and a discussion of how team quality impacts player performance follow.

Chapter 7 gives an overview of the important findings. Conclusions are presented, based on the presented results in light of the motivation behind the thesis.

Chapter 8 presents ideas for extending the presented work in future research.

Basic Theory

The basic theory behind the methods applied in this thesis is presented in this chapter. Markov chains are introduced first, followed by the extension named Markov games. The theory behind ordinal logistic regression is presented next, followed by a description of k -fold cross-validation, the Brier score and the k -nearest neighbours methodology.

2.1 Markov Chains

A Markov chain is a stochastic process which fulfills the Markovian property. A Markov chain model consists of a defined state space where each state occurs with a certain probability. The Markovian property says that the next state, given the current state, only depends on the current state (Hillier and Liebermann, 2015). Observable systems, such as a football game, can be modelled as a Markov chain by defining a state space and observing the occurrences of states and state transitions. The Markov chain reward model is an extension of the Markov chain model, which is used when a reward is associated with certain states. It is natural to use the Markov chain reward model approach when modeling football, where rewards can be used to model goals. Formal definitions of Markov chains and Markov reward models are presented next.

2.1.1 Definitions

Adapting the notation from Littman (1994), a Markov reward model is defined by the tuple $\langle \mathcal{S}, T, R, \beta \rangle$, where

- $\mathcal{S} = \{s_1, s_2, \dots, s_m\}$ is a finite set of states, where each state describes the current state of the stochastic process.
- $T : \mathcal{S} \rightarrow \Pi(\mathcal{S})$ is the transition function which defines the probability $T(s, s')$ of ending in state $s' \in \mathcal{S}$ given the initial state $s \in \mathcal{S}$.
- $R : \mathcal{S} \rightarrow \mathbb{R}$ is the reward function, which gives the immediate reward, $R(s)$, given the state s .

- $0 \leq \beta < 1$ is an optional discount factor which can be used to make short term rewards more valuable than long term rewards.

Algorithm 1 Markov Chain Value Iteration

Require: Maximum iterations M , convergence criterion c , Markov game model

```

lastValue ← 0
currentValue ← 0
converged ← false
for  $i = 1; i \leq M; i++$  do
  if  $converged = false$  then
    for all states  $s$  in Markov game model do
       $V(s)_{temp} \leftarrow \beta \frac{1}{Occ(s)} \sum_{s' \in \mathcal{S}} Occ(s, s')(R(s') + V(s'))$ 
       $currentValue \leftarrow currentValue + |V(s)_{temp}|$ 
    end for
    for all states  $s$  in Markov game model do
       $V(s) \leftarrow V(s)_{temp}$ 
    end for
    if  $\frac{currentValue - lastValue}{currentValue} < c$  then
       $converged \leftarrow true$ 
    end if
  end if
   $lastValue \leftarrow currentValue$ 
   $currentValue \leftarrow 0$ 
end for

```

2.1.2 State Valuation and Transition Probabilities

A state s is valued based on the immediate reward and value of the possible successive states. Formally, the value of a state s is given by

$$V(s) = \beta \sum_{s' \in \mathcal{S}} (T(s, s')(R(s') + V(s'))) \quad (2.1)$$

where s' is the state which directly follows s . The state and transition probabilities from an observed system can be estimated from their actual occurrences. Following the notation from Routley (2015), where

- $Occ(s)$ is the number of occurrences of state s .
- $Occ(s, s')$ is the number of times state s lead directly to state s' .

Then

$$T(s, s') = \frac{Occ(s, s')}{Occ(s)} \quad (2.2)$$

Equation 2.1 can then be written as

$$V(s) = \beta \sum_{s' \in \mathcal{S}} \frac{Occ(s, s')}{Occ(s)} (R(s') + V(s')) = \beta \frac{1}{Occ(s)} \sum_{s' \in \mathcal{S}} Occ(s, s') (R(s') + V(s')) \quad (2.3)$$

Following (Hillier and Liebermann, 2015), an absorbing state is a state which the process will never leave once it has been entered, meaning $T(s, s') = 0$ for all $s' \in \mathcal{S}$ when s is absorbing. When absorbing states exist, and the transition probabilities are known from the occurrences, each $V(s)$ can be learnt iteratively through dynamic programming. The dynamic programming algorithm iterates every state and state-transition repeatedly until the state values converge. The pseudocode of the algorithm is shown in Algorithm 1.

2.2 Markov Games

The Markov game concept originates from game theory, and is an extension of the Markov chain reward model described in Section 2.1. A Markov game model aims to model one or more agents acting in a defined environment. The environment is described by a state space, and each decision maker may perform an action from a defined action set. The agents' actions initiate a transition which brings the agents to the next state. Each agent has a reward associated with each transition, and the goal for each agent is to maximize its future reward. The zero-sum Markov game is a special case of a Markov game, where two agents interact in an environment with diametrically opposite goals. This resembles a football game, where one goal up for one team means one goal down for the other team. Another special case of a Markov game is the Markov Decision Process (MDP). In the case of an MDP, there is only one active agent, meaning a stationary environment is assumed. A formal description of Markov games and MDPs is presented next.

2.2.1 Definitions

Using the notation from Littman (1994), an n -player Markov game is defined by the tuple $\langle \mathcal{S}, \mathcal{A}_1, \dots, \mathcal{A}_n, T, R_1, \dots, R_n, \beta \rangle$, where

- $\mathcal{S} = \{s_1, s_2, \dots, s_m\}$ is a finite set of states which makes up the environment.
- $\mathcal{A}_i = \{a_{1_i}, \dots, a_{k_i}\}$ is a finite set of actions available to agent i where k_i is the number of actions available to agent i .
- $T : \mathcal{S} \times \mathcal{A}_1 \times \mathcal{A}_2 \times \dots \times \mathcal{A}_n \rightarrow \Pi(\mathcal{S})$ is the transition function which defines the probability $T(s, a_1, a_2, \dots, a_n, s')$ of ending in state $s' \in \mathcal{S}$ given the initial state $s \in \mathcal{S}$ and actions $a_1 \in \mathcal{A}_1, \dots, a_n \in \mathcal{A}_n$ by the agents.
- $R_i : \mathcal{S} \times \mathcal{A}_1 \times \mathcal{A}_2 \times \dots \times \mathcal{A}_n \rightarrow \mathbb{R}$ is the reward function, which gives the expected immediate reward of player i , $R_i(s, a_1, a_2, \dots, a_n)$, given the initial state $s \in \mathcal{S}$ and actions $a_1 \in \mathcal{A}_1, \dots, a_n \in \mathcal{A}_n$.
- $0 \leq \beta < 1$ is an optional discount factor which can be used to make short term rewards more valuable than long term rewards.

A two-player zero-sum Markov game simplifies the tuple to $\langle \mathcal{S}, \mathcal{A}, \mathcal{O}, T, R, \beta \rangle$, where \mathcal{A} is the set of actions available to player one and \mathcal{O} is the set of actions available to player two, referred to as the opponent. Only one reward function, R , is needed, because the players have diametrically opposite goals. One player seeks to maximize the reward, while the other player seeks to minimize the reward. An MDP can be simplified further to $\langle \mathcal{S}, \mathcal{A}, T, R, \beta \rangle$, as there is only one active decision making agent (Littman, 2001). The remainder of Section 2.2 focus on two-player zero-sum Markov games MDPs, as that is most relevant for this thesis.

2.2.2 State and Action Valuation

Following Littman (2001), a policy π describes the behaviour of an agent. $\pi : \mathcal{S} \rightarrow \Pi(\mathcal{A})$ defines the probability distribution over actions in \mathcal{A} for an agent, given a state in \mathcal{S} . Specifically, $\pi(s, a)$ is the probability of choosing action a in state s . $Q_i^\pi(s, a, o)$, referred to as a Q -value, denotes the expected future reward for agent i when the agents choose actions $a \in \mathcal{A}$ and $o \in \mathcal{O}$ in state $s \in \mathcal{S}$, and the game continues according to policy π . $V_i^\pi(s)$, referred to as a state value, denotes the expected future reward for player i in state s , when the game continues according to policy π . When the game is zero-sum, the values are simply $Q^\pi(s, a, o)$ and $V^\pi(s)$, where higher values are desirable for the maximizing agent and lower values are desirable for the minimizing agent.

The Q and state values regardless of policy are defined as $Q(s, a, o)$ and $V(s)$, respectively, and correspond to the average values over all policies. In an MDP there is only one active action maker, which makes $Q(s, a)$ the expected value of the state-action pair consisting of the performed action a in state s . Following Littman (2001), the Q -values are found by solving

$$Q(s, a, o) = R(s, a, o) + \beta \sum_{s' \in \mathcal{S}} T(s, a, o, s') \sum_{a' \in \mathcal{A} o' \in \mathcal{O}} \pi(s', a') \pi(s', o') Q(s', a', o') \quad (2.4)$$

In the case of an MDP, the equation is simplified to

$$Q(s, a) = R(s, a) + \beta \sum_{s' \in \mathcal{S}} T(s, a, s') \sum_{a' \in \mathcal{A}} \pi(s', a') Q(s', a') \quad (2.5)$$

The expected value of a state, $V(s)$, is a function of the value of each possible action in the state given by

$$V(s) = \sum_{a \in \mathcal{A}} \pi(s, a) Q(s, a) \quad (2.6)$$

for an MDP. Equation 2.5 can then be written as

$$Q(s, a) = R(s, a) + \beta \sum_{s' \in \mathcal{S}} T(s, a, s') V(s') \quad (2.7)$$

2.2.3 On Policy Learning and Value Iteration

Markov games are commonly used to learn optimal policies, hence learning the optimal action in each state and maximizing $Q(s, a)$. An alternative approach is to watch the

policies carried out by the players, and the corresponding rewards, in order to compute the expected Q and state values. This approach is called on policy learning (Poole and Mackworth, 2010), and has been used by Bjertnes et al. (2016) and Routley (2015) in similar theses.

Following the notation from Routley (2015)

- $Occ(s)$ is the number of occurrences of state s .
- $Occ(s, a)$ is the number of occurrences of action a in state s .
- $Occ(s, a, s')$ is the number of times action a in state s lead directly to state s' .

Then

$$T(s, a, s') = \frac{Occ(s, a, s')}{Occ(s, a)} \quad (2.8)$$

and

$$\pi(s, a) = \frac{Occ(s, a)}{Occ(s)} \quad (2.9)$$

The following equation is obtained by substituting Equation 2.8 into Equation 2.7

$$\begin{aligned} Q(s, a) &= R(s, a) + \beta \sum_{s' \in \mathcal{S}} \frac{Occ(s, a, s')}{Occ(s, a)} V(s') \\ &= R(s, a) + \beta \frac{1}{Occ(s, a)} \sum_{s' \in \mathcal{S}} Occ(s, a, s') V(s') \end{aligned} \quad (2.10)$$

Let $R(s)$ be the actual immediate reward of state s , then

$$Q(s, a) = \beta \frac{1}{Occ(s, a)} \sum_{s' \in \mathcal{S}} Occ(s, a, s') (R(s') + V(s')) \quad (2.11)$$

and by substituting Equation 2.9 into Equation 2.6, the following is obtained

$$V(s) = \frac{1}{Occ(s)} \sum_{a \in \mathcal{A}} Occ(s, a) Q(s, a) \quad (2.12)$$

Algorithm 2 Markov Game Value Iteration

Require: Maximum iterations M , convergence criterion c , Markov game model

```

lastValue  $\leftarrow 0$ 
currentValue  $\leftarrow 0$ 
converged  $\leftarrow false$ 
for  $i = 1; i \leq M; i++$  do
  if converged = false then
    for all state-action pairs  $(s, a)$  in Markov game model do
       $Q(s, a) \leftarrow \beta \frac{1}{Occ(s, a)} \sum_{s' \in \mathcal{S}} Occ(s, a, s') (R(s') + V(s'))$ 
      currentValue  $\leftarrow$  currentValue +  $|Q_i(s, a)|$ 
    end for
    for all states  $s$  in Markov game model do
       $V(s) \leftarrow \frac{1}{Occ(s)} \sum_{a \in \mathcal{A}} Occ(s, a) Q(s, a)$ 
    end for
    if  $\frac{currentValue - lastValue}{currentValue} < c$  then
      converged  $\leftarrow true$ 
    end if
  end if
  lastValue  $\leftarrow$  currentValue
  currentValue  $\leftarrow 0$ 
end for

```

The concept of learning the state and Q -values on policy is known as reinforcement learning. Value iteration is a reinforcement learning algorithm which iterates through all states and actions in the Markov game repeatedly. State and Q -values are computed using Equations 2.10 and 2.6 until the values converge. Value iteration was used by Bjertnes et al. (2016) and Routley (2015), reviewed in Chapter 3. The pseudocode of the value iteration is shown in Algorithm 2.

2.3 Ordinal Logistic Regression: The Proportional Odds Model

Ordinal logistic regression is applicable when the dependent variable is discrete and ordered (Benoit, 2012). Following Benoit (2012), let the data set consist of n observations on the form $\langle Y_i, X_{1i}, X_{2i}, \dots, X_{ki} \rangle$ for observations $i = 1, 2, \dots, n$. The dependent variable, Y , takes one of C ordered values $j = 1, 2, \dots, C$ with probability $\pi^{(j)} = P(Y = j)$. X_1, \dots, X_k denotes the k explanatory variables. The set of observed Y_i values are assumed statistically independent of each other.

Let

$$\gamma^{(j)} = P(Y \leq j) = \pi^{(1)} + \pi^{(2)} + \dots + \pi^{(j)} \quad \text{for } j = 1, 2, \dots, C - 1 \quad (2.13)$$

be the $C - 1$ cumulative probabilities. $\gamma^{(C)} = P(Y \leq C) = 1$, thus it does not need to be

modelled. The following holds for each observation i in each category $j = 1, 2, \dots, C - 1$

$$\log \left(\frac{\gamma_i^{(j)}}{1 - \gamma_i^{(j)}} \right) = \log \left(\frac{P(Y_i \leq j)}{P(Y_i > j)} \right) = \alpha^{(j)} - \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} \quad (2.14)$$

The $C - 1$ intercept terms are denoted $\alpha^{(1)}, \dots, \alpha^{(C-1)}$ and the k variable coefficients are denoted β_1, \dots, β_k . The intercept terms and variable coefficients are estimated using maximum likelihood. The proportional odds model assumes that the relationship between each pair of outcomes is the same. This means that only one coefficient, β , is computed for each X , and that the $\gamma^{(1)}, \gamma^{(2)}, \dots, \gamma^{(j)}$ curves are parallel. Equation 2.14 can be rewritten to get the following expression for the cumulative probabilities

$$\gamma^{(j)} = P(Y \leq j) = \frac{\exp(\alpha^{(j)} - (\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k))}{1 + \exp(\alpha^{(j)} - (\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k))} \quad (2.15)$$

Each intercept term must be $\alpha^{(1)} < \alpha^{(2)} < \dots < \alpha^{(j)}$, to guarantee that $\gamma^{(1)} < \gamma^{(2)} < \dots < \gamma^{(j)}$. The probabilities of each category is

$$\begin{aligned} P(Y = 1) &= \gamma^{(1)} && \text{for } j = 1 \\ P(Y = J) &= \gamma^{(j)} - \gamma^{(j-1)} && \text{for } j = 2, \dots, C - 1 \\ P(Y = C) &= 1 - \gamma^{(C-1)} && \text{for } j = C \end{aligned} \quad (2.16)$$

2.4 k -fold Cross-Validation

Cross-validation is a technique used to measure the predictive performance of a statistical model. The model can be trained and fitted to a data set, but may not be able to predict new independent data on which the model is not trained. This is called overfitting, and a trade-off between a good fit and good predictive power is introduced. The model should fit the data, but overfitting the model will reduce its predictive power. Cross-validation aims to limit the problems of overfitting by leaving out a portion of the data when fitting the model, and testing the model against this portion.

k -fold cross-validation is one variation of cross-validation where the data is randomly partitioned into k equally sized subsamples, referred to as folds. In each iteration, one of the k folds are used as test data while the remaining $k - 1$ folds are used as training data. A statistical model, e.g. a regression model, is built on the training data, and then tested on the test data to measure the predictive power of the model. The process is repeated k times until all k folds are used as test data exactly once. k is an unfixed parameter, but 10-fold cross-validation is commonly used (Geoffrey et al., 2004).

2.5 Brier Score

The Brier score measures the accuracy of probabilistic predictions. The set of possible outcomes can either be binary or categorical, and the assigned probabilities must sum to

one. Following Wilks (2011), the Brier score is essentially the mean squared error of the predictions, where observation $o^{(j)} = 1$ if event j occurred, and $o^{(j)} = 0$ if event j did not occur. Letting $\gamma^{(j)}$ denote the probability that event j occurs, the Brier score is given by

$$BS = \frac{1}{C} \sum_{j=1}^C (\gamma^{(j)} - o^{(j)})^2 \quad (2.17)$$

where C is the number of possible outcomes. The Brier score takes values in the range $0 \leq BS \leq 1$. A perfect prediction receives $BS = 0$, and less accurate predictions receive higher Brier scores.

The Brier skill score is often computed to evaluate a Brier score. The Brier skill score is given by

$$BSS = 1 - \frac{BS}{BS_{ref}} \quad (2.18)$$

where BS_{ref} is the Brier score of a reference prediction, often given by the long term probabilities or a benchmark prediction. The Brier skill score describes the relative skill of a prediction over that of the reference prediction. The Brier skill score takes values in the range $-\infty < BSS \leq 1$, where negative values mean that the prediction is less accurate than the reference, $BSS = 0$ means no skill compared to the reference, and a positive value means a better prediction than the reference prediction.

2.6 k -Nearest Neighbours

The k -nearest neighbours is an instance-based learning algorithm used for classification, meaning no model is built and the instances represent all the knowledge. The algorithm assumes that all instances correspond to points in the p -dimensional space \mathbb{R}^p , and is commonly based on the Euclidean distances between a test instance and the instances in a specified training sample (Mitchell, 1997). Let \mathbf{X} be an input sample consisting of n instances with p features.

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, \mathbf{x}_i \in \mathbb{R}^p \quad (2.19)$$

The Euclidean distance between a test instance, \mathbf{x}_t , and an instance in the training sample, \mathbf{x}_s , is then given by

$$d(\mathbf{x}_t, \mathbf{x}_s) = \sqrt{\sum_{j=1}^p (x_{tj} - x_{sj})^2} \quad (2.20)$$

Each instance in the training sample belongs to a class $\omega = \omega_1, \omega_2, \dots, \omega_m$, where m is the number of different classes. Following the k -nearest neighbour rule, the k nearest neighbours of a test instance are found by calculating the Euclidean distance to all

instances in the training sample. By majority voting, the test instance is assigned to the class to which the majority of the k nearest neighbours belong to (Peterson, 2009). In the special case where number of instances equals the number of different classes, $n = m$, the algorithm identifies the k instances which are most similar to the test instance. If the features have different scales, the values have to be normalized to ensure that the distance measure is not dominated by features with a large scale.

Literature Review

The use of data analysis, aimed at decision support in sports, has increased rapidly in recent years, and is commonly referred to as sports analytics. As the field is considered premature, and might be unfamiliar, a general introduction to sports analytics is presented in the first section of this chapter. The remaining sections concern literature directly relevant to this thesis. To the authors knowledge, no research is published on identifying similar players nor the impact of team quality on individual performances. As such, the reviewed literature concerns data driven player performance evaluation and applications of Markov models in sports. Following the introduction to sports analytics, previous works on player performance evaluation in football is presented. Studies that apply Markov models to football is presented next, followed by a review of Markov models in other sports.

3.1 Introduction to Sports Analytics

The purpose of sports analytics is to unveil competitive advantages through the use of data analysis. Today, most professional sports teams have their own analytics department (Steinberg, 2015). The increased attention towards sports analytics was sparked by the book *Moneyball*, written by Michael Lewis, released in 2003. *Moneyball* is the story of the baseball team Oakland Athletics, who successfully built a competitive player squad, based on statistics, despite spending noticeably less on player salaries than their competitors. Coleman (2012) reviews 140 journals, and states that sports analytics research has been published for over 50 years. It is found that the number of sports analytics publications, in proportion to other research areas, has increased rapidly during the last 20 years. Furthermore, it is argued that the field is maturing, based on refereed journal articles.

Data is readily available from the professional leagues themselves and third-party providers. This enables sophisticated, statistical analysis. Analytics can be applied in many domains, including game and player performance, player selection, customer relationships, business management and injury prevention (Davenport, 2014). Baseball is considered an early adapter of analytics, and the baseball analytics organization Society for American Baseball Research was founded in 1972 (Neese, 2015). The nature of base-

ball facilitates the identification and analysis of key performance indicators (KPIs). The flow of the game can be simplified to discrete events with low loss of information, compared to most major team sports. Most baseball situations can be modelled as one versus one situations, whereas several players often interact simultaneously in other team sports. In addition to baseball, several other sports have been studied by academics, including ice hockey, basketball, football and others.

Benjamim Reep is dubbed the father of football analytics by Anderson and Sally (2014), and his work on scoring probability and directness of play introduced the statistical approach to football. Football analytics has increased along with analytics in other sports, and several aspects of the sport have been studied. Physical aspects such as energy demand (Mohr et al., 2011), injury prevention (Caraffa et al., 1996) and physical output (Nørregaard et al., 1991) has received attention from academics. A much discussed metric is expected goal, which is a measure of the probability of a shot ending in goal, given input parameters such as distance, angle and others. One such study is Bjertnes et al. (2016), where 10 significant shot outcome variables are identified. Several studies look at KPIs, and the number of shots, shooting accuracy, possession percentage and passing accuracy seem to be important determinants of match outcome (Bjertnes et al., 2015; Oberstone, 2009; Lago-Peñas et al., 2010).

3.2 Data Driven Player Performance Evaluations

The evaluation of player performances has traditionally been carried out by humans, through the staff in sports clubs, experts or journalists in the media. The rise of sports analytics has introduced data driven sports rating systems, which are supposed to remove cognitive bias and tend towards objectivity.

McHale et al. (2012) describe the development of The EA Sports Player Performance Index. The index evaluates football player performances, and was the official player rating system in the English Premier League, the English Championship and the Scottish Premier League for several seasons. The rating system is based on six separately constructed and calculated subindices, each capturing different aspects of the game. The captured aspects are match contribution, winning performance, match appearance, goals scored, assists and clean sheets. A player's match contribution subindex is a measure of how many points the player won for his team during the game. A Poisson model is used to estimate the number of goals a team will score, given their number of shots, their shot accuracy, the opponent's blocking capability and the opponent's goalkeeping capability. A linear regression on how much each player's actions improve the team's likelihood of a shot is then used to estimate the player's match contribution in terms of points. Data from seasons 2002-2003 and 2003-2004 was used in the development of the index. The data was event-based and followed the actions on the ball. The final index is a weighted sum of the six subindices, and was the first index to rate individual players in a team sport using a single common score, regardless of playing position (McHale et al., 2012).

Brooks et al. (2016) describe a player rating system based entirely on the value of completed passes. These values are derived by looking at the relationship between pass locations and generated shot opportunities. Event-based data from the 2012-2013 La Liga season is used. A supervised machine learning model is used to learn the relative proba-

bilities of generating a shot later in the possession. By mapping these values to the player who performed the passes, the authors rank all players with more than 200 passes. The authors claim that the derived rankings are largely consistent with general perceptions. The ranking receive high correlations with offensive score metrics such as number of goals scored and number of assists.

Sæbø (2015) and Sæbø and Hvattum (2015) evaluate match contributions of football players by applying an adjusted plus-minus model. The plus-minus statistic of a player is given by the number of goals his team scores minus the number of goals his team concedes, while the player is on the pitch. Regression-based player ratings are derived, based on the plus-minus statistic. The player ratings are used as input in a regression model, along with other variables, for analyzing transfer fees and evaluating the efficiency of the transfer market. To measure the predictive power of the plus-minus statistic, Sæbø (2015) uses the statistic to forecast the outcome of football matches. All matches in the 2013-2014 Premier League season are simulated, which results in a simulated league table. The derived ratings manage to identify top players in European football and a player's rating is identified as important when determining the transfer fee. Bjertnes et al. (2016) develop player ratings based on the estimated value of each players' individual involvements. As this work uses Markov models to estimate involvement values, it is discussed in the next section.

3.3 Markov Models in Football

Bjertnes et al. (2016) use Markov reward models to model football, based on data from Eliteserien. Two different Markov reward models, referred to as Model 1 and Model 2 in this section, are built and validated. States and state transitions are constructed from event based data from the 2014 and 2015 seasons. Possible actions are defined, such as pass, shot and tackle, which lead to state transitions. By counting the occurrences of each state and each state transition, the probability of ending in a certain state, given the current state, is computed. Goals are defined as absorbing states, yielding a reward of 1 and -1 for the home and away team, respectively. As such, the home team seek to maximize the reward, while the away team seek to minimize the reward. The expected values of states, $V(s)$, and state-action pairs, $Q(s, a)$, are estimated using value iteration.

Bjertnes et al. (2016) define the state in Model 1 as which team had the ball, which zone the ball was in, the current score, and the manpower difference. Model 2 adds event type to the state definition, as well as a binary outcome variable and the zone in which the ball ended. This means that Model 2's state space is larger. However, Model 1 has more complex state transitions as actions are incorporated in the transitions. Model 1 resembles a Markov game, while Model 2 resembles a Markov chain. The expected values of states and state-action pairs are used to rate individual players. This is done by mapping every action of each player to the corresponding state-action pair in Model 1 and the corresponding state in Model 2. The ratings are normalized per 90 minutes in order to compare players. In Model 1, the involved player is rewarded the value $Q(s', a') - Q(s, a)$ for each performed involvement, where $Q(s', a')$ is the expected value of the state-action pair which directly follows the player's involvement. The argument for looking at the next state-action pair is that this captures how a player involvement actually influenced the game. However, a drawback of this method is that the player performing the next action

may choose an action with a low $Q(s', a')$ value. In this case a player can be punished for the next player's poor decision. Model 2 rewards the players the value $V(s)$ for each performed involvement. This means that the players are not punished for the next player's action, but it also means that a player may be rewarded too heavily in certain cases.

Bjertnes et al. (2016) compare the player ratings from Model 1 and Model 2 using inter-season correlations. The correlation between seasons in Model 2 is 0.90, meaning that the players are rated similarly in one year compared to another. The correlation is 0.19 in Model 1. It is argued that player performances are expected to be consistent between seasons, meaning Model 2 evaluates players more reliably than Model 1. The models' evaluations are also compared to player ratings by journalists. The journalist ratings are susceptible to cognitive bias, but it is argued that valid player ratings should resemble the media ratings to some extent. Both models yield correlation coefficients of approximately 0.5 with respect to the media ratings. A drawback, admitted by the authors, is that defenders appear to receive lower ratings than offensive players. It seems like the models are struggling to reward tackles and interceptions, typical actions for defenders. Goals are heavily rewarded in both models, which means the strikers score the highest values.

Hirotsu and Wright (2002) model football as a Markov process with four states. These states are team A scores, team A is in possession of the ball, team B scores and team B is in possession of the ball. The probability distribution of the final score from any position in a match is estimated, and used to estimate the optimal time to change tactics and make substitutions. The work is continued in Hirotsu and Wright (2003a) and Hirotsu and Wright (2003b), where the same type of model is used to evaluate and visualize offensive and defensive strengths of teams in the English Premier League. Dynamic programming is used to derive the optimal substitution strategy and to determine how many of each type of player should start a match.

3.4 Markov Models in Other Sports

Routley (2015) models ice hockey as a Markov game, and the work resembles that of Bjertnes et al. (2016). The state context is defined by goal differential, manpower difference, game period and the action history. The addition of action history creates a more complex state space, compared to Bjertnes et al. (2016). A larger state space implies that more data is needed to achieve reliable results. The model is built on seven years of data, which is significantly more than Bjertnes et al. (2016), meaning a more complex state space might be viable. The complexity is mitigated by a small set of actions, and ending sequences frequently. The average sequence consists of 4.87 states, which limits the length of the action histories in the states. Routley (2015) use the model to derive player ratings. The players are rewarded the value $Q(s, a) - V(s)$ for each of their involvements, where $Q(s, a)$ is the expected value of performing action a in state s , and $V(s)$ is the expected value of the state in which the action was performed. The model yields player ratings with a correlation of 0.71 between seasons, argued to support the validity.

Several studies have used Markov chains to model basketball. A two-state model is presented in Kvam and Sokol (2006), used to predict the winner of college basketball matches. The team's previous performances were used to estimate the transition probabilities. The resulting predictions are described as better than standard and subjective ranking

systems. Shirley (2007) builds a more extensive Markov chain model, defining a state by the team in possession, how the team won possession and how many points were scored in the last possession. All possible ways of winning position is captured by the five incorporated values, namely inbound pass, steal, offensive rebound, defensive rebound and free throw. 0, 1, 2 or 3 points were scored in the last possession and one of two teams has possession. Consequently, $5 \times 4 \times 2 = 40$ states are defined. However, if possession is won by a steal, no points were scored in the last position, meaning only 30 states are actually possible. Data from professional basketball games is used to estimate the the expected number of points scored in each state. In order to estimate win probabilities, 1000 games were simulated for each team. The author questions some of the results, argued to relate to the the limited data set of 4.5 games. Nevertheless, the predicted win probabilities are described as close to the actual win percentages. Strumbelj and Vracar (2012) use the same state definition as Shirley (2007), but build the model from one whole season of play-by-play data. Transition probabilities for each team are observed in the data. Monte Carlo simulation is then used to predict game outcomes, based on the involved teams' transition probabilities. The quality of the predictions are described as inferior to implied betting probabilities, but comparable to other statistical methods.

Batting order is a widely discussed topic in baseball, and choosing the right batting order is considered an important aspect of baseball strategy. Bukiet et al. (1997) use a Markov chain to find optimal batting orders, run distributions and expected wins per team. 25 Markov states are defined, describing the number of runners on the bases and the number of missed batting attempts. Statistical information about the batters are used to compute the set of transition probabilities for each batter. Using the batting order and the transition probabilities, the number of expected runs for a team can be computed. In a specific game, the distribution of expected runs for the two teams is used to produce winning probabilities for each team. The process is repeated for many games, resulting in expected wins per team. Asaro (2016) use a similar approach to find optimal batting orders and expected wins, but extend the work by estimating the effect of trading players.

Markov Models

Two different Markov models are presented in this chapter. Markov models have previously been used to model team sports, including football, as described in Section 3.3 and 3.3. The presented models are inspired by Bjertnes et al. (2016), but several improvements are proposed. Football games are modelled as a Markov chain in the first model, referred to as Model 1, and modelled as a Markov game in the second model, referred to as Model 2. The Markov modelling approach implies that football games are divided into distinct states, depending on the current environment of the game. Despite football being a dynamic game, the discretization required by a Markovian approach has shown promising results (Bjertnes et al., 2016). The values of states and state-action pairs can be deduced from the Markov models, learned through value iteration. With a large set of data, i.e. data from many football games, the learned values are believed to closely resemble the actual values of the modelled situations. Data from Opta Sports is used to build the models. To make it easier to understand how the models are built, the data is introduced first. Model 1 is then presented, followed by Model 2.

4.1 Data

The Markov models are built on data collected by Opta Sports, which is a credible, international sports data provider. The data set includes the Opta feeds for every match in Eliteserien, seasons 2014, 2015 and 2016. The structure of the data follows Opta's F24 data format. Each game has its separate XML file with a list of events in chronological order. Most of the events are related to actions on the ball, and each file has approximately 1600 events. The events describe which player acts on the ball and the event type. Some events are unrelated to the ball, such as substitutions, bookings, and end of a period. Each event includes information about where the event occurred, whether it was successful or not, and the ball's end position. Figure 4.1 illustrates the structure of the data, where the red box highlights a single event in the data. The data also includes separate data files with squad lists and information about the players in Eliteserien, e.g. age, nationality and playing position.

```

<Games timestamp=" " >
  <Game id=" " away_team_id=" " away_team_name=" " competition_id=" " competition_name=" "
game_date=" " home_team_id=" " home_team_name=" " matchday=" " period_1_start=" "
period_2_start=" " season_id=" " season_name=" ">
  <Event id=" " event_id=" " type_id=" " period_id=" " min=" " sec=" " team_id=" " outcome=" " x=" " y=" "
timestamp=" " last_modified=" " version=" ">
    <Q id=" " qualifier_id=" " value=" " />
    <Q id=" " qualifier_id=" " value=" " />
    <Q id=" " qualifier_id=" " value=" " />
    <Q id=" " qualifier_id=" " value=" " />
    <Q id=" " qualifier_id=" " value=" " />
    <Q id=" " qualifier_id=" " value=" " />
  </Event>
  <Event id=" " event_id=" " type_id=" " period_id=" " min=" " sec=" " team_id=" " outcome=" " x=" " y=" "
timestamp=" " last_modified=" " version=" ">
    <Q id=" " qualifier_id=" " value=" " />
    <Q id=" " qualifier_id=" " value=" " />
    <Q id=" " qualifier_id=" " value=" " />
    <Q id=" " qualifier_id=" " value=" " />
    <Q id=" " qualifier_id=" " value=" " />
    <Q id=" " qualifier_id=" " value=" " />
  </Event>
  <Event id=" " event_id=" " type_id=" " period_id=" " min=" " sec=" " team_id=" " outcome=" " x=" " y=" "
timestamp=" " last_modified=" " version=" ">
    <Q id=" " qualifier_id=" " value=" " />
  </Event>
</Games>

```

Figure 4.1: Illustration of Opta’s F24 data structure. A single event is highlighted in the red box.

Three Opta expert analysts collect the data during each game, utilizing specialized software. One expert collects home team data, one collects away team data, and the third checks data consistency. In addition to the live data collection, there is a post-match data check within 48 hours to ensure data accuracy (Opta Sports, 2016). Although rigorously controlled by Opta, the data collection process may introduce some human collection errors or inaccuracies.

The data is read, restructured and organized in a relational database. Every XML file is read, using the Java Document Object Model interface, and all events and event attributes are stored in the database. The event class in the database stores the data in an intelligible and manageable format opposed to the XML files, and contains most of the information needed to build the Markov models. The event feed, represented by the event class in the database, is read when building the models and the associated states, further described in the model descriptions.

4.2 Model 1

Model 1 resembles the second model presented by Bjertnes et al. (2016), reviewed in Section 3.3. The games from Eliteserien seasons 2014, 2015 and 2016 are modelled as a Markov chain reward model, following the theory introduced in Section 2.1. There are no agents taking decision based actions to move from one state to another, which is why this model is not considered a Markov game. A description of the state definition is presented next, followed by how Model 1 was constructed and the value iteration process.

4.2.1 State Definition

As explained in Section 2.1, the state describes the environment which the stochastic process is currently in. The state definition includes context variables and event type, illustrated in Figure 4.2. The context variables explain the circumstances in which an event

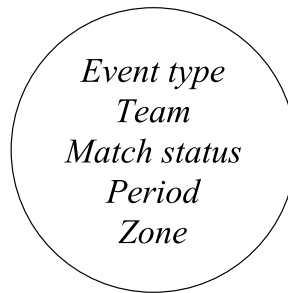


Figure 4.2: Model 1's Markov chain state definition.

occurred, and are summarized in Table 4.1 with their respective, possible values. The event types are listed in Table 4.2.

Table 4.1: The context variables of Model 1 and their respective, possible values.

Variable	Possible values
<i>Team</i>	[Home, Away]
<i>Match status</i>	[-1, 0, 1]
<i>Period</i>	[1, 2, 3, 4]
<i>Zone</i>	[1, . . . , 21]

Team tells which team performs the event described by the event type variable, and takes one of two values, home or away. *Match status* takes the value 0 when the score is even, -1 when the away team leads and 1 when the home team leads. The frequency of different event types are believed to be influenced by the score. For instance, a team that already leads may take fewer chances, leading to a lower frequency of shots. The expected value of otherwise equal states may also be affected by the score, e.g. certain passes may have a higher expected value if the team already leads. The opponent may push more players forward, leaving them vulnerable in defence, in turn making a goal more likely.

The pitch is divided into 21 zones, illustrated in Figure 4.3. *Zone* is assigned to a state from the perspective of the team given by the *Team* variable, where this team plays from left to right. This means that zones 1-3 are the most defensive zones, and zones 18-21 are the most offensive. A higher number of zones will result in more accurate modelling, but also rapid state space growth. Too many states will lead to few occurrences of each state, which is likely to cause unreliable state values. The difference between zones are thought to be bigger in the offensive half. Therefore, the offensive half is divided into 12 zones, which is three more than the defensive half. In comparison, Bjertnes et al. (2016) used 15 zones where only three were in the defensive half. The increased number of zones is considered an improvement which is likely to result in more precise modelling.

Period distinguishes otherwise equal states at different points in time of the games. The variable takes value 1 during the first 23 minutes, value 2 during the remaining time of the first half, value 3 during the first 23 minutes of the second half and value 4 during the remainder of the game. An otherwise equal state may have a different expected value

1	4	7	10	14	18
2	5	8	11	15	19
3	6	9	12	16	20
			13	17	21

Figure 4.3: Illustration of how the pitch is divided into zones. The active team plays from left to right and *Zone* takes the value of the zone where the event happens.

in one period compared to another. Players are expected to be more tired towards the end of the game. This might make otherwise equal states more likely to lead to a goal later, compared to earlier, in the game.

The possible *Event type* values are shown in Table 4.2. Most of the event types are given by Opta. Opta defines a long pass as a pass over 32 metres and a cross as a ball played into the box from wide areas. These definitions are used in the model. Two variables are added to the existing Opta feed by the authors, namely Ball carry and Ball received. Ball carry is added because there are gaps between events in the data. Opta do not consider simply moving the ball as an event unless the player takes on an opponent. However, it is included in the model because bringing the ball into a more dangerous position can be valuable. Ball received is included to fill another gap in the data. Opta includes if a pass reaches a player on the same team or not, but does not actually note an event when the ball is received. This led to some unintuitive event sequences where Pass was directly followed by Tackle. This is solved by adding ball received. A similar ball carry event type was introduced by Bjertnes et al. (2016), but ball received has not been used in the previous work.

Table 4.2: The event types which the *Event type* variable can take in Model 1.

Pass	Long pass	Cross
Free kick pass	Corner	Throw in
Take on	Ball carry	Ball touch
Aerial duel	Ball received	Ball recovery
Shot	Headed shot	Shot blocked
Tackle	Interception	Clearance
Foul committed	Fouled	

States with the Aerial duel event type always appear in pairs, one for each team. The second aerial duel state is then followed by a pass or shot state from the team that won the duel. Another such pair is Fouled and Foul committed. Ball recovery occurs when neither team was in control of the ball and usually occurs after a tackle, a clearance or a blocked shot. Shots are split into two different variables, headed shot and shot. Shot is used for every shot which is not a header, as classified by Opta. Distinguishing headers from regular shots is considered an improvement on the models by Bjertnes et al. (2016).

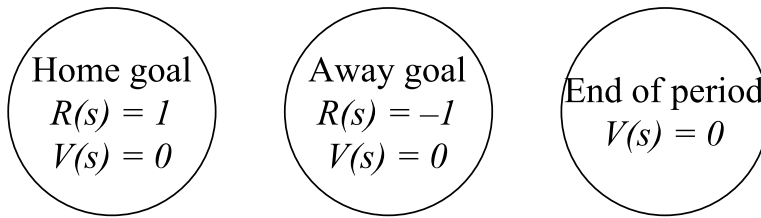


Figure 4.4: The absorbing Markov model states. The state values are set to 0 and the goal states have reward 1 or -1 .

The model considers the probability of scoring in a given state as the total number of goals divided by the number of attempts. It is generally accepted that, all else equal, it is harder to score from a header than from a regular shot. This is supported by the expected goal models by Bjertnes et al. (2016) and Caley (2015). Consequently, headed shots will have a too high expected value, while regular shots will have a too low expected value, when headed and regular shots are considered equal. This leads to inaccurate, and sometimes unfair, evaluation of shots, in turn leading to inaccurate player evaluations.

All shot and headed shot states that result in a goal lead to one of two artificial states, Home goal or Away goal. Two separate goal states are necessary because they have different rewards. Home goal has reward $R(s) = 1$ and Away goal has reward $R(s) = -1$. Every other possible state has reward $R(s) = 0$. The Home goal and Away goal states are absorbing, meaning no transitions are made out of these states. A third absorbing state, End of period, is used at the end of each half of each game. The absorbing states do not have context variables, and are illustrated in Figure 4.4.

Shot blocked was not included by Bjertnes et al. (2016), but is included in an attempt to better reward defenders. States with blocked shot as *Event type* will always follow states with Shot or Headed shot as *Event type*. Interception and Ball touch are fairly similar event types, and both represent events where players are not in control of the ball. Interception is used when a player stops a pass from reaching its target, while ball touch is used in every other situation where a player touches the ball without controlling it. The remaining event types in Table 4.2 are self explanatory.

A manpower difference variable was applied in the Markov models by Bjertnes et al. (2016). This variable would show the difference in number of players on the pitch for the two teams, primarily affected by red cards. Red cards are rare in Eliteserien, which means states with manpower difference different from zero have few occurrences. Consequently, the state values are likely to be unreliable, and the manpower difference variable is not used in this thesis.

A total of $2 \times 3 \times 4 \times 21 = 504$ combinations of *Team*, *Match Status*, *Period* and *Zone* exists. Multiplied by the 20 different event types, $504 \times 20 = 10080$ states are possible. Additionally, the three artificial absorbing states makes the total number of theoretically possible states 10083. However, some states are very unlikely, such as shots in the defensive zones. The actual observed number of states in seasons 2014, 2015 and 2016 in Eliteserien is 7738.

4.2.2 Constructing the Markov Chain

In order to construct the state space of the model, the event feed in the database is read in the order which the events occurred, game by game. The state definition of Model 1 allows each event in the database feed to map to exactly one state. However, several different events may map to the same state. States are constructed the first time they appear while reading the data. The attributes of each event are compared to the already constructed states while iterating the events. If the event maps to an already constructed state, the occurrence of that state, $Occ(s)$, increments and the event is updated with the state's identifier in the database. If the event does not match an already constructed state, a new state is constructed with $Occ(s) = 1$.

When all events in the database are updated with a state identifier, the complete state space is constructed. Next, state transitions are constructed. The event feed is re-read, focusing on each event's newly added state identifier, which points to a specific state. Two subsequent events are processed at a time. A state transition is defined by two states, and represents a transition from one state, s , to another state, s' . For each pair of events, the state identifier of the event that occurred first is treated as s , and the state identifier of the following event as s' . State transitions are constructed the first time two events corresponding to the states s and s' occur consecutively. Each pair of events map to exactly one state transition, but several event pairs may map to the same state transition. The event pairs are compared to the existing state transitions. A new state transition is constructed if none of the previously constructed state transitions has the identifier of s as start and the identifier of s' as end. The occurrence, $Occ(s, s')$, of new transitions is set to 1 and is incremented if the state transition has already been constructed.

In addition to the states observed in the data, the three artificial states, representing *Home goal*, *Away goal* and *End of period*, illustrated in Figure 4.4, are inserted into the database. These states represent absorbing states, meaning that no state transition starts in these states and the state values of these states will never be updated. The artificial goal states are the only states which have rewards different from 0.

4.2.3 Value Iteration

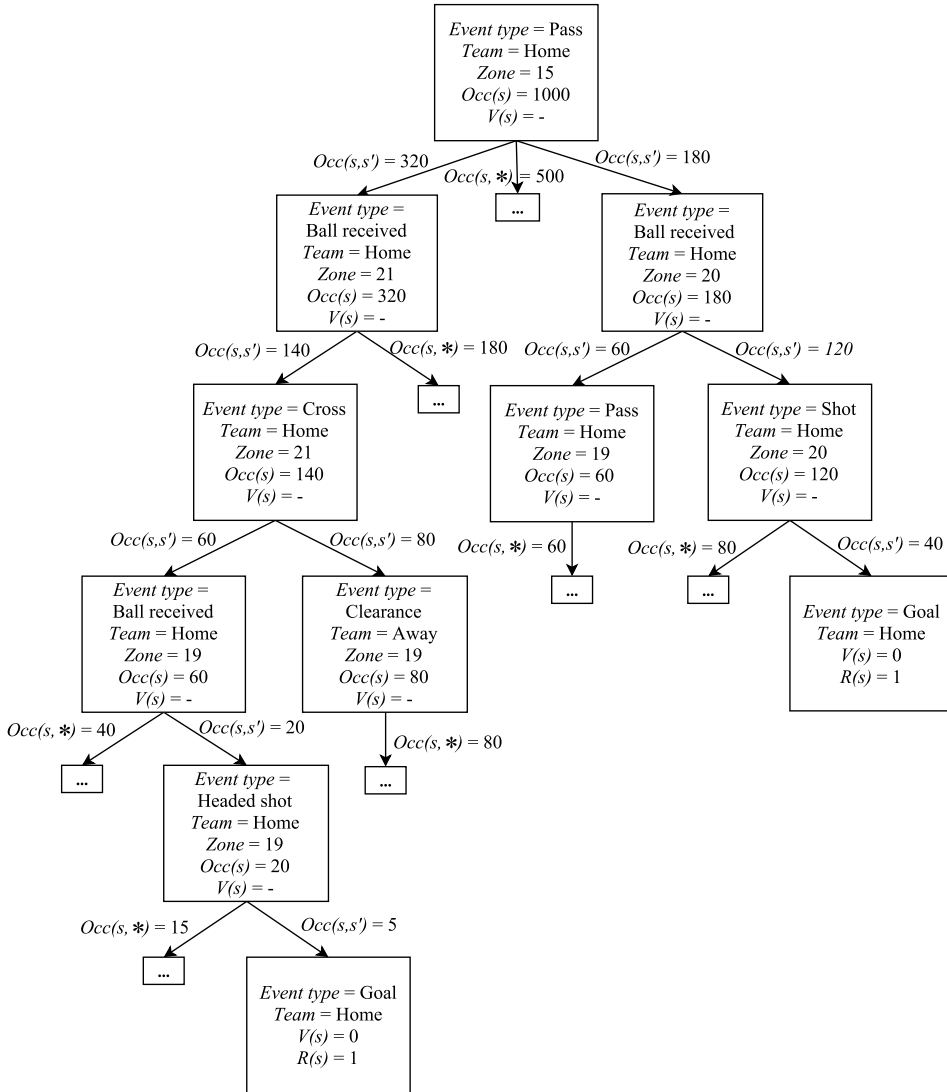
The state values, $V(s)$, can be learned from the observed states and state transitions, with the respective occurrences, stored in the database. Algorithm 1, presented in Section 2.1, is applied to learn the state values. A convergence criterion, which determines when the algorithm terminates due to a small change in the state values, of 0.00001, and a maximum number of 10000 iterations is used. The values converge after 247 iterations. Each iteration of the outermost loop in the algorithm runs through all state transitions listed in the database. For each state transition, the state value of the start state is updated. In one iteration, the state value of a state s , is updated for each end state, s' , where a transition from s to s' exists. As the value of a goal in football is independent of how many events have occurred before the goal, an undiscounted model is used. Choosing the model to be undiscounted, practically means the discount factor β is set to 1. Consequently, the state values are updated according to the equation

$$V(s) = \frac{1}{Occ(s)} \sum_{s' \in \mathcal{S}} Occ(s, s')(R(s') + V(s')) \quad (4.1)$$

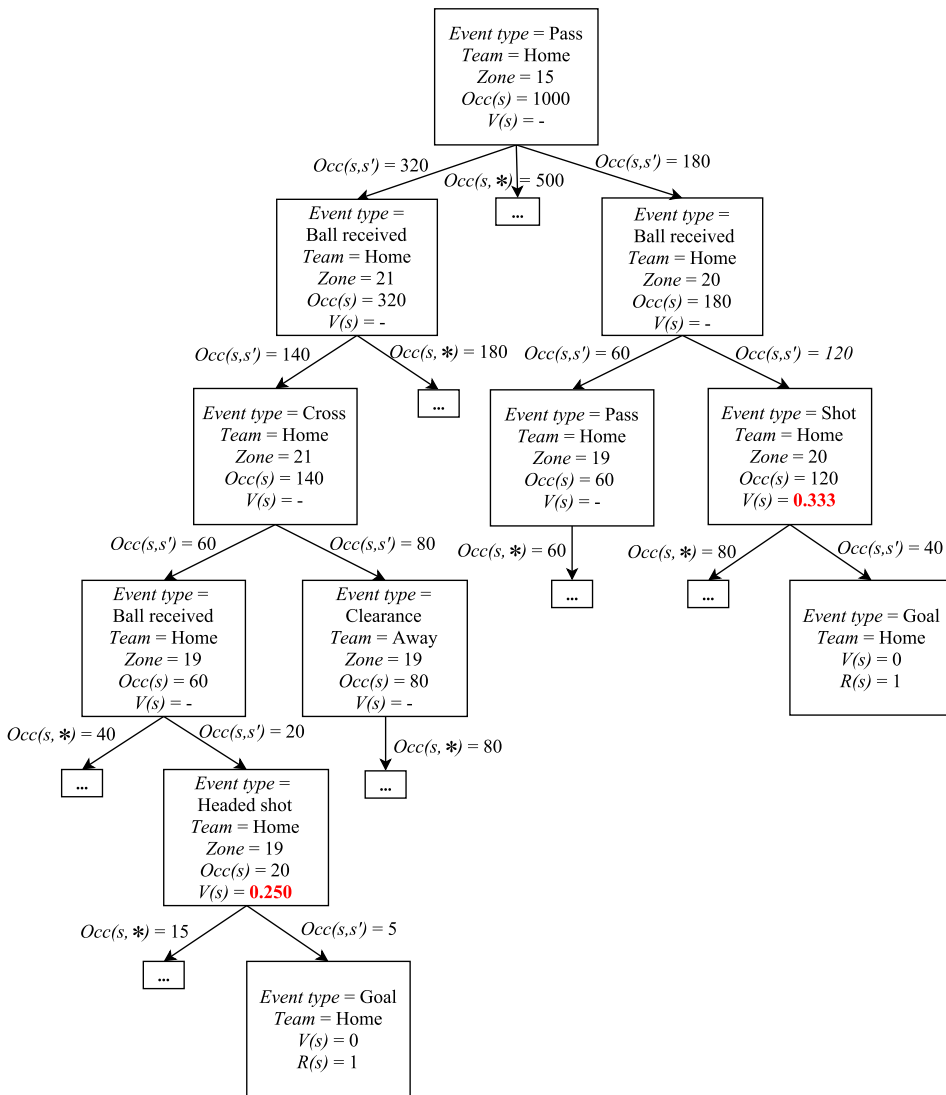
which is equivalent to Equation 2.3 with $\beta = 1$.

While within a specific iteration, the state values are stored temporarily. The state values are updated after each iteration is finished. Only the goal states has a reward different from 0, and only the absorbing states has an initial state value, which is set to 0. Consequently, the only state values updated in the first iteration are the values of the states where a transition between the state and a goal state exists. Figure 4.5 illustrates the algorithm. Figure 4.5a shows the states and the state values before running the algorithm. In order to simplify the figure, the context variables *Match status* and *Period* are not shown. After iteration one, the state values of states leading directly to a goal state have been updated, as shown in Figure 4.5b. The red numbers in the figure illustrate the values that have been updated in the last iteration. As more iterations are completed, more state values are updated, until the algorithm eventually converges. The algorithm has not yet converged after five iterations in the example, shown in Figure 4.5c. $Occ(s, *)$ denotes the occurrence of transitions from state s to any state not shown in the example. The intermediate iterations are illustrated in Figure A.1 in Appendix A.

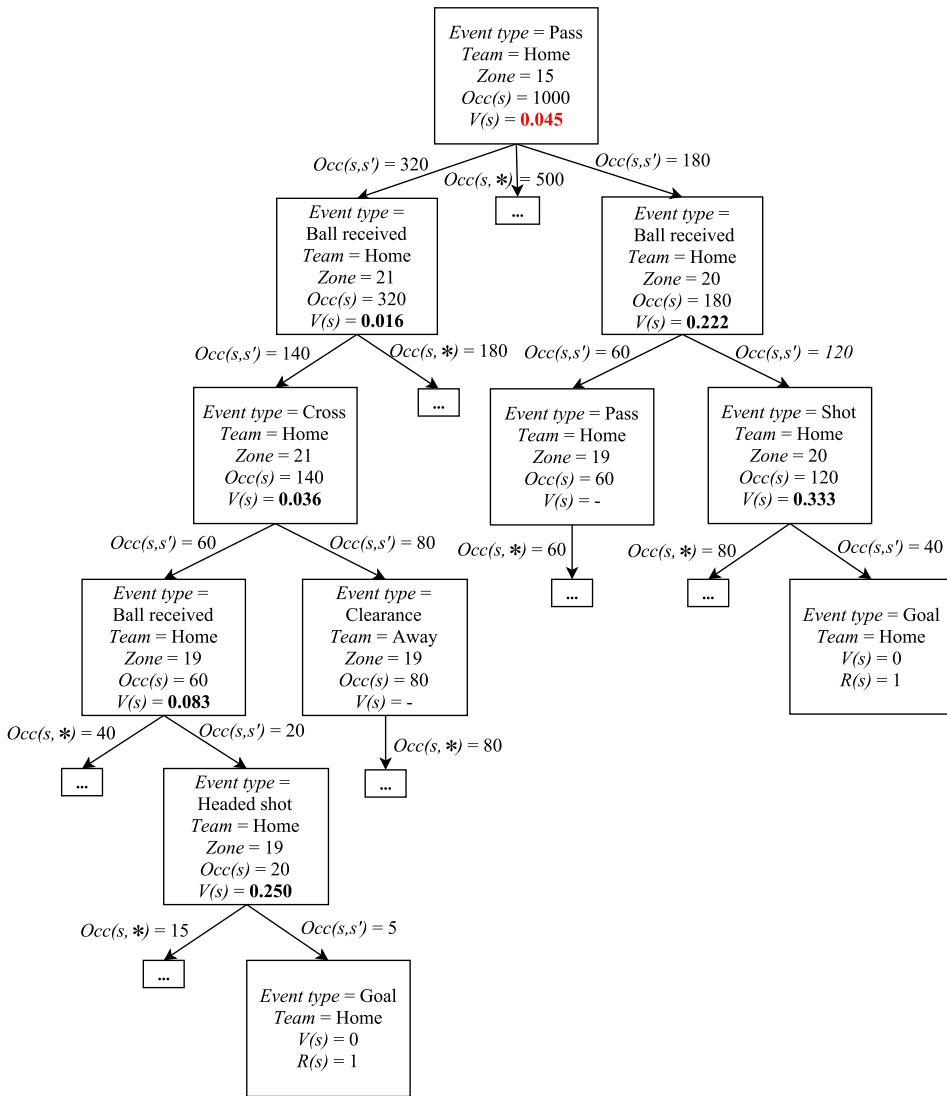
Figure 4.5: Model 1 value iteration example, illustrating Algorithm 1.



(a) Before running the algorithm. No state values are set, except the goal states, which have a value of 0 and reward of 1 or -1 . States not illustrated explicitly in the figure are denoted $*$.



(b) After iteration 1. The state values of the states leading directly to a goal are updated. The updated values are highlighted in red. One state lead to a goal in one out of four occurrences. Another state lead to a goal in one out of three occurrences. Thus, after iteration 1, these states are valued 0.250 and 0.333, respectively.



(c) After iteration 5. The values of most states shown in the figure have been updated. The state value updated in the current iteration is highlighted in red, while state values in bold have been updated in previous iterations. As subsequent states not illustrated explicitly in the figure, denoted *, are yet to be updated, the state values in the figure will be updated in later iterations. The algorithm converges when the state values changes by less than the predetermined convergence criterion.

4.3 Model 2

Model 2 resembles the first model presented in Bjertnes et al. (2016), reviewed in Section 3.3, but introduces several proposed improvements to the previous work. The games from Eliteserien, seasons 2014, 2015 and 2016 are modelled as a zero-sum Markov game, introduced in Section 2.2. The state definition and agent actions are presented next, followed by how the Markov game was constructed and the value iteration process.

4.3.1 State Definition and Actions

As explained in Section 2.2, the state defines the current environment of the game. The state definition is illustrated in Figure 4.6. Four context variables are shown in Table 4.3 with their respective possible values. The state definition differs from Model 1 by event type not being included in the state definition. An event initiates a transition in Model 2, and maps to a Markov game action. Consequently, Model 2 is considered a Markov game, while Model 1 is a Markov chain. The modelling approach translates each team involved in a game to a Markov game agent. This means that two agents are playing a zero-sum Markov game, as the two teams have diametrically opposite goals. Each team is represented by its players, meaning that the players take actions on behalf of their team. Because only one team makes an action at a time, the Markov game is considered a Markov decision process, following the theory presented in Section 2.2.

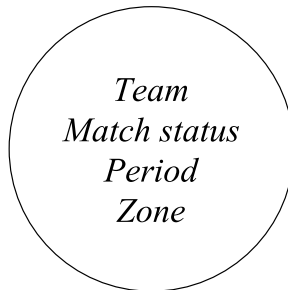


Figure 4.6: Model 2's Markov game state definition.

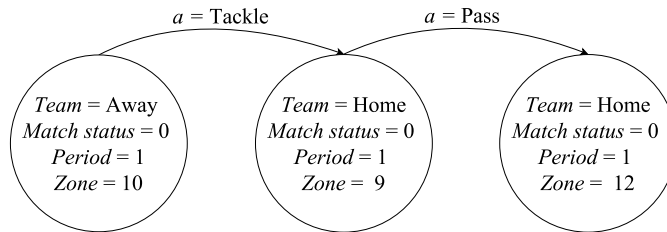
Zone is the same as in Model 1 and takes values between 1 and 21. Figure 4.3 in Section 4.2 illustrates how the pitch is divided into zones. The zones are assigned to the states from the perspective of the team who is in possession of the ball, given by the *Team* variable, where the team plays from left to right.

Team is included to describe which team is in possession of the ball. This is different from Model 1, where *Team* describes the team which performed the event. Figure 4.7 illustrates this difference by an example where a home team player tackles an away team player and wins possession. Figure 4.7a shows the example in terms of Model 2. The away team has possession, shown by the *Team* variable. The home team chooses the tackle action, which initiates the transition. The tackle leads to the home team having possession, meaning $Team = \text{Home}$ in the next state. Figure 4.7b illustrates the same example in terms of Model 1. The difference is that the team variable in Model 2 takes

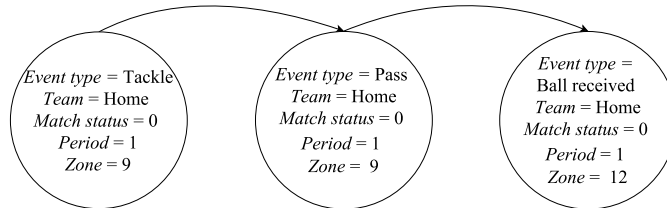
Table 4.3: The context variables of Model 2 and their respective, possible values.

Variable	Possible values
<i>Team</i>	[Home, Away, None]
<i>Match status</i>	[-1, 0, 1]
<i>Period</i>	[1, 2, 3, 4]
<i>Zone</i>	[1, . . . , 21]

the value of the team who performs the tackle, i.e. the home team in the example. A state transition is made to the state which corresponds to the next event in Model 1, where *Event type* in the end state takes the value of the next event, which is a pass in the example. The use of the team variable in Model 2 is believed to more closely model the flow of football games, and to make the state transitions more intuitive, compared to Model 1.



(a) Model 2 tackle illustration. The tackle by a home team player initiates a transition from a state where *Team* = Away to a state where *Team* = Home.



(b) Model 1 tackle illustration. A home team player performs the tackle, hence the first state has *Event type* = Tackle and *Team* = Home. The next state has *Team* = Home because the tackle was successful.

Figure 4.7: Illustration of how a sequence involving a tackle is modelled differently in Model 2 compared to Model 1. A home team player tackles an away team player and the home team player wins the ball. The home team player then makes a successful pass. The important difference is the value of the *Team* variable in the first state.

The purpose of the team variable in Model 2 is to model who is in possession of the ball. Therefore, the third variable value None is added, used when neither team controls the ball. For instance, Opta’s ball recovery event is used when neither team was in control of the ball. Consequently, whenever a ball recovery action occurs, it must initiate a transition

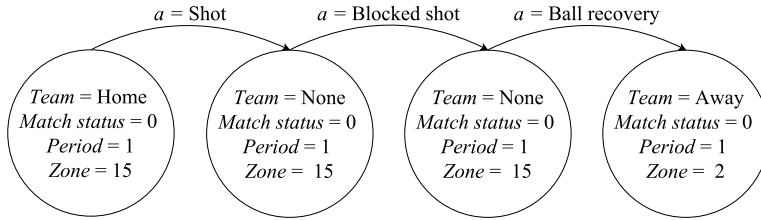
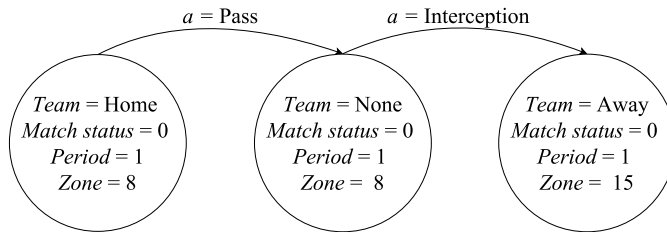
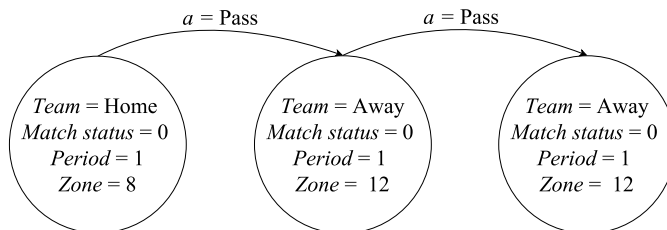


Figure 4.8: Ball recovery illustration. The ball recovery happens after a shot which is blocked. No-one controls the ball prior to a ball recovery, hence the state before the ball recovery has $Team=None$. $Team=Away$ in the next state as an away player recovered the ball.

from a state with $Team = None$ to a state where $Team$ is the team who performed the ball recovery. This is illustrated by Figure 4.8, where the home team performs the ball recovery action. In the example, the ball recovery follows a shot which was blocked. None as a $Team$ value has not been applied in previous work and is considered an improvement. By introducing this value, the model will resemble the actual flow of the game more closely. This is thought to result in more precise valuation of states and state-action pairs.



(a) Unsuccessful pass followed by non-offensive action. The transition initiated by the pass leads to a state where $Team=None$ because neither team had control of the ball after the pass.



(b) Unsuccessful pass followed by offensive action. The transition initiated by the pass leads to a state where $Team=Away$ because the away team captured and controlled the ball after the unsuccessful pass.

Figure 4.9: Illustration of how unsuccessful passes are modelled in Model 2. The $Team$ variable in the second state depends on the second action.

None is also used when passes are unsuccessful, which is given by the data, unless it leads to the other team performing an offensive action directly. Figure 4.9a illustrates

an unsuccessful pass followed by an interception. A player is not in control of the ball when he intercepts, hence the pass action must initiate a transition to a state with $Team = None$, and the interception must initiate a second transition to a third state. The away team takes control of the ball after the interception in the example. Figure 4.9b illustrates an unsuccessful pass by the home team which is directly followed by a long pass by the away team. In order for the away team to perform the long pass, the unsuccessful pass from the home team must have lead to the away team having possession directly, as illustrated in the figure. $Team = None$ in the start state whenever a state transition is initiated by an aerial duel. An aerial duel means the ball was in the air and neither team was in possession, hence $Team = None$ is appropriate.

The *Period* and *Match status* variables are similar to that of Model 1. *Period* takes the value 1 during the first 23 minutes, value 2 during the remaining time of the first half, value 3 during the first 23 minutes of the second half and value 4 during the remainder of the game. *Match status* takes the value 1 when the home team leads, -1 when the away team leads, and 0 when the game is even.

Table 4.4: The action set \mathcal{A} in Model 2.

Pass	Long pass	Cross
Free kick pass	Corner	Throw in
Take on	Ball carry	Ball touch
Aerial duel	Foul committed	Ball recovery
Shot	Headed shot	Shot blocked
Tackle	Interception	Clearance

Table 4.4 shows the set of available actions, \mathcal{A} . Both teams can choose from the same set, but only one team will make an action at a time. However, when a certain state occurs, the set available to each team is practically a subset of \mathcal{A} and will be different for the two teams. For instance, the team in possession will not choose to tackle, and the team not in possession can not choose to pass. \mathcal{A} largely resembles the set of event types in Model 1. Because of how *Team* behaves in Model 2, Ball received is redundant and therefore not used. Fouled is also left out because it does not fit with how the flow of the game is modelled. Instead, only Foul committed is included. For a more detailed explanation of the actions, see the description of the different event types in Section 4.2.

The number of possible values of each context variable, shown in Table 4.3, means that the number of possible states is $3 \times 3 \times 4 \times 21 = 756$. Additionally, the three artificial, absorbing states, *Home goal*, *Away goal* and *End of period*, are included, making the number of possible states 759. 18 different actions are available, meaning the number of possible state-action pairs is $756 \times 18 = 13608$. All 759 states occur in the data, while 8085 different state-action pairs are observed.

4.3.2 Constructing the Markov Game

To construct the Markov game, the event feed is read from the database, game by game. The events from each game are read in the order which they occurred. States, state transitions and state-action pairs are built while iterating the events. The introduction of None as a *Team* value makes the process of building state and state transitions more complex,

compared to Model 1. For instance, before a ball recovery can be inserted, an additional state must be inserted with $Team = None$, and the previous transition must be manipulated to lead to this state. This is also the case for the different pass events which may lead to an aerial duel. Furthermore, unsuccessful passes that initiate a transition to a state with $Team = None$ are separated from passes that initiate a transition to a state where $Team$ takes the value of the opposing team.

Each event maps to a state transition, and the start state and action is given by the event. The end state is based on the context of the game and where the ball ends after the event. Depending on the action type, the next state is either given by the end coordinates of the current event or by the start coordinates of the next event. States and state transitions are stored in the database with unique identifiers, and each event is marked with the identifier of the corresponding state transition in the database. A state transition consists of a start state s , an end state s' and an action a . A new state is created and stored in the database the first time it occurs as a start state. The occurrence, $Occ(s)$, is set to 1. Similarly, new state transitions and state-action pairs are created the first time they occur, with the respective occurrences, $Occ(s, a, s')$ and $Occ(s, a)$, set to 1. If an event corresponds to a start state, a state transition or a state-action pair that already exists, the corresponding occurrence is incremented by 1. The three absorbing states, Home goal, Away goal and End of period are added manually. The goal states are the only states with rewards different from 0, and the absorbing states have value $V(s) = 0$.

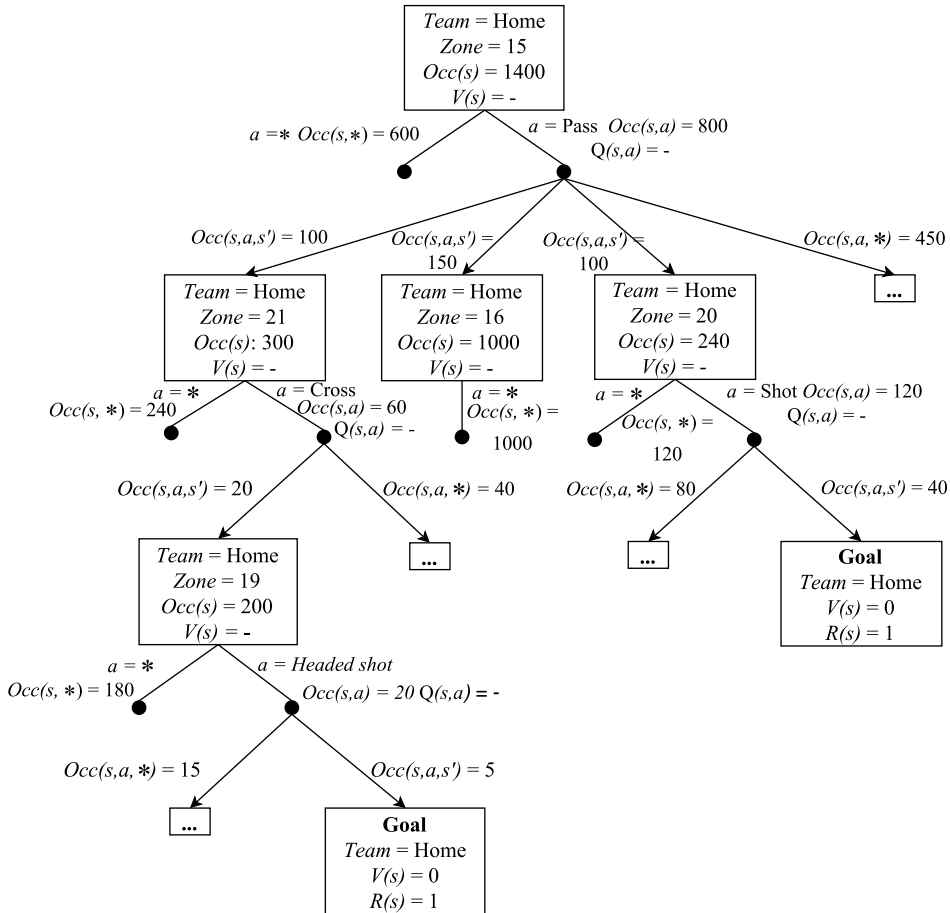
4.3.3 Value Iteration

Algorithm 2, presented in Section 2.2.3, is applied to learn the state values, $V(s)$, and the Q -values, $Q(s, a)$. A convergence criterion, which determines when the algorithm terminates due to a small change in the state values, of 0.00001, and a maximum number of 10000 iterations is used. The values converge after 162 iterations. The outermost loop iterates all state-action pairs in each iteration and updates the Q -values. Each Q -value is updated according to the equation

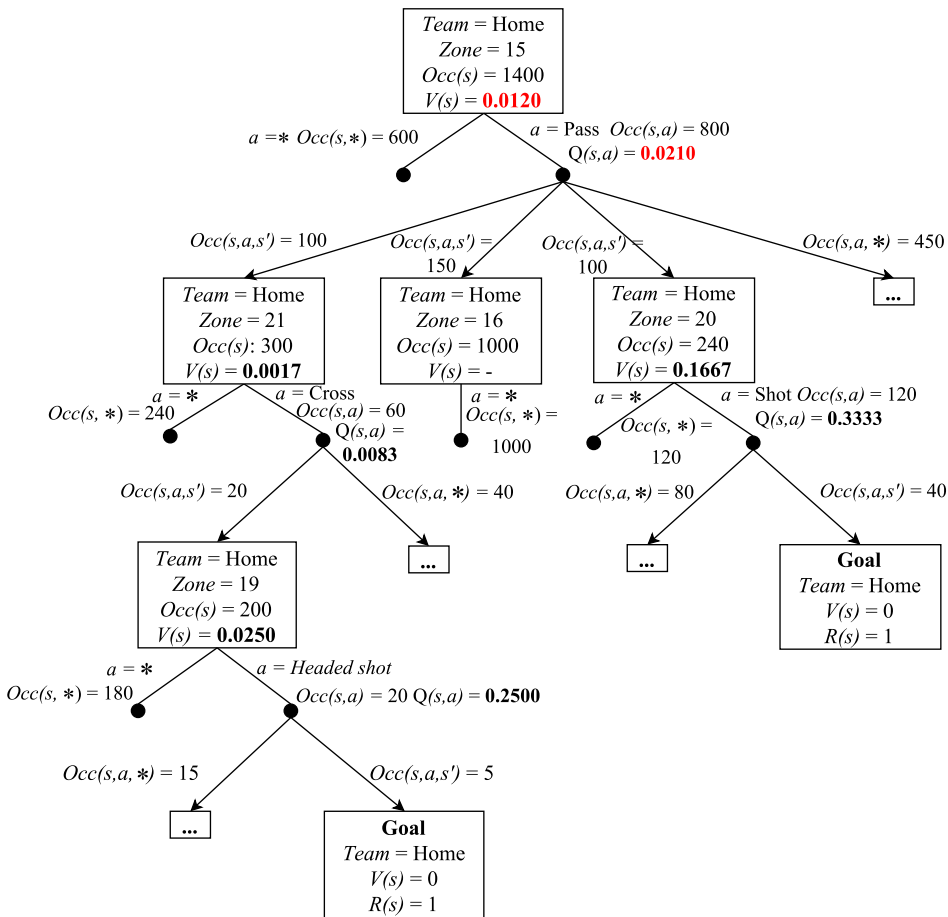
$$Q(s, a) = \frac{1}{Occ(s, a)} \sum_{s' \in \mathcal{S}} Occ(s, a, s') (R(s') + V(s')) \quad (4.2)$$

which is equivalent to Equation 2.11, setting the discount factor $\beta = 1$. As the value of a goal in football is independent of the number of prior events, $\beta = 1$ is appropriate. The state values, $V(s)$, are updated after each iteration, according to Equation 2.12. Figure 4.10 illustrates an example of running the value iteration algorithm for Model 2. As shown in Figure 4.10a, only the absorbing home goal state has an initial value, set to 0, and reward $R = 1$. To simplify the figure, the context variables $Match\ status$ and $Period$ are not shown, and * denotes any state or action not shown in the example. After iteration 3, illustrated in Figure 4.10b, most of the state values and Q -values have been updated, but the values will continue to change as long as there are other states and state-action pairs being updated. The intermediate iterations are illustrated in Figure A.2 in Appendix A.

Figure 4.10: Model 2 value iteration example, illustrating Algorithm 2.



(a) Before running the algorithm. No state values and Q -values are set, except the goal states, which have a state value of 0 and reward of 1 or -1 . Subsequent states and actions not explicitly illustrated in the figure are denoted $*$.



(b) After iteration 3. Most state values and Q -values shown in the figure are updated. Values updated in the current iteration are highlighted in red, while values in bold have been updated in previous iterations. Subsequent states and actions not explicitly illustrated in the figure are denoted $*$. As the values of subsequent states and state-action pairs, are yet to be updated, the illustrated values will be updated in later iterations.

4.4 Modelling Examples

The previous sections have presented and described the characteristics of the two Markov models. As illustrated, Model 1 and Model 2 model the game differently. In order to further clarify the differences between the models, the modelling of a longer sequence is illustrated in this section. This is illustrated with two different figures, one representing Model 1 and one representing Model 2. The sequence is taken from the match between Lillestrøm and Rosenborg in Eliteserien 2016, Rosenborg being the away team. In the modelled sequence, a Rosenborg player carries the ball from the right side of the pitch, inwards across the halfway line. The ball is passed short, and then passed further forward

into the penalty area. Here, the ball is passed short, before a shot is made. The shot is blocked by a Lillestrøm player, another Lillestrøm player recovers the ball outside the penalty area and then the player makes a pass.

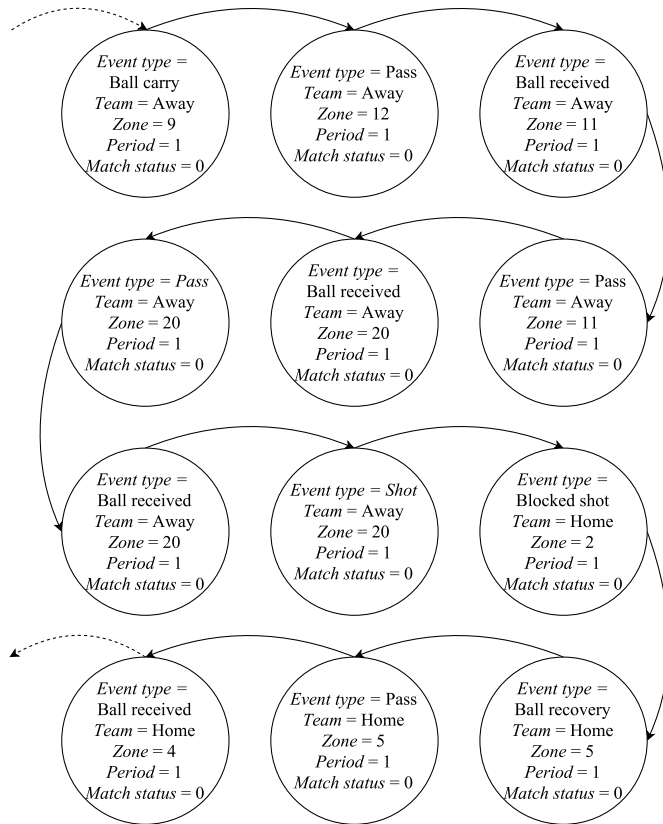


Figure 4.11: An illustration of how the game sequence example is modelled according to Model 1.

Figure 4.11 illustrates how Model 1 models the sequence. Recall that Model 1 is a Markov chain, and thus no agents are taking actions to move from one state to another. The first state in the figure illustrates a state with *Event type* = Ball carry, which is one of the event types added by the authors. A transition to the next state is made, where *Event type* = Pass. The next state describes the away player receiving the ball in zone 11. This event type is the second event type added by the authors to make a more realistic model of the game flow. Four states describing passes and players receiving the ball follow, before a state with *Event type* = Shot occurs. The shot is blocked by the home team in the next state, and then recovered in the following state. The home team then makes a pass and receives the ball in the next states before the game continues beyond the illustration.

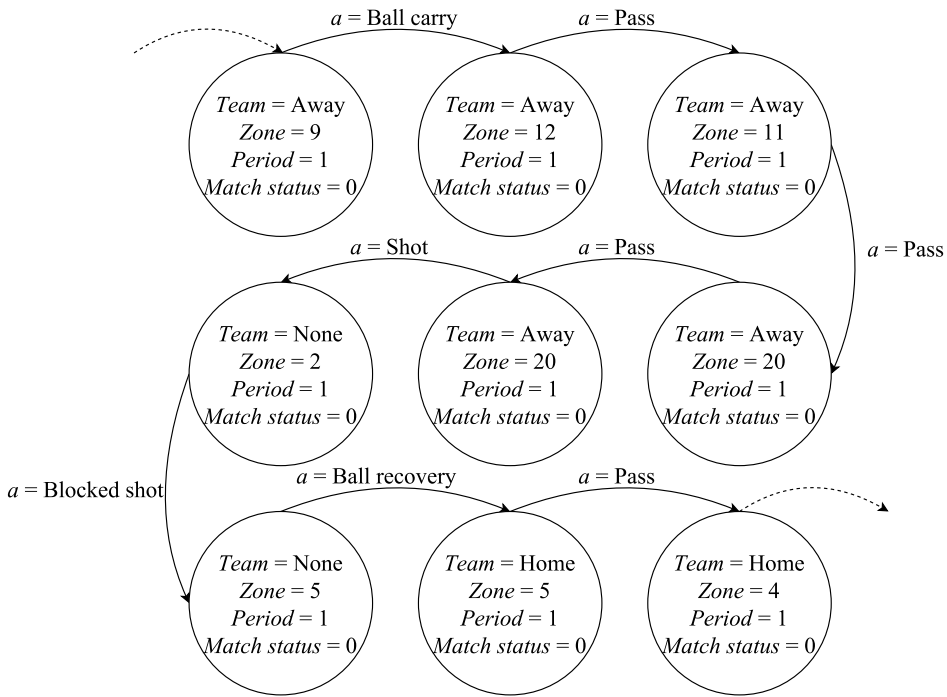


Figure 4.12: An illustration of how the game sequence example is modelled according to Model 2.

Figure 4.12 illustrates the same sequence, modelled by Model 2. As Model 2 is a Markov game, actions initiate transitions between states. Fewer states are needed to model the sequence, compared to Model 1, because Model 1 requires states with *Event type* = Ball received. Additionally, the team variable can take the value None in Model 2. The first state in Figure 4.12, illustrates a home player in possession of the ball in zone 9. Then an action, $a = \text{Ball carry}$, is made, moving the game into the next state in zone 12. The figure illustrates two passes moving the game into zone 11 and then into zone 20. Another pass is made, but the following state is the same as the previous state. Then a shot is made, moving the game into a new state where $Team = \text{None}$. In this state, neither the home nor the away team is in possession of the ball. A new action is made, taking the value Blocked shot. In the following state, the team variable is still None. The next action is a ball recovery, which moves the game into a state where the home team is in possession of the ball in zone 5. A pass is made, and the home team possesses the ball in zone 4, illustrated by the last state in the figure.

Experimental Setup

The experimental setup is described in this chapter. The experiment is designed to answer the research questions presented in Section 1.2. First, four impact functions, measuring the values of player involvements, are presented. The state and state-action values, $V(s)$ and $Q(s, a)$, from the Markov models, presented in Chapter 4, are used to estimate player impact. The impact functions are used to evaluate player performances, and aggregated and normalized impact values are proposed as data driven player ratings. Next, the validation methods are presented. The player ratings must be carefully validated, in order to ensure that player performance is described by the ratings. The goal of the validation is to identify which of the four impact functions result in the most accurate player evaluations. Next, an approach to identify similar players is presented. Similarity is measured by utilizing the player ratings and the nearest neighbour methodology. Last, an approach to measure teams' impact on individual player performances is presented. Here, players who have played for more than one team and players that have played for the same team in successive seasons are studied independently.

5.1 Player Performance Evaluation

The proposed procedure for evaluating player performance is presented in this section. The players are rated based on their individual involvements throughout the games which they have played. The values of states, $V(s)$, in Model 1 and the values of states and state-action pairs, $V(s)$ and $Q(s, a)$, in Model 2 make up the foundation used to evaluate individual player involvements. The value of an individual player involvement is referred to as impact. Four different impact functions are proposed as measures of the value provided by a player towards his team through his involvement. The impact functions are denoted $I_{i,j}$, where $i = 1, 2$ is the Markov model number, and j is the impact equation number of model i . The states in Model 1 include the *Event type* variable in the state definition, meaning player involvements are represented by states in Model 1. Consequently, Model 1 has only state values and no state-action values. Player involvements are represented by Markov game actions in Model 2. As such, the state values of Model 1 closely

resemble the state-action values of Model 2.

Two impact functions are constructed from Model 1 and two from Model 2. Recall that the two teams have diametrically opposite goals, meaning positive state and state-action values are preferred for the home team, while negative state and state-action values are preferred for the away team. As such, the impact functions presented below apply when a home team player makes an involvement, while the impact functions must be negated for away team players.

$$I_{1,1} = V(s) \tag{5.1}$$

$$I_{1,2} = V(s') - V(s) \tag{5.2}$$

$$I_{2,1} = Q(s, a) \tag{5.3}$$

$$I_{2,2} = V(s') - Q(s, a) \tag{5.4}$$

$V(s)$ and $Q(s, a)$ represent the values of the state and state-action pair corresponding to the current player involvement, and $V(s')$ correspond to the value of the next state. The general idea is that players are rewarded when their involvement has a positive expected value and punished when their involvement has a negative expected value. A positive expected value indicates that the player's team has the highest probability of scoring the next goal, while a negative expected value indicates that the opposing team has the highest probability of scoring the next goal. As the reward of an away goal is -1 , negative state values are beneficial for the away team. The impact functions are thus negated when giving impact to the away players. Following impact function $I_{1,1}$, for instance, an away player is given an impact of 0.25 if $V(s) = -0.25$ and -0.25 if $V(s) = 0.25$. For each event that maps to a player involvement, the required state and state-action values are extracted from the database, and the involved player receives the corresponding impact values. A higher positive impact is always better for the involved player.

$I_{1,1}$ rewards players the expected value of their involvement without looking at the next state. A drawback of this method is that it does not distinguish between successful and unsuccessful involvements. For instance, a specific pass, as defined by the state, which reaches a teammate results in the same impact as if the pass went to an opponent. However, the state values represent the expected value of the state, indicating fair values in the long run. Furthermore, $I_{1,1}$ is similar to the impact function concluded to result in valid and reliable player ratings in Bjertnes et al. (2016). The second impact function of Model 1, $I_{1,2}$, takes the next state into account, and can be interpreted as how much value the player adds above what is expected. The drawback of distinguishing successful and unsuccessful involvements associated with $I_{1,1}$ is limited, but the impact is dependent on the next player's involvement. $I_{1,2}$ is believed to capture how the player involvement influenced the flow of the game to a greater extent than that of $I_{1,1}$. However, $I_{1,2}$ may lead to players being unfairly punished or rewarded for the next player's involvement.

Model 2 separates the involvement from the environment, which lead to a more natural modelling of football. This may in turn lead to more accurate measures of impact. $I_{2,1}$ resembles $I_{1,1}$, as the states in Model 1 capture the action of Model 2 in the *Event type* variable. However, as there are important differences between the models, the impact values are likely to be different. $I_{2,1}$ suffers from the aforementioned drawback of not distinguishing successful from unsuccessful involvements. Nevertheless, it resembles the

Table 5.1: Example of player ratings from a single game, grouped by involvement type. Total represents the aggregate of the N involvement types.

Player	Game	Total	Type 1	Type 2	...	Type N
1	1	0.7328	0.3513	0.0427	...	0.0715
2	1	0.6826	0.2871	0.0476	...	0.1142
3	1	0.7077	0.2511	0.0654	...	0.1585
4	1	0.0146	-0.0128	0.0313	...	0.0012
5	1	0.4219	0.1812	0.2811	...	-0.1821

impact function argued to give reliable results in Bjertnes et al. (2016), and is expected to yield fair values in the long run. $I_{2,2}$ incorporates the outcome of an involvement through the value of the next state. As actions and states are separated in Model 2, $I_{2,2}$ does not have the drawback of being dependent on the following player's involvement. For instance, for a player's successful pass, $V(s')$ will be the value of a state where the player's team has the ball. For a player's unsuccessful pass, $V(s')$ will be the value of a state where the opponent has the ball. The former case is likely to yield higher impact.

The impact values of all involvements performed by a player during a game can be aggregated, resulting in a total game impact value. This value can be interpreted as a player rating, where players with higher ratings were the better players in the game. Similarly, impact values of all involvements across a whole season can be aggregated and adjusted for minutes played, which can be interpreted as a per season player rating. The ratings per game and per season are considered data driven player performance measures, useful for comparing players. The aggregated player impacts can be grouped by the type of involvement, given by *Event type* in Model 1 and action in Model 2. An example of player ratings per game, grouped by involvement type is illustrated in Table 5.1. The overall impact per game is given by the column Total, which is the aggregate of all involvement types. The aggregate of all involvement types is the natural player performance measure and is proposed as a player rating. However, for other purposes, such as analyzing playing style or finding similar players, it can be useful to consider subsets of involvement types or direct comparison of individual involvement types. The set of ratings for each involvement type constitute a player performance profile for each player. As different players score differently across different involvements types, the performance profiles are different for each player.

Figure 5.1 illustrates the same sequence as described and illustrated in Section 4.4, modelled by Model 1. The value of each state is added to the figure, while the context variables are removed. Combined with Table 5.2, the figure illustrates how players involved in the sequence receive impact values, based on impact functions $I_{1,1}$ and $I_{1,2}$. Each state is given a number which describes the state's order in the sequence. This is to be able to refer to the correct states for the sake of illustration. Recall that the impact functions are negated for players in the away team. In the first state, the state value is -0.0044 , impact function $I_{1,1}$ awards Jonas Svensson the impact of 0.0044 . Using $I_{1,2}$, Svensson is given $-(-0.0066 - (-0.0044)) = 0.0022$. As the away team moves the ball closer to goal, the state values get more negative, and both $I_{1,2}$ and $I_{1,2}$ yield positive impacts. In the fifth state, however, $I_{1,2}$ yields a negative impact. After the ball is received in zone

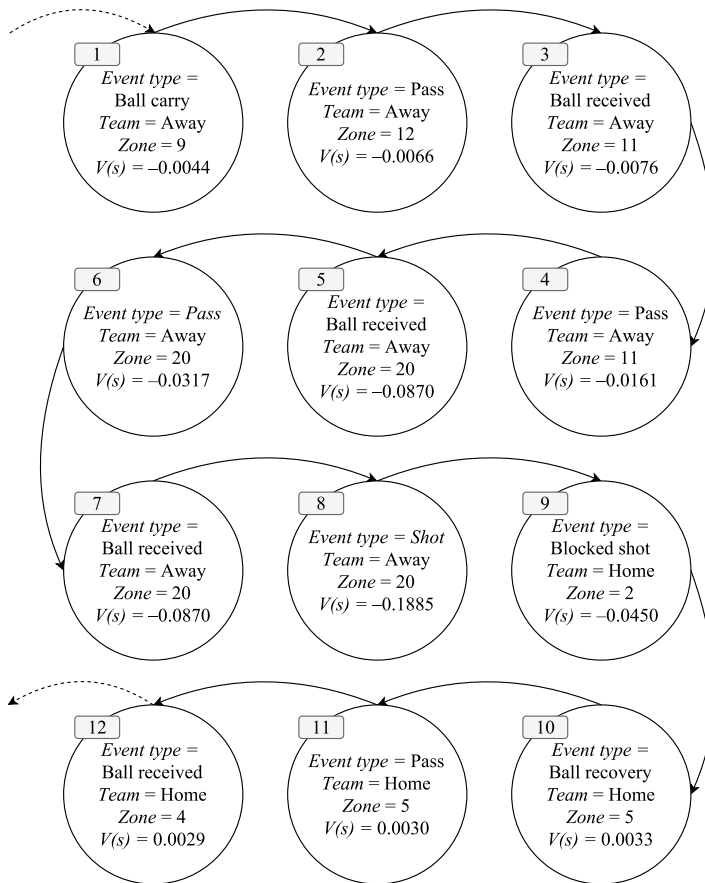


Figure 5.1: Sequence example with state values, modelled by Model 1.

20 by the away team, the following state has a less negative value, and Fredrik Midtsjø is given a negative impact by $I_{1,2}$. As the value of the fifth state is negative, $I_{1,1}$ yields positive impact. The impacts given after the eighth state show an important difference in the two impact functions. As shooting in zone 20 has high expected value, compared to the other states, Pål André Helland is given a large positive impact by $I_{1,1}$. $I_{1,2}$ however, takes the next state into account. Since Helland missed the shot, he is punished by the impact function, receiving an impact of $-(-0.0450 - (-0.1885)) = -0.1435$. If Helland had scored, he would have received an impact of $-(-1 - (-0.1885)) = 0.8115$, based on $I_{1,2}$, as the away goal state gives a reward of -1 . Helland would still receive the same impact of 0.1885 by $I_{1,1}$ if he had scored, as the function only considers the expected value of the current state.

Figure 5.2 and Table 5.3 illustrate how the sequence is modelled by Model 2 and how the impact functions $I_{2,1}$ and $I_{2,2}$ award players. The first impact is given to Jonas Svensson as he carries the ball from zone 9 to zone 12. Because Svensson plays for the

Table 5.2: Model 1 impact values from the sequence example. s is the state number from Figure 5.1

s	Event type	$V(s)$	Player	Team	$I_{1,1}$	$I_{1,2}$
1	Ball carry	-0.0044	Jonas Svensson	Away	0.0044	0.0022
2	Pass	-0.0066	Jonas Svensson	Away	0.0066	0.0011
3	Ball received	-0.0076	Christian Gytkjær	Away	0.0076	0.0085
4	Pass	-0.0161	Christian Gytkjær	Away	0.0161	0.0709
5	Ball received	-0.0870	Fredrik Midtsjø	Away	0.0870	-0.0553
6	Pass	-0.0317	Fredrik Midtsjø	Away	0.0317	0.0553
7	Ball received	-0.0870	Pål André Helland	Away	0.0870	0.1015
8	Shot	-0.1885	Pål André Helland	Away	0.1885	-0.1435
9	Blocked shot	-0.0450	Marius Amundsen	Home	-0.0450	0.0484
10	Ball recovery	0.0033	Bonke Innocent	Home	0.0033	-0.0004
11	Pass	0.0030	Bonke Innocent	Home	0.0030	-0.0001

away team, $I_{2,1}$ gives the negated value of the state-action-pair. Svensson receives an impact off -0.0007 . $I_{2,2}$ gives the value of the next state minus the value of the state-action-pair. For Svensson, this is $-(0.0003 - 0.0007) = 0.0010$. $I_{2,1}$ and $I_{2,2}$ calculates impact of the shot by Helland differently, as was also the case for the impact functions from Model 1. $I_{2,2}$ captures that the shot is missed, and punishes the player by giving negative impact. $I_{2,1}$ gives the expected value of shooting in state 20, which is positive.

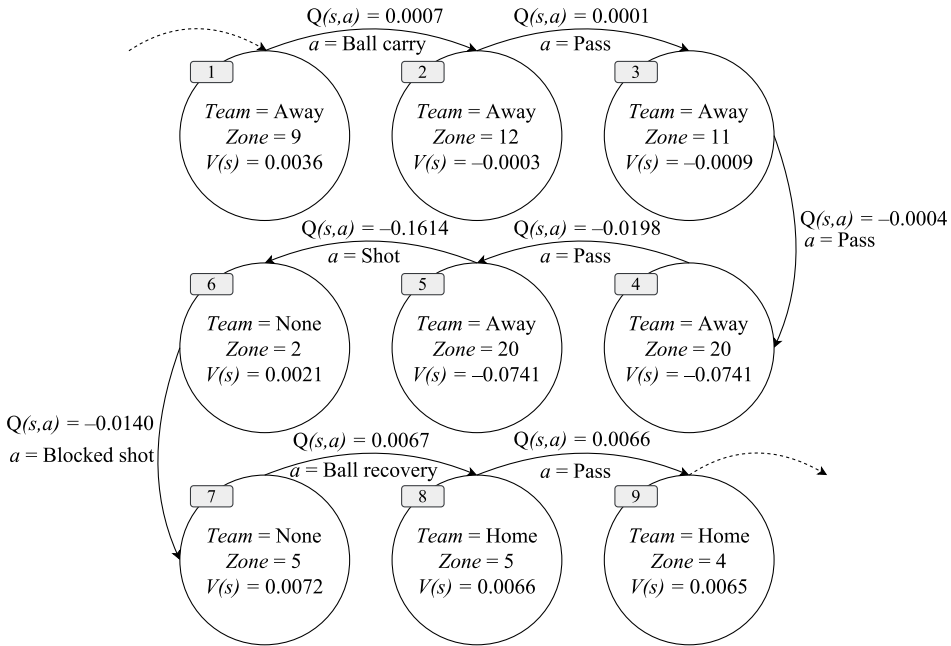
**Figure 5.2:** Sequence example with state and state-action values, modelled by Model 2.

Table 5.3: Model 2 impact values from the sequence example. s is the state number from Figure 5.2

s	$V(s)$	a	$Q(s, a)$	<i>Player</i>	<i>Team</i>	$I_{2,1}$	$I_{2,2}$
1	0.0036	Ball carry	0.0007	Jonas Svensson	Away	-0.0007	0.0010
2	-0.0003	Pass	0.0001	Jonas Svensson	Away	-0.0001	0.0010
3	-0.0009	Pass	-0.0004	Christian Gytkjær	Away	0.0004	0.0738
4	-0.0741	Pass	-0.0198	Fredrik Midtsjø	Away	0.0198	0.0543
5	-0.0741	Shot	-0.1614	Pål André Helland	Away	0.1614	-0.1634
6	0.0021	Blocked shot	-0.0140	Marius Amundsen	Home	-0.0140	0.0212
7	0.0072	Ball recovery	0.0067	Bonke Innocent	Home	0.0067	-0.0001
8	0.0066	Pass	0.0066	Bonke Innocent	Home	0.0066	-0.0002

5.2 Validation Methods

The purpose of the player evaluation is to objectively and accurately rate the players in Eliteserien. Player ratings are derived from the impact functions $I_{1,1}$, $I_{1,2}$, $I_{2,1}$ and $I_{2,2}$, introduced in Section 5.1. Three different methods for validating the player ratings are presented in this section. k -fold cross-validation is presented first, followed by correlation with benchmark ratings and inter-season correlations.

5.2.1 k -fold Cross-Validation

The first step in the validation process is k -fold cross-validation, introduced in Section 2.4. $k = 10$ folds is used. For each player rating, derived from the impact functions, 10-fold cross-validation is performed. The models are supposed to award the best players with the best ratings. Consequently, the likelihoods of the possible outcomes of each game should depend on the average ratings of the players involved for each team, because the better team is expected to have a better chance of winning. In a game between two teams, A and B, team A's probability of winning is expected to be higher if the average rating of their players is higher than that of team B, compared to if teams A and B had equally rated players. This is tested through 10-fold cross-validation, where the 720 games from Eliteserien seasons 2014, 2015 and 2016 are randomly split into 10 equally sized folds.

Ordinal logistic regression, introduced in Section 2.3, is applied to the training data with the number of outcomes $C = 3$. The dependent variable, Y , takes the value Y_i in game i , and represents the outcome of game i defined as

$$Y_i = \begin{cases} 1 & \text{if the outcome is away victory in game } i \\ 2 & \text{if the outcome is draw in game } i \\ 3 & \text{if the outcome is home victory in game } i \end{cases} \quad (5.5)$$

Following the notation introduced in Section 2.3, the cumulative probabilities are defined as

$$\gamma^{(j)} = \begin{cases} P(Y \leq 1) = \pi^{(1)} & \text{for } j = 1 \\ P(Y \leq 2) = \pi^{(1)} + \pi^{(2)} & \text{for } j = 2 \\ P(Y \leq 3) = 1 & \text{for } j = 3 \end{cases} \quad (5.6)$$

Table 5.4: Structure of the ordinal regression input data.

Game	Outcome	Rating difference
Game 1	Y_1	X_1
Game 2	Y_2	X_2
\vdots	\vdots	\vdots
Game n	Y_n	X_n

One independent variable is used in the regression, denoted X . This variable measures the difference in player ratings of the two participating teams. In each of the 10 cross-validation iterations, each players' average rating per game is computed over the 9 folds in the training data, i.e. the players' average impact per game over 90% of the games. Players only receive ratings in the games which they start, i.e. players substituted in during a game do not receive ratings for this specific game. For each game in the training data, let Z_H be the average rating of the ten outfield players starting the game for the home team, and Z_A be the average rating of the ten outfield players starting the game for the away team. $X = Z_H - Z_A$ is used in the regression, meaning X is the difference in average player rating of the two teams. $X > 0$ indicates that the home team has the better players, $X < 0$ indicates that the away team has the better players and $X = 0$ indicates that the two teams have equally good players in a given game. The structure of the regression input data is illustrated in Table 5.4.

An ordinal logistic regression is run on the 9 folds in the training data in each of the 10 cross-validation iterations. Hence, a regression model is fitted to 648 games and tested on the remaining 72 games in every iteration. Every game belongs to the test fold exactly once and belongs to the training set exactly nine times. Recall Equation 2.14

$$\log \left(\frac{\gamma_i^{(j)}}{1 - \gamma_i^{(j)}} \right) = \log \left(\frac{P(Y_i \leq j)}{P(Y_i > j)} \right) = \alpha^{(j)} - \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} \quad (2.14)$$

In the case of one independent variable, Equation 2.14 can be written as

$$\log \left(\frac{\gamma_i^{(j)}}{1 - \gamma_i^{(j)}} \right) = \log \left(\frac{P(Y_i \leq j)}{P(Y_i > j)} \right) = \alpha^{(j)} - \beta X_i \quad (5.7)$$

The two intercept terms, α_1 and α_2 , and the variable coefficient, β , are estimated from Equation 5.7, using the training data. Recall the definition of the cumulative distribution functions, given by Equation 2.15, expressed as $\gamma^{(j)}$ for $j = 1, 2$

$$\gamma^{(j)} = P(Y \leq j) = \frac{\exp(\alpha^{(j)} - (\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k))}{1 + \exp(\alpha^{(j)} - (\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k))} \quad (2.15)$$

which becomes

$$\gamma^{(j)} = P(Y \leq j) = \frac{\exp(\alpha^{(j)} - \beta X)}{1 + \exp(\alpha^{(j)} - \beta X)} \quad (5.8)$$

with one independent variable. The outcome probabilities, $P(Y = 1) = \pi^{(1)}$, $P(Y = 2) = \pi^{(2)}$ and $P(Y = 3) = \pi^{(3)}$, are then given by

$$\pi^{(1)} = \gamma^{(1)} \tag{5.9}$$

$$\pi^{(2)} = \gamma^{(2)} - \pi^{(1)} \tag{5.10}$$

$$\pi^{(3)} = 1 - \pi^{(1)} - \pi^{(2)} \tag{5.11}$$

Using α_1 , α_2 and β , the probabilities of each outcome of the games in the test fold is computed in each cross-validation iteration, based on the players starting the game for each team. As mentioned, X represents the difference in player ratings obtained in the training data. The test seeks to identify the predictive power of previously observed player ratings on the outcomes of unobserved games. If the player ratings actually rate player performance, the teams with the highest rated players should win the most games, and this should be reflected by the predicted outcome probabilities.

The predictive power is evaluated by computing the average Brier score, described in Section 2.5, over all predictions. For each player rating, every game is in the test set exactly once. Consequently, 720 predictions are made for each rating, and evaluated by the Brier score. The Brier scores of the different player rating methods can be compared, where the lower Brier score indicates that the player ratings better predict game outcomes. This in turn indicates that the rating with the lower Brier score more accurately measures player performance. The Brier scores of the player ratings are compared to two reference predictions, using the Brier skill score, introduced in Section 2.5. The first reference prediction is based on the observed probabilities of the outcome being home victory, away victory or draw, calculated from the actual frequencies of home victories, away victories and draws in the 720 Eliteserien games. The second reference prediction is based on the implied probabilities of betting odds for each game.

5.2.2 Benchmark Correlation

The correlations of the data driven player ratings with benchmark player ratings are investigated. A player's average impact per 90 minutes within a specific season is considered the player's rating for this season, as defined in Section 5.1. Four proposed ratings are computed for each player in every season, one based on each impact function. The correlations between the benchmark ratings and the data driven ratings are computed for every season in the data. Two benchmarks are chosen, namely Altomfotball (Altomfotball, 2014, 2015, 2016) and VG (VG, 2014, 2015, 2016). Altomfotball is a website owned by the Norwegian broadcaster TV2, and presents statistics from most major football leagues, including player ratings from Eliteserien. VG is the most read Norwegian newspaper, and player ratings from Eliteserien are available at their website. Both sources present average player ratings per season. Altomfotball's and VG's ratings are set by journalists, and are not to be considered as a definitive. As the goal is to develop objective ratings, the goal is not perfect correlation with journalist ratings. Nevertheless, a positive correlation would indicate that the data driven ratings and the journalists ratings agree to some extent. Definitive conclusions on the quality of the player ratings should not be based on correlations with journalist ratings alone. However, the correlations may be useful when used in conjunction with other validation metrics.

5.2.3 Inter-Season Correlation

The correlation of player ratings across successive seasons is useful for validating the player ratings. This correlation measures the similarity of each player's ratings in two seasons. A player's per season rating is given by the player's average impact per 90 minutes, as defined in Section 5.1. For each data driven rating, based on the four impact functions, the inter-season correlation is calculated. As data from three seasons is used, two pairs of successive seasons exist, the first being 2014 and 2015, and the second being 2015 and 2016. Valid player evaluations are believed to have high inter-season correlation. This will imply that the performance of most players do not change much between two successive seasons. The quality of some players will improve and the quality of some players will decline. However, the quality, hence the performance, of most players are expected to be consistent from one season to the next.

5.3 Identifying Similar Players

Recall from Section 5.1 that the aggregated results from the player performance evaluations result in ratings for each involvement type, which constitute a performance profile for each player. As such, the involvement types form a multidimensional space, and each performance profile can be represented by a point in this space. Different players will score differently across the different involvement types, but the idea is that more similar players have more similar performance profiles. To identify similar players, the k -Nearest Neighbours algorithm, introduced in Section 2.6, is used to calculate the Euclidean distances between each point in the multidimensional space. As the scale of the different involvement type ratings varies, the ratings have to be normalized before running the algorithm. A lower distance between two player performance profiles, indicates higher similarity between the two players. Looking at each player performance profile as a different class, the algorithm will identify the k most similar players for each player.

5.4 Team Impact on Player Performance

The procedure for investigating the impact of a player's team on the player's performance is described in this section. The ratings of individual players are expected to be influenced by the players' respective teams. A player rating is meant to rate a player's performance, but player performance does not necessarily equal player quality. For instance, two equally good players playing for different teams may perform differently because the quality of the other players in the respective teams is different. This is tested by looking at the players who have played for more than one team in Eliteserien, throughout seasons 2014, 2015 and 2016. The following values are computed for each such player

- P_a : The player's average rating while playing for team a .
- P_b : The player's average rating while playing for team b .
- T_a : The average rating of the other players in team a while the player played for team a .

- T_b : The average rating of the other players in team b while the player played for team b .

Per game average ratings are used in the above variables. $\Delta P_{a,b}$ is expressed as

$$\Delta P_{a,b} = P_a - P_b \quad (5.12)$$

and $\Delta T_{a,b}$ as

$$\Delta T_{a,b} = T_a - T_b \quad (5.13)$$

$\Delta P_{a,b}$ is interpreted as the difference in player performance of player p in the two teams a and b . $\Delta P_{a,b} > 0$ indicates that the player performed better in team a , $\Delta P_{a,b} < 0$ indicates that the player performed better in team b , and $\Delta P_{a,b} = 0$ indicates that the player performed equally well in both teams. $\Delta T_{a,b}$ is interpreted as the difference in average player performance between teams a and b , indicating the difference in team strength. $\Delta T_{a,b} > 0$ indicates that team a is the better team, $\Delta T_{a,b} < 0$ indicates that team b is the better team, and $\Delta T_{a,b} = 0$ indicates that the teams are equally good.

The correlation of $\Delta P_{a,b}$ and $\Delta T_{a,b}$ explains if the two values tend to move in the same direction. Positive correlation indicates that players tend to perform better for better teams, negative correlation indicates that players tend to perform worse for better teams, and correlation close to zero indicates that player performance is independent of team quality. Furthermore, a linear regression with $\Delta P_{a,b}$ as the independent variable and $\Delta T_{a,b}$ as the dependent variable, may provide further insight into the magnitude of the effect of team quality on individual player performances. The magnitude of such an effect, if it exists, is useful when scouting players from different teams. It is then possible to estimate the expected performance of a player in a new team, based on the quality of the two teams and the player's performance in the previous team.

As team quality might influence the performance of players, it is important to ensure that individual player performances are actually measured rather than team performances. This is investigated by considering the performances of the players who have played for the same team in two consecutive seasons. The following values are computed for each such player for each pair of consecutive seasons

- P_1 : The player's average rating in the first season.
- P_2 : The player's average rating in the second season.
- T_1 : The average rating of the other players in the player's team in the first season.
- T_2 : The average rating of the other players in the player's team in the second season.

In the remainder of this section, T_1 and T_2 is referred to as team performance. Per game ratings are averaged when computing P_1 , P_2 , T_1 and T_2 . A player's performance, compared to his team's performance in seasons 1 and 2, is denoted S_1 and S_2 , defined as

$$S_1 = P_1 - T_1 \quad (5.14)$$

and

$$S_2 = P_2 - T_2 \quad (5.15)$$

A positive difference indicates the player performed better than the team average, a negative difference indicates the player performed worse than the team average, and zero difference indicates the player performed similar to the team average.

Players' performances compared to their team's performances are expected to be consistent between seasons. Some deviation is natural, but the best players in a team in one season are generally expected to be the best also in the following season. The same relationship is expected for average and below average players. This is tested by computing the correlation of S_1 and S_2 . Positive correlation indicates that individual player performance is consistent, compared to the team's performance, between seasons. A correlation close to zero indicates that there is no tendency for player performance to stay consistent, compared to team performance, between seasons. Negative correlation indicates that above average players in one season tend to perform below average in the following season, and vice versa. In turn, zero or negative correlation indicates that teams are rated rather than individual players. This would mean that each team is rated, while the ratings are split somewhat randomly between each teams' players. As such, a positive correlation is expected, as it would mean that the individual player performances are actually evaluated.

Evaluating Results

The results of the experiment, described in Chapter 5, are presented and discussed in this chapter. The validation of the four proposed player ratings is presented in the first section. Based on the validation results, the four impact functions are narrowed down to one. The chosen impact function is then used to produce player ratings, and the top ten rated players in each position are presented in the following section. Next, the results of identifying similar players is presented and discussed, followed by results and discussion of team's impact on player performance. The chapter ends with a discussion of the results in light of the research questions, presented in Section 1.2.

6.1 Validation

The four player ratings are validated in this section. The k -fold cross-validation results are presented and discussed first, followed by correlations with benchmark ratings and the inter-season correlations. Impact functions are rejected throughout the validation process, justified by the validation results. Lastly, the overall validity of the ratings are summarized, and the most appropriate impact function is chosen to produce the results for the remainder of the thesis.

6.1.1 k -fold Cross-Validation

Four player ratings are proposed, based on impact functions $I_{1,1}$, $I_{1,2}$, $I_{2,1}$ and $I_{2,2}$, presented in Section 5.1. Recall that $I_{1,1}$ and $I_{1,2}$ are based on Model 1, and $I_{2,1}$ and $I_{2,2}$ are based on Model 2. The player ratings are validated using k -fold cross-validation with $k = 10$, following the setup in Section 5.2.1. Ten sets of regression coefficients, $\alpha^{(1)}$, $\alpha^{(2)}$, β , are computed for each player rating, one set for each of the ten folds. Table 6.1 summarizes the regression results. Each player rating is represented by its underlying impact function, and the average coefficient values are shown along with the average and max p -values of each coefficient. For complete regression results, including all coefficients and p -values, see Appendix B.

Table 6.1: k -fold cross-validation results represented by coefficients and p -values.

Impact function	Average coefficient values			Average p -values			Max p -values		
	$\alpha^{(1)}$	$\alpha^{(2)}$	β	$\alpha^{(1)}$	$\alpha^{(2)}$	β	$\alpha^{(1)}$	$\alpha^{(2)}$	β
I_{11}	-0.941	0.233	-3.535	0.000	0.008	0.000	0.000	0.015	0.000
I_{12}	-0.865	0.137	1.955	0.000	0.097	0.445	0.000	0.193	0.831
I_{21}	-0.898	0.273	-6.488	0.000	0.002	0.000	0.000	0.005	0.000
I_{22}	-0.948	0.168	-16.098	0.000	0.053	0.000	0.000	0.127	0.000

Recall that β is the independent variable coefficient, and that the independent variable, X , represents the difference in average player ratings of the two teams in a specific match. $X > 0$ means the home team has higher rated players, $X < 0$ means the away team has higher rated players and $X = 0$ means the teams have equally rated players. The p -value of β explain the statistical significance between the difference in player ratings and the outcome of the game. Following the equations in Section 5.2.1, a negative β is expected, because that would mean having higher rated players increase the chance of winning. The threshold coefficients, $\alpha^{(1)}$ and $\alpha^{(2)}$, capture the base probabilities of the three possible match outcomes. The home field advantage in football is well documented, see for instance Carmichael and Thomas (2005) and Bjertnes et al. (2015). All else equal, a home victory is always the most likely outcome in a a match between two equally good teams. β is considered the most important coefficient when considering the quality of the rating methods, but the threshold coefficients are important in the prediction step of k -fold cross-validation.

$I_{1,1}$ yields coefficients significant at $p < 0.05$ and β significant at $p < 0.01$ across all folds. This indicates a relationship between the player ratings and match outcomes. $I_{1,2}$ results in less significant coefficients, and β is not significant at $p < 0.10$ in any of the folds. The average β coefficient is positive, meaning that having the higher rated players decrease the chance of winning. This suggests the ratings based on $I_{1,2}$ poorly describe player performances. $I_{2,1}$ results in all coefficients being significant at $p < 0.01$, suggesting that the player ratings help explain the match outcomes. For $I_{2,2}$, β is found significant at $p < 0.01$. This suggests a relationship between the player ratings and match outcomes. However, the significance of $\alpha^{(2)}$, considering $I_{2,2}$, varies. As mentioned, low significance of the threshold terms may result in less accurate predictions, despite the information provided by X being useful.

The regression coefficients are used to predict the outcomes of the matches in the test fold, and the accuracies of the predictions are measured by the Brier score. Table 6.2 shows the Brier scores of the player ratings, represented by the impact functions, and of the reference predictions, denoted R_1 and R_2 . R_1 is based on the outcome frequencies. Table 6.3 shows the outcome frequencies, and the implied probabilities, from the 720 matches in the data set. The implied probabilities are used in R_1 's prediction. The second reference prediction, R_2 , is made from the underlying probabilities of betting odds. The underlying probabilities from the betting odds are normalized so that the sum equals one. Historical odds are found at BetExplorer (BetExplorer, 2014, 2015, 2016), which provide average odds from most online odds providers. The Brier scores of the player ratings are compared to the Brier scores of the reference predictions, using the Brier skill score. The Brier skill scores are shown in Table 6.4.

Table 6.2: Brier scores of the predictions made from the impact functions and the two reference predictions.

Predictor	Brier score
$I_{1,1}$	0.195
$I_{1,2}$	0.213
$I_{2,1}$	0.196
$I_{2,2}$	0.220
R_1	0.213
R_2	0.197

Table 6.3: Match outcome distribution of the predicted matches.

	Total	Home victories	Draws	Away victories
No.	720	336	171	213
Probability	1.000	0.467	0.236	0.296

A perfect predictor will achieve a Brier score of 0. This is unrealistic when predicting the outcome of football matches, as chance plays a big role (Anderson and Sally, 2014). If no information about the involved teams is known, R_1 would make the best predictions in the long run. In reality, however, the better teams are expected to have a better chance of winning, adjusted for the home field advantage. As such, a predictor based on an acceptable player rating system should outperform R_1 . This is not the case for the predictors based on $I_{1,2}$ and $I_{2,2}$. It is not surprising that $I_{1,2}$ performs poorly, as the regression coefficients had high p -values. $I_{1,2}$ is rejected as a measure of player impact suitable for rating player performance. $I_{2,2}$ is also rejected, due to the poor prediction. However, the low significance of $\alpha^{(2)}$ may be the reason of the poor prediction, rather than $I_{2,2}$ being a poor impact measure. More data may result in a more reliable threshold estimate, allowing for more accurate predictions, based on $I_{2,2}$ ratings. Nevertheless, with the data available, $I_{1,1}$ and $I_{2,1}$ seem to be more valid measures of player impact, thus more suitable for evaluating player performances. $I_{1,1}$ and $I_{2,1}$ achieve similar Brier scores, and the Brier scores are analogous to R_2 . Predictor performance close to that of betting odds strengthens the validity of the player ratings, based on $I_{1,1}$ and $I_{2,1}$, as the bookmaker is dependent on accurate probability estimates to survive.

Table 6.4: Brier skill scores of the impact functions against the two reference predictions.

	R_1	R_2
$I_{1,1}$	0.085	0.010
$I_{1,2}$	0.000	-0.081
$I_{2,1}$	0.080	0.005
$I_{2,2}$	-0.033	-0.117

6.1.2 Benchmark Correlation

In this section, the correlations between the data driven player ratings and journalists' player ratings are presented and discussed. Impact functions $I_{1,2}$ and $I_{2,2}$ were rejected in the k -fold cross validation, and only $I_{1,1}$ and $I_{2,1}$ are considered from this point forward. Tables 6.5, 6.6 and 6.7 show the correlations between the different player ratings in seasons 2014, 2015 and 2016, respectively. The correlation between $I_{1,1}$ and $I_{2,1}$ is close to 1 in all three seasons, meaning the resemblance between the two ratings is high. The correlations with VG and Altomfotball (AOF) vary, but is positive for both $I_{1,1}$ and $I_{2,1}$ every season. Journalist ratings are subject to cognitive bias, and the goal is to construct a data driven player evaluations which tend towards objectivity. Under the assumption that journalist ratings are not perfect, the goal is not perfect correlations. However, positive correlations show that the data driven player evaluations agree to some extent with the journalists, and is considered to strengthen the validity of the data driven ratings. Based on the correlations with VG and AOF, no conclusion is made on whether $I_{1,1}$ or $I_{2,1}$ produce the most valid ratings.

Table 6.5: Benchmark correlation of impact function $I_{1,1}$ and $I_{2,1}$, season 2014.

	$I_{1,1}$	$I_{2,1}$	VG	AOF
$I_{1,1}$	1.000			
$I_{2,1}$	0.968	1.000		
VG	0.413	0.456	1.000	
AOF	0.434	0.466	0.803	1.000

Table 6.6: Benchmark correlation of impact function $I_{1,1}$ and $I_{2,1}$, season 2015.

	$I_{1,1}$	$I_{2,1}$	VG	AOF
$I_{1,1}$	1.000			
$I_{2,1}$	0.958	1.000		
VG	0.567	0.607	1.000	
AOF	0.507	0.531	0.878	1.000

Table 6.7: Benchmark correlation of impact function $I_{1,1}$ and $I_{2,1}$, season 2016.

	$I_{1,1}$	$I_{2,1}$	VG	AOF
$I_{1,1}$	1.000			
$I_{2,1}$	0.959	1.000		
VG	0.261	0.282	1.000	
AOF	0.294	0.351	0.801	1.000

6.1.3 Inter-Season Correlation

The inter-season correlations of the ratings based on impact functions $I_{1,1}$ and $I_{2,1}$ are discussed in this section. A positive correlation means that players tend to be rated similarly in the two seasons, and a negative correlation means that players tend to be rated differently. Similar ratings in two seasons mean that, if the actual player performance quality is captured by the rating, players perform at a similar level in the two seasons. Player performance is expected to be fairly stable across seasons, meaning that a high inter-season correlation is expected.

Table 6.8: Inter-season correlations of impact functions $I_{1,1}$ and $I_{2,1}$ and the benchmark ratings.

	2014/2015 Correlation	2015/2016 Correlation
$I_{1,1}$	0.868	0.874
$I_{2,1}$	0.766	0.758
VG	0.607	0.668
AOF	0.547	0.629

Table 6.8 shows the inter-season correlations for player ratings based on $I_{1,1}$, $I_{2,1}$, VG and AOF. The journalist ratings are included as a reference. $I_{1,1}$ and $I_{2,1}$ have higher inter-season correlations than VG and AOF, thus the data driven ratings seem more stable. Journalist ratings are based on opinions, and different journalists may have rated a specific player in the different games which might explain some of the variation. The high inter-season correlations of the data driven ratings is considered to strengthen their validity. Even though $I_{1,1}$ has higher inter-season correlation than $I_{2,1}$, the resemblance is too high to conclude which impact function produces the most accurate player ratings, based solely on inter-season correlations.

6.1.4 Overall Validity Assessment

An overall validity assessment of the ratings based on the four impact functions, $I_{1,1}$, $I_{1,2}$, $I_{2,1}$ and $I_{2,2}$, is presented in this section. The ratings based on $I_{1,2}$ and $I_{2,2}$ were rejected as appropriate player ratings in Section 6.1.1, and thus not investigated further in Section 6.1.2 and 6.1.3. Ratings based on both $I_{1,1}$ and $I_{2,1}$ were positively correlated with journalist ratings, and had high inter-season correlations. It is hard to separate impact functions $I_{1,1}$ and $I_{2,1}$, based on the applied validation methods. However, as argued in Section 4.3.1, Model 2 resembles the flow of football games in a more accurate and intuitive way, compared to Model 1. Furthermore, one measure of player performance is sufficient to answer the research questions presented in Section 1.2. Consequently, $I_{2,1}$, which is based on Model 2, is used to produce the results presented in the remaining sections.

6.2 Player Ratings

The highest rated players in Eliteserien 2016, evaluated by impact function $I_{2,1} = Q(s, a)$, is presented in this section. Player ratings are calculated as the sum of the impact of every involvement by a player throughout the season, normalized per 90 minutes played. As such, a player's rating is interpreted as the player's average impact per 90 minutes. Players are grouped by their most played position, using the five groups central defenders, full backs, central midfielders, wingers and strikers. The players' total ratings are displayed, as well as a subset of the involvements types in the player performance profiles. The displayed subset of involvement types is based on which involvements are believed most relevant for each position.

Table 6.9: Top 10 rated central defenders in Eliteserien 2016.

Player	Team	Total	Pass	Long pass	Ball carry	Tackle	Aerial duel	Clearance	Blocked shot
G. Valsvik	Strømsgodset	0.4846	0.2823	0.0631	0.0895	-0.0077	0.0076	-0.0249	-0.0101
T. Reginiussen	Rosenborg	0.4761	0.2745	0.0228	0.0722	-0.0035	0.0025	-0.0180	-0.0092
F. Semb Berge	Odd	0.4470	0.2503	0.0590	0.1022	-0.0021	-0.0022	-0.0167	-0.0112
M. Høibraten	Strømsgodset	0.4394	0.3315	0.0297	0.0768	0.0005	0.0018	-0.0147	-0.0054
S. Hagen	Odd	0.4137	0.2305	0.0612	0.1229	-0.0053	-0.0007	-0.0224	-0.0089
S. Larsen	Vålerenga	0.3996	0.1850	0.0402	0.0595	-0.0022	0.0006	-0.0136	-0.0157
O. Heieren Hansen	Sarpsborg 08	0.3964	0.1953	0.0447	0.0429	-0.0001	0.0084	0.0076	-0.0088
J. Gundersen	Tromsø	0.3931	0.1614	0.0371	0.0666	-0.0091	0.0062	-0.0046	-0.0106
J. Grønner	Brann	0.3316	0.1637	0.0723	0.0625	-0.0008	0.0004	-0.0089	-0.0218
H. O. Eyjolfsson	Rosenborg	0.3172	0.2068	0.0145	0.0779	-0.0125	0.0009	-0.0090	-0.0084
Top 10 average		0.4045	0.2228	0.0452	0.0723	-0.0049	0.0027	-0.0105	-0.0107

Table 6.9 lists the ten highest rated central defenders. Pass, long pass and ball carry are the involvement types which contribute the most to the total rating for all the listed players. Intuitively, central defenders are expected to be good at defending actions, such as tackles and blocking shots. However, most of the listed players have negative ratings in these involvement types. Careful consideration of the impact function $I_{2,1} = Q(s, a)$, helps explain this result. A negative rating indicates that the involvements are expected to lead to more goals against than for the player's team. When a defender performs a defensive involvement, the ball is usually closer to his goal than the opponent's. Intuitively, this makes a goal against more likely than a goal for, in turn resulting in negative impact. The impact function seems to measure the offensive contribution of an involvement, rather than the defensive. Incorporating the involvement's effect on the opponent's chance of scoring might yield impact values for defensive involvements more comparable to those of offensive involvements.

Concerning central defenders, avoiding goals against is generally considered more important than contributing to goals for. The top three teams in 2016, RBK, Odd and Brann, are represented by five of the top ten central defenders. As these teams conceded the least goals in Eliteserien 2016, this result is not surprising. However, as shown, offensive involvements contribute more to the ratings than defensive involvements. Thus, the proposed ratings might be a less accurate measure of performance for central defenders, compared to offensive players. Central defenders in stronger teams are likely to be more offensively involved than central defenders in weaker teams, which might explain why the central defenders in stronger teams are higher rated. Superior teams generally have more time on the ball, given by possession percentage, than weaker teams (Bjertnes et al.,

2015; Oberstone, 2009), meaning that the players can make a higher number of offensive involvements, in turn receiving higher ratings.

Table 6.10: Top 10 rated full backs in Eliteserien 2016.

Player	Team	Total	Pass	Long pass	Crossing	Ball carry	Take on	Tackle	Shot
E. Ruud	Odd	0.7328	0.3513	0.0427	0.0658	0.0917	0.0035	-0.0081	0.0715
A. Gersbach	Rosenborg	0.7290	0.2763	0.0411	0.0794	0.1451	0.0423	-0.0038	0.0265
P. E. Flo	Molde	0.7077	0.3651	0.0270	0.0493	0.0796	0.0108	-0.0114	0.0847
J. Parr	Strømsgodset	0.6826	0.2871	0.0199	0.0729	0.0939	0.0096	-0.0079	0.1284
R. Lindkvist	Vålerenga	0.6736	0.2688	0.0146	0.0849	0.0753	0.0102	-0.0113	0.1100
J. V. Nilsen	Odd	0.6557	0.3669	0.0307	0.0217	0.0932	0.0065	-0.0123	0.0911
K. A. Antonsen	Tromsø	0.6385	0.2016	0.0476	0.0543	0.0658	0.0110	-0.0105	0.0939
T. Grøgaard	Odd	0.6197	0.3181	0.0145	0.0464	0.0485	0.0073	-0.0049	0.0283
J. Svensson	Rosenborg	0.6114	0.3113	0.0379	0.0432	0.1100	0.0121	-0.0191	0.0656
L.C. Vilstvik	Strømsgodset	0.6057	0.3089	0.0301	0.0643	0.0695	0.0136	-0.0010	0.0731
Top 10 average		0.6657	0.3055	0.0306	0.0582	0.0872	0.0127	-0.0091	0.0773

Table 6.10 lists the ten highest rated full backs. Pass is the involvement type contributing most to the total ratings, while the ratings in the remaining involvement types appear to vary among the listed players. RBK and Odd are represented by two and three players each, and Strømsgodset have both of their preferred full backs on the list. All players have negative tackle ratings, as was the case for central defenders. As previously argued, this is because the impact of tackles seem to be the offensive, rather than the defensive contribution of the involvement. Defensive involvements make up an important part of full backs' duties, but full backs are generally expected to contribute more offensively than central defenders. As such, the player ratings are likely a better representation of the performances of full backs than that of central defenders, and particularly useful for assessing the offensive capabilities of full backs.

Table 6.11: Top 10 rated central midfielders in Eliteserien 2016.

Player	Team	Total	Shot	Pass	Long pass	Crossing	Ball carry	Take on	Tackle
M. Jensen	Rosenborg	1.0289	0.3167	0.3837	0.0381	0.0312	0.1274	0.0093	0.0018
G. Thorarinnsson	Rosenborg	0.9092	0.1270	0.4785	0.0487	0.0242	0.1460	0.0086	-0.0025
B. Boateng	Strømsgodset	0.8591	0.1735	0.4630	0.0359	0.0097	0.0988	0.0055	0.0004
A. Trondsen	Sarpsborg 08	0.8566	0.1042	0.3981	0.0501	0.0173	0.1552	0.0169	0.0036
A. Konradsen	Rosenborg	0.8456	0.2057	0.3565	0.0616	0.0055	0.1207	0.0227	-0.0045
F. Midtsjø	Rosenborg	0.8291	0.1381	0.3740	0.0285	0.0372	0.1712	0.0341	0.0004
K. Barmen	Brann	0.8236	0.1477	0.4215	0.0399	0.0153	0.0950	0.0015	-0.0018
J. Samuelsen	Odd	0.8083	0.1671	0.3971	0.0474	0.0199	0.1382	0.0137	-0.0051
H. Singh	Molde	0.7938	0.1770	0.3965	0.0735	0.0187	0.1321	0.0096	-0.0051
M. Abu	Strømsgodset	0.7889	0.0336	0.5402	0.0847	0.0106	0.1302	0.0039	-0.0072
Top 10 average		0.8543	0.1591	0.4209	0.0508	0.0190	0.1315	0.0126	-0.0020

Table 6.11 lists the ten highest rated central midfielders. Pass is the dominant contributing involvement type towards the total ratings. This is unsurprising, as passing is generally considered one of the most important skills of a central midfielder. Jensen has the highest shot rating, noticeably higher than Konradsen, the second highest rated shooter. Jensen was the highest scoring central midfielder in 2016, scoring eight goals, followed by Singh with five and Konradsen with four. Unsurprisingly, the number of goals seem to influence the shot ratings. The highest rated passer is Abu, who together with Thorarinnsson and Boateng appear noticeably higher rated than the rest. An interesting observation is that

the three Rosenborg midfielders Jensen, Thorarinnsson and Midtsjø are the highest rated crossers, indicating that they push wide more often than most central midfielders. Rosenborg dominate the central midfielder list, while no other team has more than one player among the top ten. It is worth mentioning that the central midfielder with the most assists in 2016, Haugen from Brann with eight, is the eleventh highest rated, merely missing the top ten list.

The average total ratings of the highest rated central midfielders are higher than those of the defenders. This is most likely due to midfielders operating further up the pitch. Intuitively, goals are more likely in more offensive positions. As such, a more offensively positioned player is likely to receive higher impact values for his involvements, compared to a more defensive player. The total ratings must be interpreted accordingly, and an obvious interpretation is that a midfielder can not be said to have performed better than a defender, based on the players' total ratings. The reasonable application of the ratings appear to be comparison of players in similar positions and roles.

Table 6.12: Top 10 rated wingers in Eliteserien 2016.

Player	Team	Total	Shot	Headed shot	Pass	Long pass	Crossing	Ball carry	Take on
P. A. Helland	Rosenborg	1.2226	0.4058	0.0117	0.3175	0.0251	0.0787	0.2130	0.0367
M. Keita	Strømsgodset/Stabæk	1.1794	0.2414	0.0064	0.4486	0.0565	0.0655	0.2011	0.0444
G. Zahid	Vålerenga	1.0734	0.2888	0.0138	0.4599	0.0235	0.0166	0.1829	0.0466
Y. E. de Lanlay	Rosenborg	0.9987	0.3646	0.0059	0.2883	0.0141	0.0664	0.1898	0.0412
M. Elyounoussi	Molde	0.9926	0.4392	0.0849	0.2835	0.0147	0.0238	0.1227	0.0335
S. Adegbenro	Viking	0.9851	0.3107	0.0681	0.2310	0.0068	0.0481	0.1926	0.0832
T. Nguen	Strømsgodset	0.9602	0.2852	0.0344	0.3147	0.0078	0.0206	0.1978	0.0477
Ø. Storflor	Strømsgodset	0.9396	0.1524	0.0000	0.4346	0.0331	0.0701	0.1025	0.0264
A. Sigurdarson	Tromsø	0.9091	0.3657	0.0127	0.2390	0.0118	0.0521	0.1692	0.0271
G. Koomson	Sogndal	0.8824	0.2162	0.0135	0.3142	0.0169	0.0466	0.1597	0.0447
Top 10 average		1.0143	0.3070	0.0251	0.3331	0.0210	0.0489	0.1731	0.0432

Table 6.12 lists the ten highest rated wingers. Shot, pass and ball carry contribute the most to the total ratings. Helland has the highest total rating, regardless of position, with shots as the most contributing involvement type. Keita and Zahid, who were the most scoring wingers in 2016 with seven and eight goals, respectively, are ranked second and third. Elyounoussi has the highest headed shot rating, followed by Adegbenro, both rated noticeably higher than the rest. Koomson, the player with most assists in 2016, is only the sixth highest rated passer among the top ten.

The wingers appear to generally have higher total ratings than central midfielders, strengthening the point that overall performance comparisons should not be made of players in different positions. However, for more detailed studies of isolated involvement types, such comparisons may be appropriate. For instance, central midfielder Jensen is noticeably higher rated on shots than Storflor, implying that he impacts games more in terms of shots. A natural assumption is then that Jensen is expected to score more than Storflor, supported by their eight and four respective goals in 2016.

Table 6.13 lists the top ten rated strikers. Unsurprisingly, shot is the most contributing involvement type to the total ratings. Some strikers also have high headed shot ratings, such as Gytkjær, Viljálmsón, Moussa and Otoo. Otoo is the highest rated striker, with nearly half of his total coming from shots, and pass being the second highest contributor. This result is surprising, as Otoo only scored three goals in 2016. 97 players scored at least as many goals as Otoo, and 70 players scored more. Only one of the five strikers who

Table 6.13: Top 10 rated strikers in Eliteserien 2016

Player	Team	Total	Shot	Headed shot	Pass	Long pass	Crossing	Ball carry	Take on
M. Otoo	Sogndal	1.2146	0.5507	0.1002	0.3089	0.0215	0.0153	0.1470	0.0200
F. Friday	Lillestrøm	1.0555	0.4939	0.0482	0.2640	0.0097	0.0098	0.1513	0.0411
S. Abdullahi	Viking	0.8221	0.3990	0.0328	0.2133	0.0023	0.0368	0.0967	0.0349
O. Omojuanfo	Stabæk	0.8212	0.2060	0.0569	0.3240	0.0101	0.0218	0.1326	0.0220
M. Vilhjálmsson	Rosenborg	0.8108	0.1230	0.1390	0.3845	0.0251	0.0046	0.0693	0.0060
S. Svendsen	Molde	0.8039	0.2850	0.0436	0.2884	0.0189	0.0188	0.0870	0.0163
C. Gytkjær	Rosenborg	0.8018	0.3235	0.1820	0.1983	0.0072	0.0023	0.0422	0.0041
D. Brown	Vålerenga	0.7993	0.4606	0.0608	0.1546	0.0087	0.0076	0.0495	0.0128
P. A. Kirkevold	Sarpsborg 08	0.7761	0.2735	0.0587	0.3037	0.0094	0.0103	0.0860	0.0225
S. Moussa	Tromsø	0.7764	0.2967	0.1013	0.2728	0.0105	0.0003	0.0649	0.1340
Top 10 average		0.8680	0.3394	0.0722	0.2686	0.0120	0.0157	0.1018	0.0240

scored at least 10 goals in 2016, Gytkjær, appear on the top ten list, as Mos Abdellaoue, Azemi, Ocean and Agdestein are left out. As scoring goals is considered an important task for strikers, this is a questionable result. Because the impact function, $I_{2,1} = Q(s, a)$, does not incorporate the outcome of an involvement, players can receive a high shot rating, regardless of goals scored. If a player attempts many shots where the associated value $Q(s, a)$ is high, his impact from shots will be high. This in turn results in high shot ratings, regardless of the shots ending in a goal or not. As $Q(s, a)$ represents the expected value of the action a in state s , it can be argued that Otoo were unlucky to score only three goals, and that the number of goals will get closer to the expected number in the long run. However, some players may need fewer shots per goal, which might explain why some of the highest scoring strikers are not the highest rated. Shooting skills, and better evaluations of when to shoot, is likely to influence the number of shots per goal. As the presented shot ratings largely depend on the number of shots, this is not accounted for. $I_{2,2}$ would capture this better, compared to $I_{2,1}$, as illustrated in Section 5.1, but this impact function was rejected in Section 6.1.1. However, there is more to the striker role than scoring goals. As such, the presented ratings might provide a valid overall evaluation of strikers, taking the impact of other involvements than shots and goals into account.

The validation, presented in Section 6.1, indicate acceptable accuracy of the ratings presented in this section. However, as discussed, offensive involvements seem more fairly rewarded than defensive involvements. Furthermore, players in more offensive positions are higher rated than defensive players. Thus, the ratings appear most applicable for comparing players in similar positions. It is worth mentioning that defensive contributions can be important also for players in offensive positions, something which is not captured well by the presented ratings. Consequently, an assessment of defensive contributions could be considered as a supplement to the presented ratings. In the presented results, most teams are represented by at least one player, meaning that players are recognized regardless of team. However, in line with general expectations, the better teams have more highly rated players than the weaker teams.

6.3 Similar Players

The k -Nearest Neighbours (k NN) algorithm, introduced in Section 2.6, is used to find the $k = 10$ most similar players to the outfield players in RBK's preferred starting eleven

in Eliteserien 2016. $k = 10$ is chosen, which seems to result in an appropriate number of similar players. If the range of the distances is small, a smaller k could mean that similar players are not listed. On the other hand, the similarity diminishes with a larger k . Although only RBK players are presented, the k most similar players can easily be obtained for any outfield player in Eliteserien. Recall that a player performance profile consists of a player's rating for each involvement type. Matches from the 2016 season is used in the calculation of the player performance profiles, and only players with more than 900 minutes played are included to ensure that the performance profiles are stabilized. The performance profiles are calculated per 90 minutes played, representing the per season ratings of the players. Furthermore, the scale of the different involvement types varies. Therefore, the ratings for each involvement type are normalized, in order to avoid certain involvement types dominating the distances.

Table 6.14: Explanation of position abbreviations

Abbreviation	Position
CD	Central Defender
FB	Full Back
CM	Central Midfielder
W	Winger
ST	Striker

The tables in this section presents the Euclidean distances between the performance profile of the player in the first row and the performance profile of the player listed in a respective row. All players are listed with their most played position. The position abbreviations are explained in Table 6.14, and the positions are illustrated in Figure C.1 in Appendix C. A short Euclidean distance means that the similarity of the compared performance profiles is high, and a long Euclidean distance means that the similarity is low. In turn, a high similarity means that two players are rated similarly across all the involvement types in the performance profiles. Both the distances and the range of the distances vary for the different players, meaning some players have more unique performance profiles. For players that are much better or worse than the k most similar players, the distances are higher. Players who are rated similarly across all involvement types are likely to resemble each other, meaning that one could potentially be a replacement of the other. Being able to identify similar players is important in the process of scouting replacement players when a player has been sold or gets injured. The presented approach is able to identify similar players, however, replacing a player with a better player is usually considered ideal. The same set of involvement types may be the highest rated within two players' performance profiles, but if one is sufficiently higher rated than the other, this approach will not consider the players similar. Consequently, a generally considered similar, but better player may not be identified. Nevertheless, the presented similarity analysis may provide valuable information, useful as an initial screening of replacement player candidates.

Table 6.15 presents the ten most similar players to RBK's two central defenders Tore Reginiussen and Hólmar Örn Eyjólfsson. The k NN-algorithm does not take playing position into account, and the lists in Table 6.15 consist of both defenders and midfielders. Some of the listed players are similar to both Reginiussen and Eyjólfsson, and Eyjólfsson is

Table 6.15: 10 most similar players to RBK's most playing central defenders season 2016.

Tore Reginiussen	CD	Distance	Hólmar Örn Eyjólfsson	CD	Distance
Peter Orry Larsen	CM	0.760	Christian Landu Landu	CM	0.892
Marius Høibraten	CD	0.768	Matti Lund Nielsen	CM	0.903
Marius Lundemo	CM	0.883	Ruben Gabrielsen	CD	0.909
Vegard Skjerve	CD	0.916	Jostein Gundersen	CD	0.911
Ruben Gabrielsen	CD	0.927	Vegard Skjerve	CD	0.922
Hólmar Örn Eyjólfsson	CD	0.997	Eirik Birkelund	CM	0.923
Christophe Psyche	CD	1.001	Oliver Berg	CM	0.928
Joachim Jørgensen	CD	1.031	Abdisalam Ibrahim	CM	0.943
Enar Jääger	CD	1.034	Luc Kassi	CM	0.945
Claes Phillip Kronberg	FB	1.054	Christophe Psyche	CD	0.946

among the ten players most similar to Reginiussen. As pointed out in Section 6.2, offensive involvements seem more fairly rewarded than defensive involvements. This is not ideal for finding similar players to central defenders, as defensive duties are considered important for such players. Nevertheless, 11 of 20 players similar to Eyjólfsson and Reginiussen are central defenders. As RBK dominated Eliteserien in 2016, most teams probably had a defensive approach when playing against RBK. In turn, RBK's central defenders are allowed to push further forward than central defenders on most other teams. This might help explain why Eyjólfsson and Reginiussen are found similar to central midfielders, as well as defenders. RBK's central defenders are likely to often operate in a similar area of the pitch to that of central midfielders on weaker teams, leading to ratings similar to midfielders. Overall, the results do not seem very intuitive for RBK's two central defenders, and the proposed approach do not seem to identify replacement players for central defenders.

Table 6.16: 10 most similar players to RBK's most playing full backs season 2016.

Jonas Svensson	FB	Distance	Jørgen Skjelvik	FB	Distance
Hans Norbye	FB	1.189	Henri Toivomäki	FB	0.928
Bjørn Helge Riise	CM	1.195	Claes Phillip Kronberg	FB	0.943
Ruben Kristiansen	FB	1.195	Thomas Jacobsen	FB	0.961
Jørgen Skjelvik	FB	1.223	Magnus Andersen	W	1.029
Henri Toivomäki	FB	1.282	Marius Lundemo	CM	1.041
Amin Nouri	FB	1.339	Abdisalam Ibrahim	CM	1.130
Emil Jonassen	FB	1.369	Vegard Skjerve	CD	1.139
Henrik Robstad	FB	1.391	Henrik Furebotn	CM	1.153
Jukka Raitala	FB	1.392	Riku Riski	W	1.169
Joachim Thomassen	FB	1.404	Lars Christian Kjemhus	W	1.179

Table 6.16 presents the ten most similar players to the two full backs Jonas Svensson and Jørgen Skjelvik. Although Svensson moved to the dutch club AZ Alkmaar in January 2017, and Skjelvik currently plays centre back, these two players are presented as they were the regular full backs in 2016. The range of the distances to the most similar players are different for the two players. For Skjelvik, the range goes from 0.928 to 1.179, while the range is 1.189 to 1.404 for Svensson. This indicates that it may be more difficult to find a player similar to Svensson in Eliteserien, than to find one similar to Skjelvik. Another observation is that few of the players similar to Svensson are similar to Skjelvik, although Skjelvik is among Svensson's most similar players. Whereas most players sim-

ilar to Svensson are full backs, players in several different positions are found similar to Skjelvik. Skjelvik played several matches as central defender and winger in 2016. Thus, Skjelvik's performance profile is not expected to represent a typical full back. This could be a reason why Skjelvik is found similar to wingers as well as central defenders and full backs.

Table 6.17: 10 most similar players to RBK's most playing central midfielders season 2016.

Mike Jensen	CM	Distance	Anders Konradsen	CM	Distance	Fredrik Midtsjø	CM	Distance
Fredrik Haugen	CM	1.103	Harmeet Singh	CM	1.192	Mohamed Ofkir	W	0.738
Gudmundur Thorarinsson	CM	1.190	Jone Samuelsen	CM	1.198	Fredrik Nordkvælle	CM	0.981
Herman Stengel	CM	1.242	Bismark Boateng	CM	1.250	Ghayas Zahid	W	0.982
Anders Trondsen	CM	1.282	Fredrik Aursnes	CM	1.260	Mattias Moström	W	1.016
Eirik Hestad	CM	1.314	Björn Sverrisson	CM	1.290	Francisco Junior	CM	1.045
Øyvind Storfloer	W	1.380	Daniel Fredheim Holm	W	1.319	Rafik Zekhnini	W	1.048
Christian Grindheim	CM	1.386	Mathias Antonsen Normann	CM	1.376	Ohikhuæme Omoijuanfo	ST	1.052
Petter Strand	W	1.407	Abdisalam Ibrahim	CM	1.421	Jone Samuelsen	CM	1.083
Gilbert Koomson	W	1.490	Sander Berge	CM	1.458	Chukwuma Akabueze	W	1.099
Jonas Lindberg	ST	1.508	Matti Lund Nielsen	CM	1.483	Tokmac Nguen	W	1.120

Table 6.17 presents the k NN results for RBK's central midfield trio Mike Jensen, Anders Konradsen and Fredrik Midtsjø. The trio played regularly in 2016 and has continued to do so in the first half of the 2017 season. Of the 30 players listed in the table, there are 29 different players, and none of the three RBK players are on the lists of the other two. This indicates that the RBK midfielders do not resemble each other. As all three players were among the top ten rated central midfielders listed in Table 6.11, the lack of similarity between the three RBK midfielders is unlikely to occur because of a difference in the quality of their total performance, but rather due to having different qualities. Midtsjø is found similar to several wingers, indicating, as mentioned in Section 6.2, that he tends to push wide. Konradsen has a more defensive role than Midtsjø and Jensen in RBK's 4-3-3 formation, and he appears to be similar to more CMs than the other two. The list of Konradsen's similar players largely consists of CMs playing in 4-4-2 systems, arguably a similar role to Konradsen, or players in the same 4-3-3 role. Even though Jensen and Midtsjø are found similar to wingers, several of the players categorized as wingers are wide midfielders in 4-4-2 systems. The difference between 4-4-2 and 4-3-3, and the difference of the RBK midfielders' roles, are illustrated in Appendix C. Arguably, the roles of Midtsjø and Jensen resemble the wide midfielder role more than the CM role in 4-4-2 systems. As such, the lists of similar players can be argued more accurate than suggested by the displayed playing positions, and may be useful for identifying replacement players.

Table 6.18 lists RBK's forward line, represented by Yann-Erik de Lanlay, Christian Gytkjær and Pål André Helland, and the ten most similar players to each. As Gytkjær played striker and the other two are wingers, the results are expected to differ. The calculated distances from Pål André Helland to the the most similar players are high, ranging from 1.542 to 2.086. The tables in Section 6.2 show that Helland achieved the highest overall rating in 2016. As Helland accordingly receives high ratings in several involvement types, one would expect the distances to the most similar players to be high. Furthermore, it appears that the player performance profiles largely separate strikers and wingers from players in other positions, as Gytkjær is only found similar to strikers, and de Lanlay and Helland are mostly similar to wingers.

Table 6.18: 10 most similar players to RBK's most playing forwards season 2016.

Yann-Erik de Lanlay	W	Distance	Christian Gytkjær	ST	Distance	Pål André Helland	W	Distance
Aron Sigurdarson	W	1.012	Fitim Azemi	ST	1.121	Gilbert Koomson	W	1.542
Erik Huseklepp	W	1.104	Torbjørn Agdestein	ST	1.379	Muhamed Keita	W	1.594
Mohamed Ofkir	W	1.105	Patrick Mortensen	ST	1.394	Øyvind Storflo	W	1.637
Tokmac Nguen	W	1.152	Agon Mehmeti	ST	1.421	Mike Jensen	CM	1.696
Edwin Gyasi	W	1.174	Mohammed Abdellaoue	ST	1.440	Petter Strand	W	1.824
Chukwuma Akabueze	W	1.180	Mathias Bringaker	ST	1.449	Deyver Vega	W	1.978
Rafik Zekhnini	W	1.185	Jakob Orlov	ST	1.468	Anders Trondsen	CM	1.995
Fredrik Midtsjø	CM	1.285	Deshorn Brown	ST	1.470	Aron Sigurdarson	W	2.053
Mohammed Elyounoussi	W	1.380	Runar Espejord	ST	1.490	Yann-Erik de Lanlay	W	2.063
Simen Junkerød	W	1.400	Marcus Pedersen	ST	1.496	Erik Huseklepp	W	2.086

As seen in this section, the k NN algorithm for identifying similar players seems a promising approach. The results regarding central defenders, however, are questionable, as RBK's central defenders were found similar to several midfielders. The k NN algorithm may produce better results for central defenders if defensive involvements were evaluated differently. As discussed, Svensson is found similar to other full backs, indicating that replacement full backs may be successfully identified. Furthermore, the results regarding RBK's midfielders and offensive trio help establish the impression that the approach is applicable for finding similar players. However, the similarity for players with defensive duties could be supported by an assessment of defensive capabilities for more detailed comparisons.

6.4 Team Impact on Player Performance

A player's performance is expected to depend on the quality of the player's team. Even though the better teams are generally expected to have the better players, there might be a complimentary effect when a player plays alongside higher quality players. A better team is likely to create more goals and have a higher possession percentage (Bjertnes et al., 2015; Oberstone, 2009), resulting in player ratings that are dependent on team quality. This is investigated by considering the players who have played for more than one club in Eliteserien throughout 2014, 2015 and 2016. Only players who have played more than 450 minutes for each team are considered. 63 players have played for two different teams, and three players have played for three different teams. For each pair of teams, team a and team b , for which player p has played, $\Delta P_{a,b}$ and $\Delta T_{a,b}$ are computed, as explained in Section 5.4. Recall that $\Delta P_{a,b}$ represents the difference in player p 's average rating while playing in teams a and b , and that $\Delta T_{a,b}$ represents the difference in average ratings of the other players in teams a and b when player p played for each of the two teams.

Figure 6.1 shows the plot of $\Delta P_{a,b}$ against $\Delta T_{a,b}$ for every pair of clubs a and b related to each of the 66 players. The correlation between $\Delta P_{a,b}$ and $\Delta T_{a,b}$ is 0.6238. The positive correlation implies that players tend to receive higher ratings when playing for a team where the average rating of the other players is higher, and vice versa. The interpretation of this is that players tend to perform better when playing for a stronger team, and perform worse when playing for a weaker team. Table 6.19 shows the results of a linear regression where $\Delta P_{a,b}$ is the dependent variable and $\Delta T_{a,b}$ is the only independent variable. The constant is not significant, which makes sense, logically. A player is not expected to perform better or worse in a different team without accounting for factors that

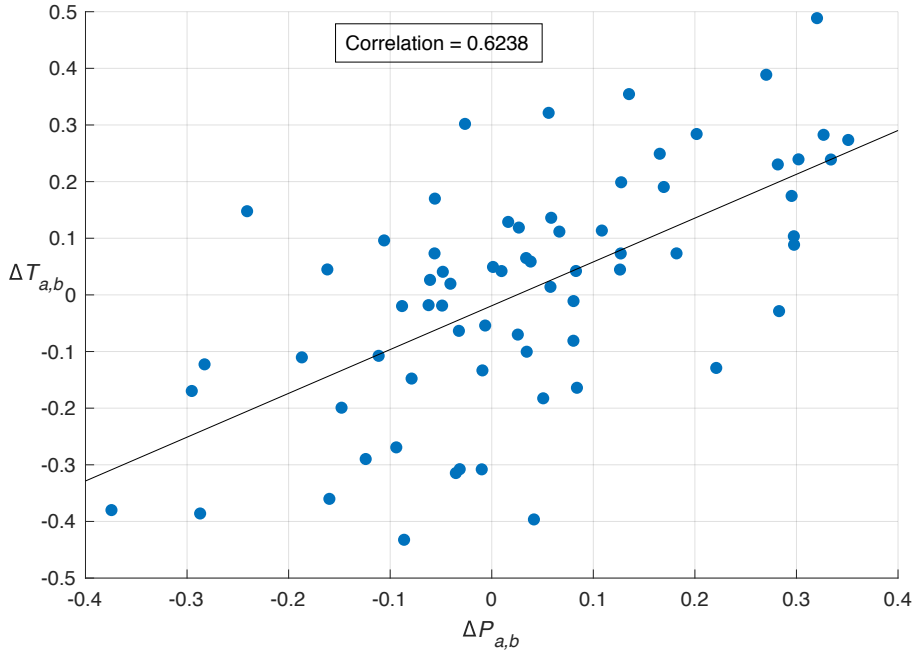


Figure 6.1: Scatter plot and correlation of change in player performance and change in team rating.

separate the two teams. The independent variable, however, is significant at $p < 0.01$. This indicates that team quality significantly influence player performance. The coefficient of 0.774 indicates that the player performance changes by 77.4% of the change in average team rating.

Table 6.19: Linear regression results using $\Delta P_{a,b}$ as the dependent and $\Delta T_{a,b}$ as the independent variable.

Variable	Coefficient	Std. error	<i>t</i> -value	<i>p</i> -value
Constant	-0.019	0.020	-0.951	0.345
$\beta_{\Delta T_{a,b}}$	0.774	0.118	6.533	0.000

Players' performances, compared to their team's performances, is discussed next, based on the approach explained in Section 5.4. The players who have played for the same team in two consecutive seasons during 2014, 2015 and 2016 are used to compute S_1 and S_2 . Recall that S_1 and S_2 measure player performance compared to team performance in season 1 and 2, respectively. 199 pairs of S_1 and S_2 values exist in the data, and a single player may be represented by two pairs of S_1 and S_2 values if he played for the same team in all three seasons.

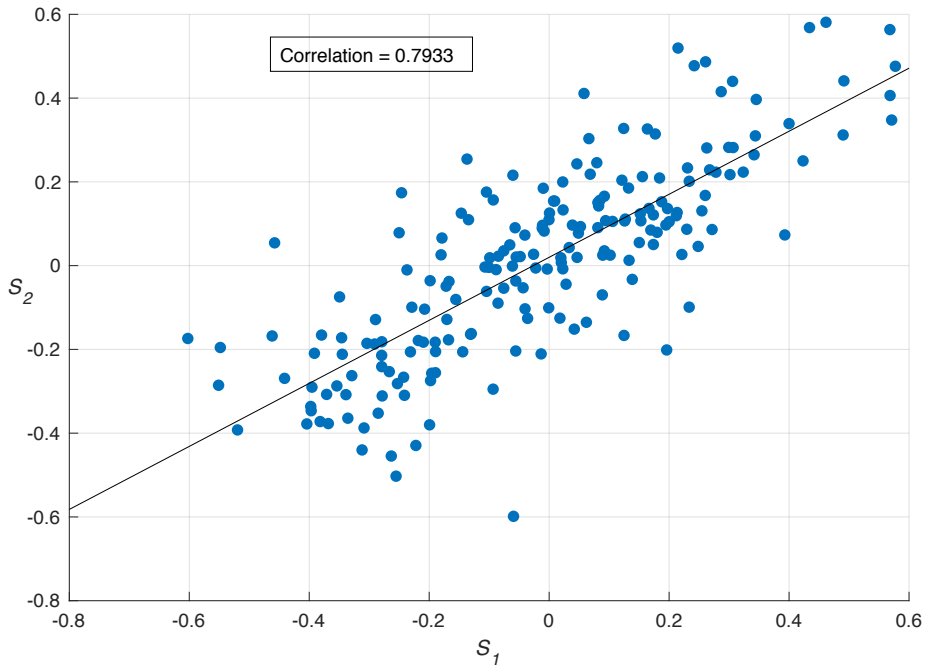


Figure 6.2: Scatter plot and correlation of player performance in one season, S_1 , against the player performance in the same team in the successive season, S_2 .

S_1 is plotted against S_2 in Figure 6.2 and the correlation between the two is 0.7933. This indicates that players perform consistently, compared to their team's performances, across seasons. As the team quality was shown to influence player performance, this is particularly important. It could be that the ratings evaluated teams rather than players, and that the team's ratings were split randomly between a team's players. It is expected that the same players are among the best within the teams from one season to another, and the positive correlation between S_1 and S_2 is a strong indication that this is reflected by the ratings. Hence, the data driven player evaluation appear to distinguish individual players, and therefore rate player performances rather than team performances.

In Section 6.2, the stronger teams were found to have more players on the top ten lists of most positions, compared to weaker teams. As shown in this section, this may be partly due to the complementary effect of playing with better players. Nevertheless, player ratings were shown to distinguish players within teams, and thus, comparing players from different teams is viable. A natural extension of the work is to attempt to adjust the player ratings for team strength, but that is reserved for further research, discussed in Chapter 8.

6.5 Evaluation of Research Questions

The research questions, presented in Section 1.2, are answered and evaluated in this section, based on the results presented in this chapter.

RQ1: Can outfield association football player performances be objectively and accurately evaluated by data driven player ratings, derived from Markov models?

Impact function $I_{2,1}$, presented in Section 5.1, was used to produce data driven player ratings. The impact function is based on Q -values from Markov Model 2, presented in Chapter 4. Ratings can be produced per game and per season, by aggregating and normalizing the impact of all involvements by each player. As the player ratings are based on Markov game Q -values, learned through value iteration, it is a fair to assume that the ratings tend towards objectivity. Furthermore, the ratings produced probability estimates similar to the underlying probabilities of betting odds, which indicates accurate player evaluation. The impression of accurate player evaluation was further established as the correlation with benchmark ratings and inter-season correlation corresponded with the expected results. However, as discussed in Section 6.2, the ratings appear most applicable for comparing players in similar positions. As goals are generally more likely in more offensive areas of the pitch, offensive players tend to receive higher ratings than defensive players. This could possibly be improved by looking into other ways of evaluating defensive involvements, as offensive involvements seem more heavily rewarded.

RQ2: Can data driven player evaluations be used to identify similar players?

The evaluation of player performances resulted in a unique player performance profile for each evaluated player. In order to identify similar players, the performance profiles were used as input in the k -Nearest Neighbours algorithm. Presuming that more similar players have more similar performance profiles, the approach presented in Section 5.3, identifies players who resemble each other. The results from the approach were presented in Section 6.3, which lists the ten most similar players to RBK players in every playing positions.

The results were promising for players in most positions. The approach managed to identify similar players in terms of playing position, but more interestingly, different players playing in the same position were found similar to different sets of players. This indicates that the approach manages to identify similar players in terms of qualities and roles, and may therefore be applicable for finding replacement players. However, a limitation is that two players who are highly rated in the same involvement types, where one player is sufficiently higher rated than the other player, might not be found similar. Consequently, players at a similar level, which also possess similar qualities, are generally identified in the analysis. The results appear less intuitive for central defenders than for offensive players. A possible explanation is that the player ratings are more influenced by offensive than defensive contributions. Although better defensive evaluations may improve the overall results, the approach seems to identify similar players and produce results applicable for replacing current players and recruiting new talent.

RQ3: Is there a relationship between team quality and individual player performances?

Using player ratings based on impact function $I_{2,1}$ as the measure of player performance, the relationship between team quality and individual player performances were investigated in Section 6.4. The performances of players who had played for at least two different teams, and the quality of these players' respective teams were considered. It was found that players tend to perform better while playing for a stronger team, and perform worse while playing for a weaker team. This indicates a complimentary effect of playing with better players. This is important for real-world applications, as it implies that the quality of the teams must be taken into account when scouting players. A player's rating might not stand out initially, but if the player's team is weaker than the potential new team, the performance is likely to improve for the new team. The magnitude of the relationship between team quality and player performance was identified by linear regression. This means that a reasonable estimate of a player's performance in a new team can be made, based on the previous performance of the player, the quality of the player's previous team and the quality of the new team.

Conclusion

This thesis has described the development of data driven player performance ratings, and potential applications of such ratings. Two Markov game models were developed by improving the models presented in Bjertnes et al. (2016). The players' ratings are based on the expected values of the players' individual involvements, obtained from the Markov models. Offensive involvements appear to be higher rewarded than defensive involvements. This results in lower ratings for defensive players, compared to offensive players. However, thorough validation indicates that the ratings accurately evaluate individual player performances, and that the ratings are believed to tend towards objectivity.

Cognitive bias may lead to badly informed decisions and undesired risks in the transfer market for professional football clubs. The presented, objective player evaluations eliminate cognitive bias, which may in turn reduce the risk associated with player transfers. Furthermore, players who are rated similarly across different involvement types are believed to resemble each other. Similar players, in terms of position, were identified, which is promising. More interestingly, however, different sets of players were found similar to different players in the same position, indicating that similar players in terms of individual qualities and roles are identified. The approach is limited to identifying similar players at a somewhat similar level, and seems to work better for more offensive players. Nonetheless, the results are promising and suggest that the approach can be useful in practice.

Objective evaluation, in combination with identification of similar players, can streamline the process of scouting players, resulting in lower costs and reduced risk. To further increase the usefulness of the player ratings, the impact of team quality on the ratings is investigated. Players are found to perform better while playing for better teams, and the magnitude of this effect is identified. Although not further investigated in this thesis, this allows for player ratings to be adjusted for team quality.

The work in this thesis, and the described approaches, can be used as a framework for professional clubs operating in the transfer market. A club can select a particular player and instantly receive his rating and identify his similar players. The ratings of the similar players can be adjusted for team quality to investigate the expected performance compared to the selected player. Only data from Eliteserien is used in this thesis, but similar data is available from at least 30 other divisions. Data from any division can be used to analyze

players from all over the world. Such a framework can be a valuable tool for Rosenborg Ballklub, the initiator of this thesis. The club operates in a highly competitive sport, and the smallest competitive advantage can be invaluable. The presented work extends the limited sports analytics research in Norway, which brings data driven tools closer to the operations of Norwegian clubs.

Recommendations for Further Research

This thesis explores how football modelled as Markov chains and Markov games can be utilized in data driven player performance evaluation. The Markov models presented in this thesis resemble the models introduced by Bjertnes et al. (2016). However, improvements and new ways to utilize the results are introduced. In addition to produce player ratings, the results from the player performance evaluation are used to find similar players. This represents a new application of Markov models in football, and further research in this direction may be worthwhile. Recommendations for further development of Markov models in football, and suggestion for better utilization of such models, are made in this chapter.

So far, the work on Markov models, used in player performance evaluation in football, has shown that the developed models do not sufficiently evaluate defensive performances. In this thesis, attempts were made to better evaluate defensive player involvements by seeking a more realistic modelling of the game flow. The results however, show that players are often not rewarded positively for performing defensive involvements. This suggests that adjustments have to be made to better evaluate such involvements. Using the same impact function for both offensive and defensive involvements does not seem to produce optimal results. Thus, further experiments regarding choice of impact functions should be conducted. In order to evaluate defensive performances more accurately, a separate impact function for evaluating defensive involvements should be considered.

Assuming future Markov models are based on event-based data only, there are still limitations for modelling the state contexts. However, the event-based Opta data contains information which is not included in the existing models. Every event in the data is registered with the angle the ball travels at during an event relative to the direction of play, presented in radians. This makes it possible to categorize the passes into different groups depending on the direction of the pass. Passes going forwards are different than passes going backwards or sideways on the pitch. Making a pass forwards in a given state may be more valuable than a pass backwards to the defender. Some impact functions may catch this difference by looking at the next state, but categorizing passes and long passes depending on the direction might improve the models.

The granularity of the models is increased by adding context variables or by increas-

ing the number of possible variable values, actions or event types. As the granularity increases, a more realistic modelling of the game may be achieved, but on the other hand, you may end up with a too large number of possible states and state-action pairs. To obtain accurate state and Q -values, the occurrences of each state and state-action pair have to be sufficiently high, in order to account for irregular outcomes and to give a reasonable representation of the real values. Hence, a trade-off between granularity and number of occurrences is introduced. It is possible to determine an optimal state definition and the optimal number of possible variable values. In this thesis, the pitch is divided into 21 different zones and four different time periods are used. The optimal numbers are not investigated, and it is not certain whether a larger or smaller number is beneficial, or how much a change will improve the results. This could be investigated by running cross-validations on different alternatives, similar to the approach presented in this thesis. If reducing the possible values of a context variable to the minimum does not change the results for the better, the variable should be considered removed from the state definition.

The analysis of similar players presented in this thesis is, to the best of the authors knowledge, the first analysis where Markov models in football are used to identify similar players. Similar players are identified entirely based on player involvements on the ball. Physical data could be included in the analysis to reveal further similarity between players. For instance, preferred foot, height, weight and age could be included to identify players which are physically similar to the player you want to replace. Number of passes in different zones, number of passes forwards, backwards and sideways may also be included to identify further similarity. Having tracking data available would make it possible to include movement patterns and distance covered per game in the analysis. Including physical and tracking data in the analysis may require an adjusted and customized implementation of the k NN-algorithm. Other distance measures has to be used to measure the similarity, if categorical or ordinal variables are included. For instance, it may be possible to develop a distance measure combining Euclidean distance and Jaccard similarity. A customized implementation also makes it possible to weigh each attribute differently, but it may be difficult to determine what gives the best results and highest level of total similarity.

The team impact on player performance was investigated in this thesis. A linear regression model was built, and the model makes it possible to predict how a player will perform in a different club, given the rating of the old and new club. This also makes it possible to adjust all player ratings as if all players played for the same club, and thus remove the team impact on the player ratings. Furthermore, it is possible to identify how each involvement type is affected by team quality. In this thesis, the impact on the total rating was identified, but different involvement types are likely to be affected differently. A player may perform better on passes in a better club, but his performances on tackles or aerial duels may be better for a weaker club. When the team impact is identified for each involvement type, these ratings can be adjusted according to the team impact for every player, resulting in player performance profiles which are team independent. In turn, these player performance profiles can be used to identify similar players, using the k NN algorithm. This approach may improve the similarity analysis and produce better results.

Bibliography

Altomfotball, 2014. [Online; accessed 27-April-2017].

URL <http://www.altomfotball.no/element.do?cmd=tournamentStatistics&tournamentId=1&seasonId=336&useFullUrl=false>

Altomfotball, 2015. [Online; accessed 27-April-2017].

URL <http://www.altomfotball.no/element.do?cmd=tournamentStatistics&tournamentId=1&seasonId=337&useFullUrl=false>

Altomfotball, 2016. [Online; accessed 27-April-2017].

URL <http://www.altomfotball.no/element.do?cmd=tournamentStatistics&tournamentId=1&seasonId=338&useFullUrl=false>

Anderson, C., Sally, D., 2014. *The Numbers Game*. Penguin Books Ltd.

Asaro, V., 2016. *Baseball Analysis Using Markov Chains*. California Polytechnic State University, San Luis Obispo.

Benoit, K., 2012. *Multinomial and Ordinal Logistic Regression*. Lecture notes ME104 Linear Regression Analysis, London School of Economics and Political Science, [Online; accessed 01-June-2017].

URL http://www.kenbenoit.net/courses/ME104/ME104_Day8_CatOrd.pdf

BetExplorer, 2014. [Online; accessed 01-May-2017].

URL <http://www.betexplorer.com/soccer/norway/tippeligaen-2014/>

BetExplorer, 2015. [Online; accessed 01-May-2017].

URL <http://www.betexplorer.com/soccer/norway/tippeligaen-2015/>

-
- BetExplorer, 2016. [Online; accessed 01-May-2017].
URL <http://www.betexplorer.com/soccer/norway/tippeligaen-2016/>
- Biggest Global Sports, 2017. A statistics-based analysis of the world's most popular sports. [Online; accessed 25-April-2017].
URL <http://www.biggestglobalsports.com/>
- Bjertnes, V., Nørstebø, O., Vabo, E., 2015. Identifying Key Performance Indicators in Norwegian Association Football. Tech. rep., Norwegian University of Science and Technology.
- Bjertnes, V., Nørstebø, O., Vabo, E., 2016. Valuing individual player involvements in norwegian association football. Master's thesis, Norwegian University of Science and Technology.
- Brooks, J., Kerr, M., Guttag, J., 2016. Developing a Data-Driven Player Ranking in Soccer Using Predictive Model Weights. In: Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16. ACM, New York, NY, USA, pp. 49–55.
URL <http://doi.acm.org/10.1145/2939672.2939695>
- Bukiet, B., Harold, E., Palacios, J., 1997. A Markov Chain Approach to Baseball. *Operations Research* 45 (1), 14–23.
- Caley, M., 2015. Premier League Projections and New Expected Goals. [Online; accessed 27-March-2017].
URL <http://cartilagefreecaptain.sbnation.com/2015/10/19/9295905/premier-league-projections-and-new-expected-goals>
- Caraffa, A., Cerulli, G., Projetti, M., Aisa, G., Rizzo, A., 1996. Prevention of anterior cruciate ligament injuries in soccer. *Knee Surg, Sports traumatol, Arthroscopy* 4 (10), 19–21.
- Carmichael, F., Thomas, D., 2005. Home-Field Effect and Team Performance. *Journal of Sports Economics* 6 (3), 264–281.
- Cochran, J. J., 2010. The Emergence of Sport Analytics. [Online; accessed 25-April-2017].
URL <http://analytics-magazine.org/the-emergence-of-sport-analytics/>
- Coleman, J., 2012. Identifying the "Players" in Sports Analytics Research. *Interfaces* 42 (2), 109–118.
- Davenport, T. S., 2014. Analytics in Sports: The New Science of Winning. Tech. rep., International Institute for Analytics.
- Geoffrey, J. M., Kim-Anh, D., Ambrose, C., 2004. Analyzing MicroAreray Gene Expression Data, 1st Edition. John Wiley & Sons, Inc, 111 River Street, Hoboken, NJ 07030.

-
- Hillier, F. S., Liebermann, J., 2015. Introduction to Operations Research, 10th Edition. McGraw-Hill Education, 2 Penn Plaza, New York, NY 10121.
- Hirotsu, N., Wright, M., 2002. Using a Markov Process Model of an Association Football Match to Determine the Optimal Timing of Substitution and Tactical Decisions. *The Journal of the Operational Research Society* 53, 88–99.
- Hirotsu, N., Wright, M., 2003a. Determining the Best Strategy for Changing the Configuration of a Football Team. *The Journal of the Operational Research Society* 54, 878–887.
- Hirotsu, N., Wright, M., 2003b. An Evaluation of Characteristics of Teams in Association Football by Using a Markov Process Model. *Journal of the Royal Statistical Society* 52, 591–602.
- Ingle, S., 2015. How Midtjylland took the analytical route towards the Champions League. *The Guardian*[Online; accessed 21-May-2017].
URL <https://www.theguardian.com/football/2015/jul/27/how-fc-midtjylland-analytical-route-champions-league-brentford-matthew-benham>
- Kvam, P., Sokol, J., 2006. A Logistic Regression/Markov Chain Model for NCAA Basketball. *Naval Research Logistics* 53, 788–803.
- Lago-Peñas, C. and Lago-Ballesteros, J., Dellal, A., Gómez, M., 2010. Game-Related Statistics that Discriminated Winning, Drawing and Losing Teams from the Spanish Soccer League. *Journal of Sports Science and Medicine* 9, 288–293.
- Lewis, T., 2014. How Analytics Took Over At Britain’s Top Football Clubs. *The Guardian*[Online; accessed 21-May-2017].
URL <https://www.theguardian.com/football/2014/mar/09/premier-league-football-clubs-computer-analysts-managers-data-winning>
- Littman, M. L., 1994. Markov Games as a Framework for Multi-agent Reinforcement Learning. Tech. rep., Brown University.
- Littman, M. L., 2001. Value-function reinforcement learning in Markov games. *Journal of Cognitive Systems Research* 2, 55–66.
- Marr, B., 2015. How Big Data and Analytics are Changing Soccer. [Online; accessed 25-April-2017].
URL <https://www.linkedin.com/pulse/how-big-data-analytics-changing-soccer-bernard-marr>
- Maxcy, J., Drayer, J., 2014. Sports Analytics: Advancing Decision Making Through Technology and Data. Tech. rep., Fox School of Business, Temple University.
- McHale, I. G., Scarf, P. A., Folker, D. E., 2012. On the Development of a Soccer Player Performance Rating System for the English Premier League. *Interfaces* 42 (4), 339–351.

-
- Mitchell, T. M., 1997. *Machine Learning*, 1st Edition. McGraw Hill Higher Education.
- Mohr, M., Krusturup, P., Bangsbo, J., 2011. Match performance of high-standard soccer players with special reference to development of fatigue. *Journal of Sports Sciences* 21 (7), 519–528.
- Neese, B., 2015. *Inside the Numbers: An Introduction to Sports Analytics*. Shorter University, [Online; accessed 22-May-2017].
URL <http://online.shorter.edu/sports-management/introduction-to-sports-analytics/>
- Nørregaard, L., Thorsø, L., Bangsbo, J., 1991. Activity profile of competition soccer. *Canadian Journal of Sports Sciences* 2 (16), 110–116.
- Oberstone, J., 2009. Differentiating the Top English Premier League Football Clubs from the Rest of the Pack: Identifying the Keys to Success. *Journal of Quantitative Analysis in Sports* 5 (10).
- Opta Sports, 2016. The data collection process. [Online; accessed 15-March-2017].
URL <http://www.optasports.com/about/how-we-do-it/the-data-collection-process.aspx>
- Peterson, L. E., 2009. [Online; accessed 01-May-2017].
URL http://scholarpedia.org/article/K-nearest_neighbor
- Poole, D. L., Mackworth, A. K., 2010. *Artificial Intelligence: Foundations of Computational Agents*. Cambridge University Press, 32 Avenue of the Americas, New York, NY, USA.
- Radicchi, E., Mozzachiodi, M., 2016. Social Talent Scouting: A New Opportunity for the Identification of Football Players? *The Journal of Josef Pilsudski University of Physical Education in Warsaw and International Society for the Social Sciences of Sport* 70, 28–43.
- Routley, K., 2015. *A Markov Game Model for Valuing Player Actions in Ice Hockey*. Master's thesis, Simon Fraser University.
- Shirley, K., 2007. A Markov Model for Basketball. Poster presentation at New England Symposium for Statistics in Sports, Boston, MA, [Online; accessed 22-May-2017].
URL http://www.nesis.org/nessis07/Kenny_Shirley.pdf
- Steinberg, L., 2015. *CHANGING THE GAME: The Rise of Sports Analytics*. [Online; accessed 14-March-2017].
URL <https://www.forbes.com/sites/leighsteinberg/2015/08/18/changing-the-game-the-rise-of-sports-analytics/#54cc24af4c1f>
- Strumbelj, E., Vracar, P., 2012. Simulating a basketball match with a homogeneous Markov model and forecasting the outcome. *International Journal of Forecasting* 28, 532–542.

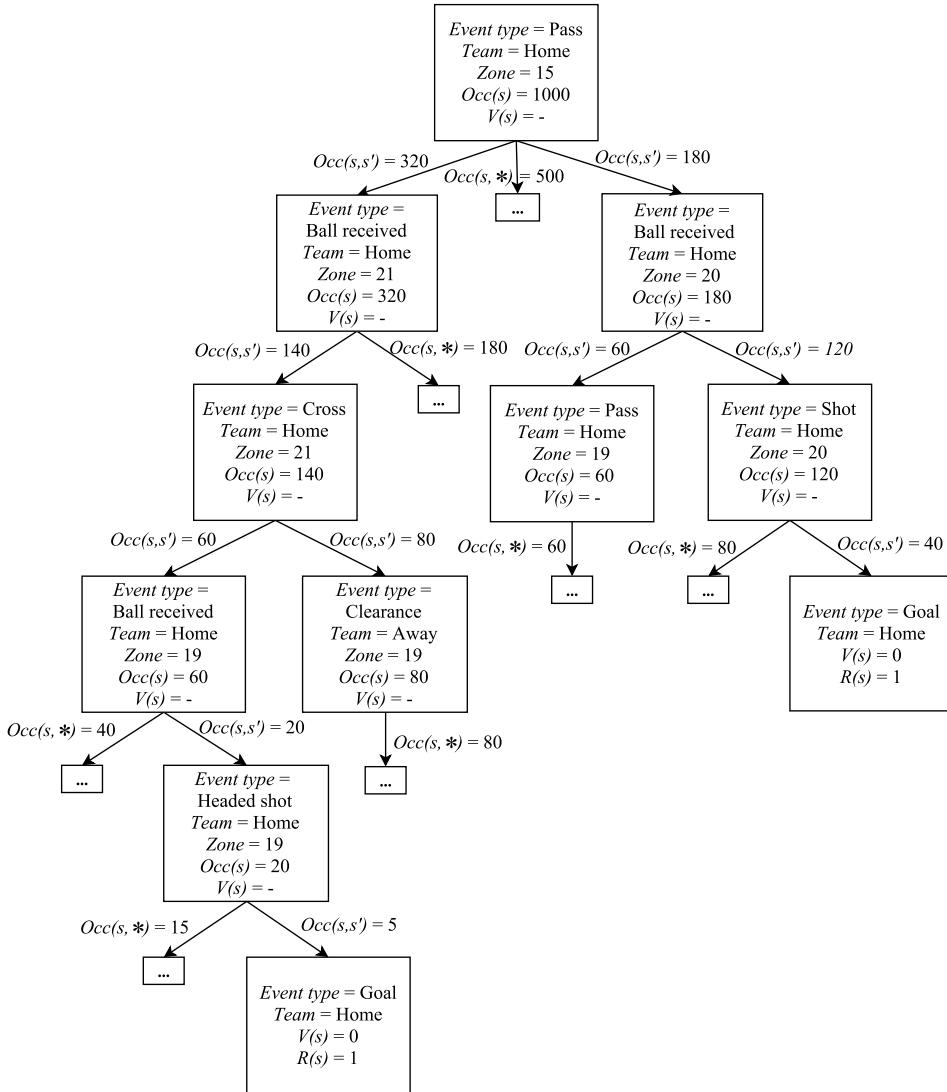
-
- Sæbø, O. D., 2015. Modelling clubs' financial investment in association football players. Master's thesis, Norwegian University of Science and Technology.
- Sæbø, O. D., Hvattum, L. M., 2015. Evaluating the efficiency of the association football transfer market using regression based player ratings. In: Norsk Informatikkonferanse 2015. pp. –.
URL <http://ojs.bibsys.no/index.php/NIK/article/view/246>
- The Football Forecast, 2016. Is Money Taking over Football? [Online; accessed 25-April-2017].
URL <https://thefootballforecast.com/2016/02/11/is-money-taking-over-football/>
- Total Sportek, 2017. 25 World's Most Popular Sports (Ranked by 13 factors). [Online; accessed 25-April-2017].
URL <http://www.totalsportek.com/most-popular-sports/>
- VG, 2014. [Online; accessed 27-April-2017].
URL <http://old.vglive.no/#frontpage=20170427&eliteguiden=s-borsen-sid-669>
- VG, 2015. [Online; accessed 27-April-2017].
URL <http://old.vglive.no/#frontpage=20170427&eliteguiden=s-borsen-sid-706>
- VG, 2016. [Online; accessed 27-April-2017].
URL <http://old.vglive.no/#frontpage=20170427&eliteguiden=s-borsen-sid-743>
- Wilks, D. S., 2011. Statistical Methods in the Atmospheric Sciences, 3rd Edition. Elsevier Inc., 225 Wyman Street, Waltham, MA 02451.

Appendix **A**

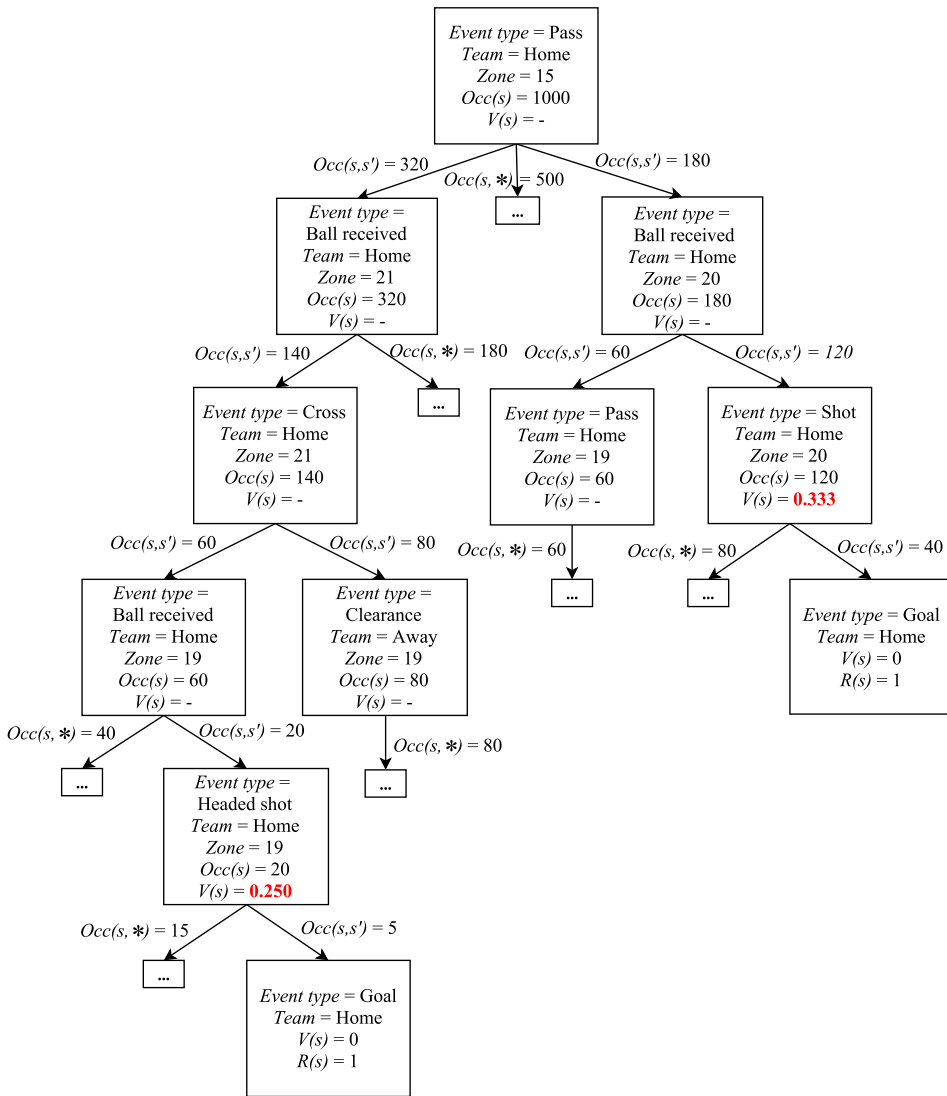
Value Iteration Examples

The intermediate iterations of the value iteration examples, illustrated in Section 4.2.3 and Section 4.3.3, are illustrated in this appendix. Figure A.1 illustrates the first five iterations of running the Markov chain value iteration algorithm, Algorithm 1. Figure A.2 illustrates the first three iterations of the Markov game value iteration algorithm, Algorithm 2. Each iteration is explained in the caption of its respective subfigure.

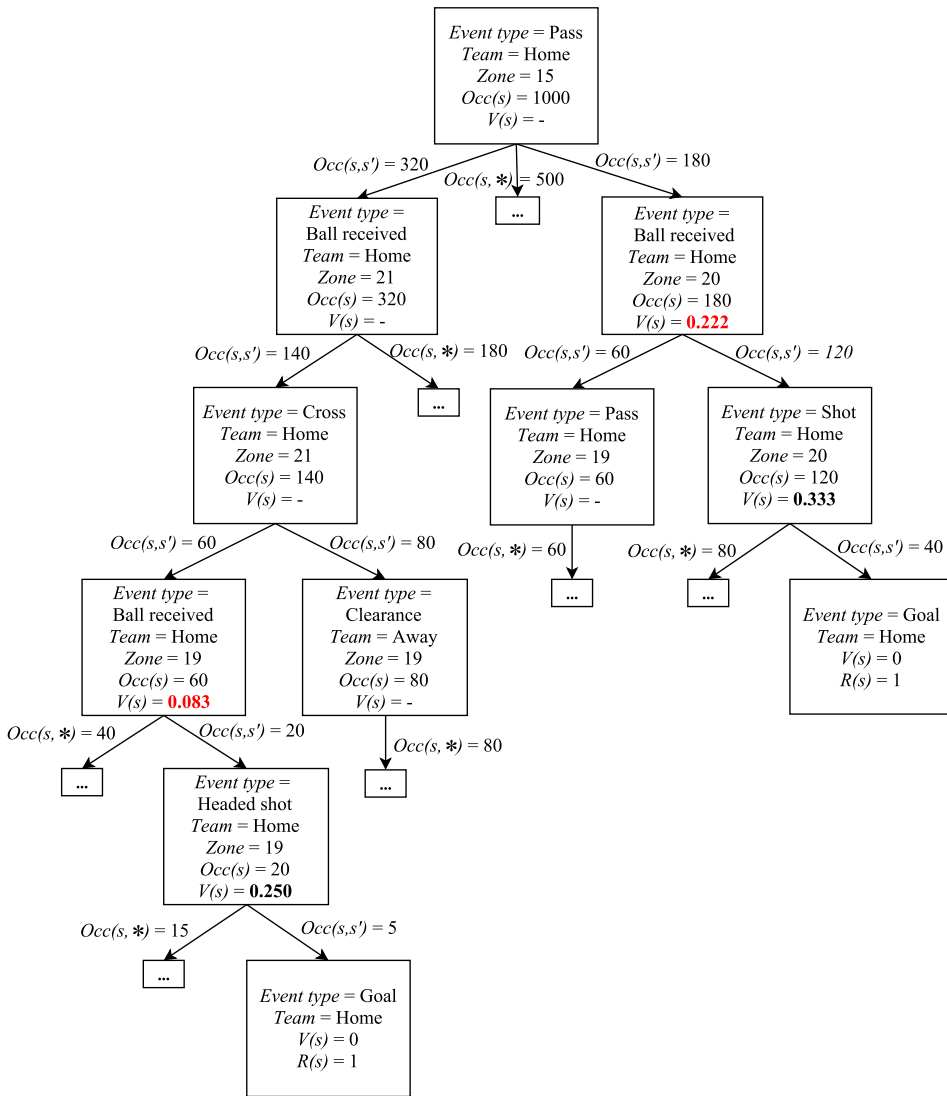
Figure A.1: Model 1 value iteration example, illustrating Algorithm 1.



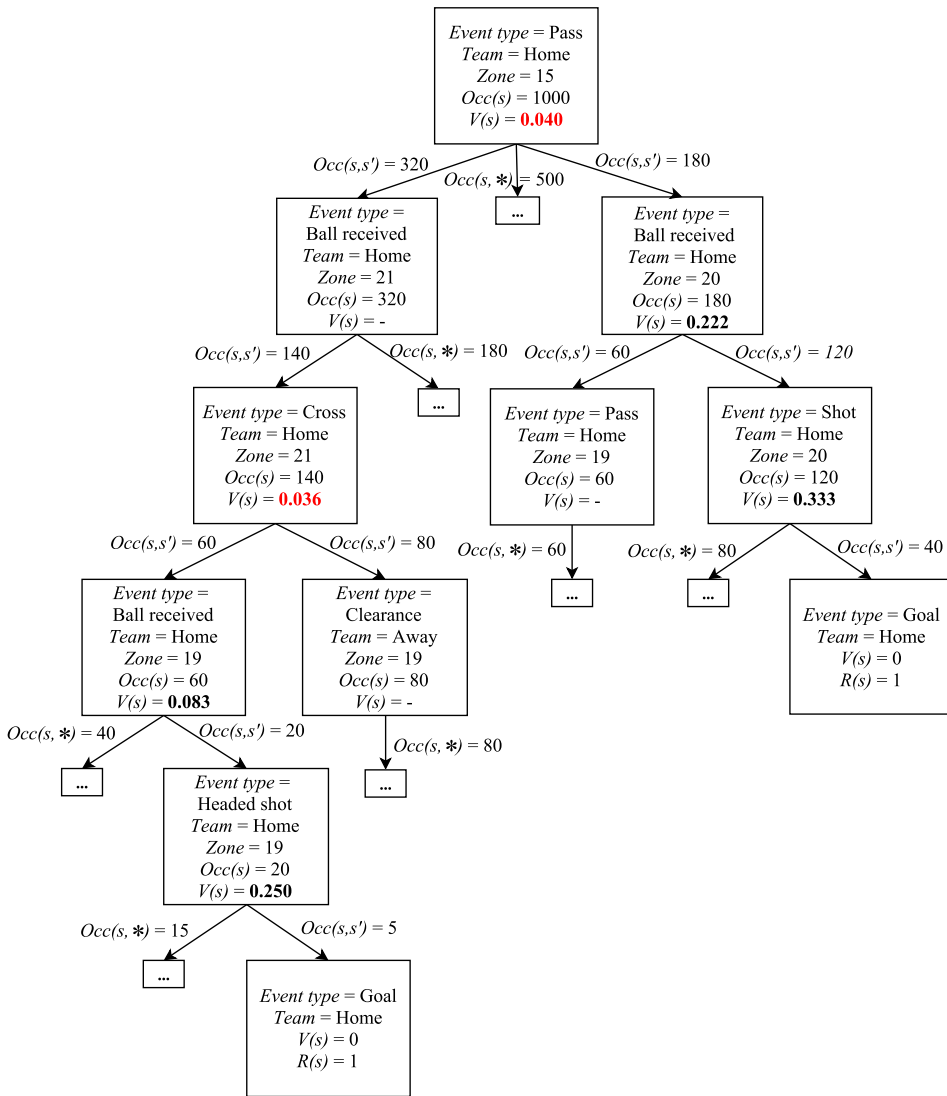
(a) Before running the algorithm. No state values are set, except the goal states, which have a value of 0 and reward of 1 or -1 . States not illustrated explicitly in the figure are denoted $*$.



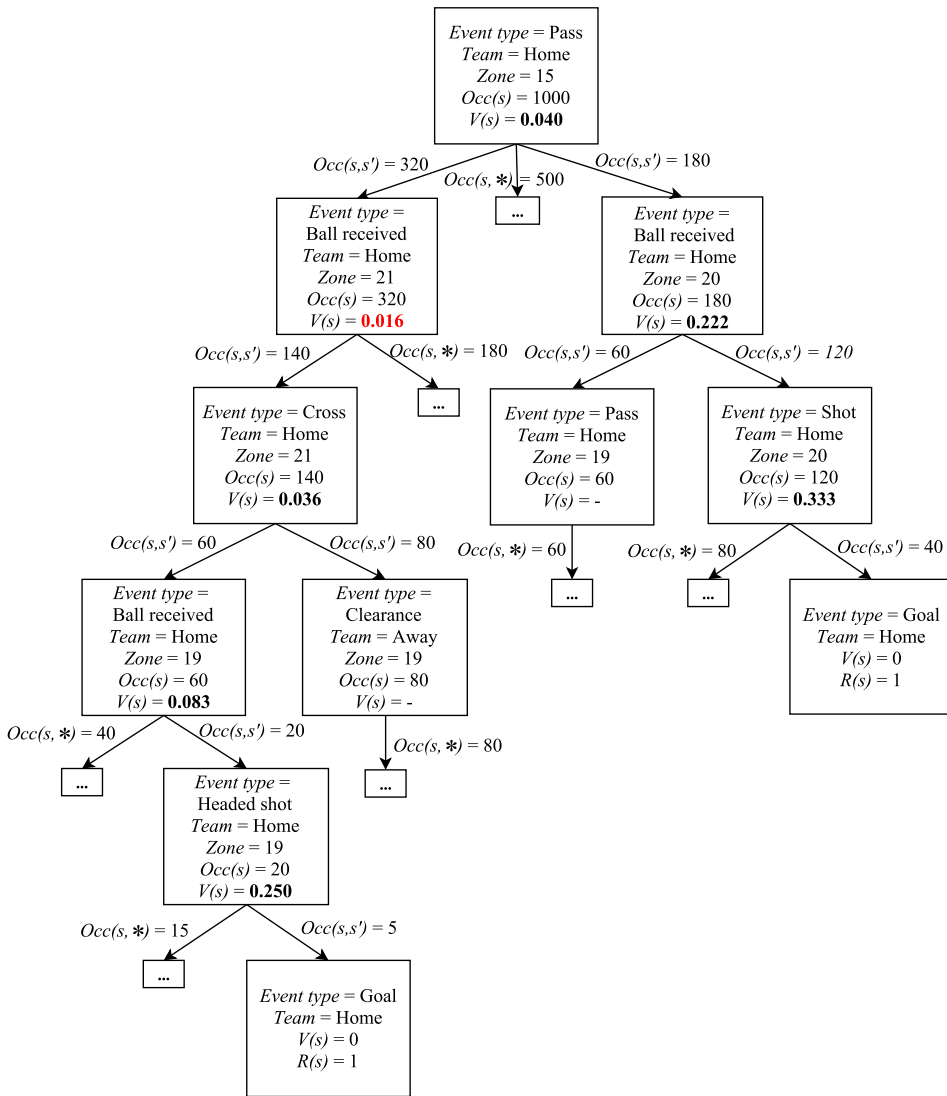
(b) After iteration 1. The state values of the states leading directly to a goal are updated. The updated values are highlighted in red. One state lead to a goal in one out of four occurrences. Another state lead to a goal in one out of three occurrences. Thus, after iteration 1, the states are valued 0.250 and 0.333, respectively.



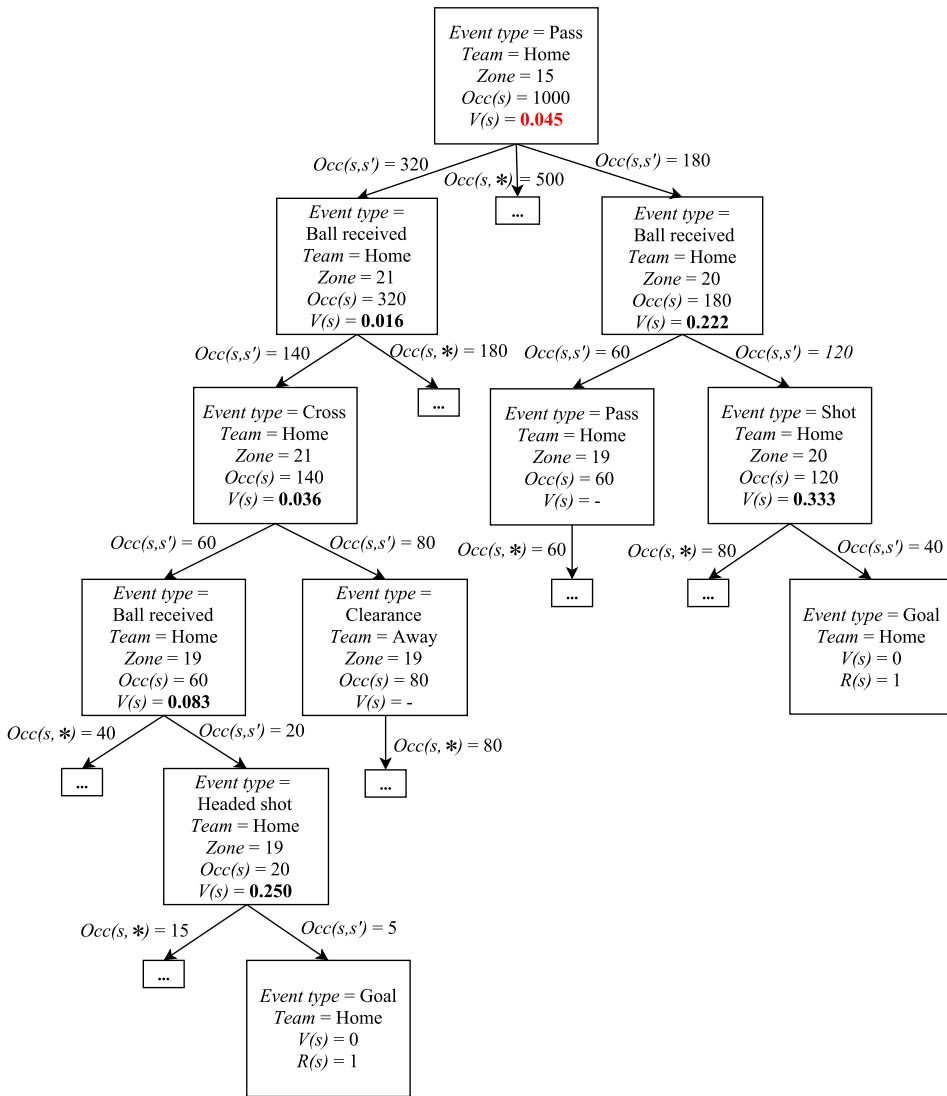
(c) After iteration 2. Two more state values, highlighted in red, are updated. One state leads to a state with a value of 0.250 in one third of the occurrences and is thus valued 0.083. The other state leads to a state with a value of 0.333 in two thirds of the occurrences and is valued 0.222. The state values updated in iteration 1, highlighted in bold, are not updated in iteration 2.



(d) After iteration 3. Two more states are updated in the third iteration. The states are updated according to the occurrences and state values of the subsequent states.

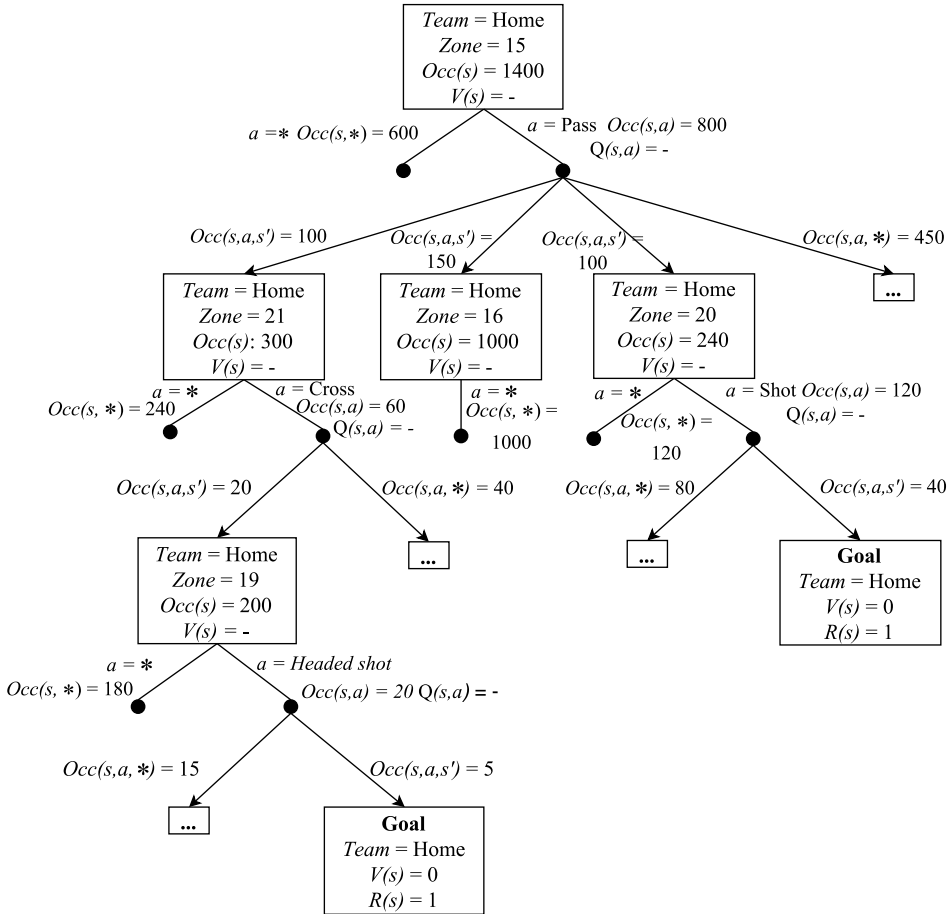


(e) After iteration 4. Only one state value is updated in the fourth iteration, where $V(s) = \frac{1}{320} \times (140 \times 0.036 + 180 \times 0) = 0.016$.

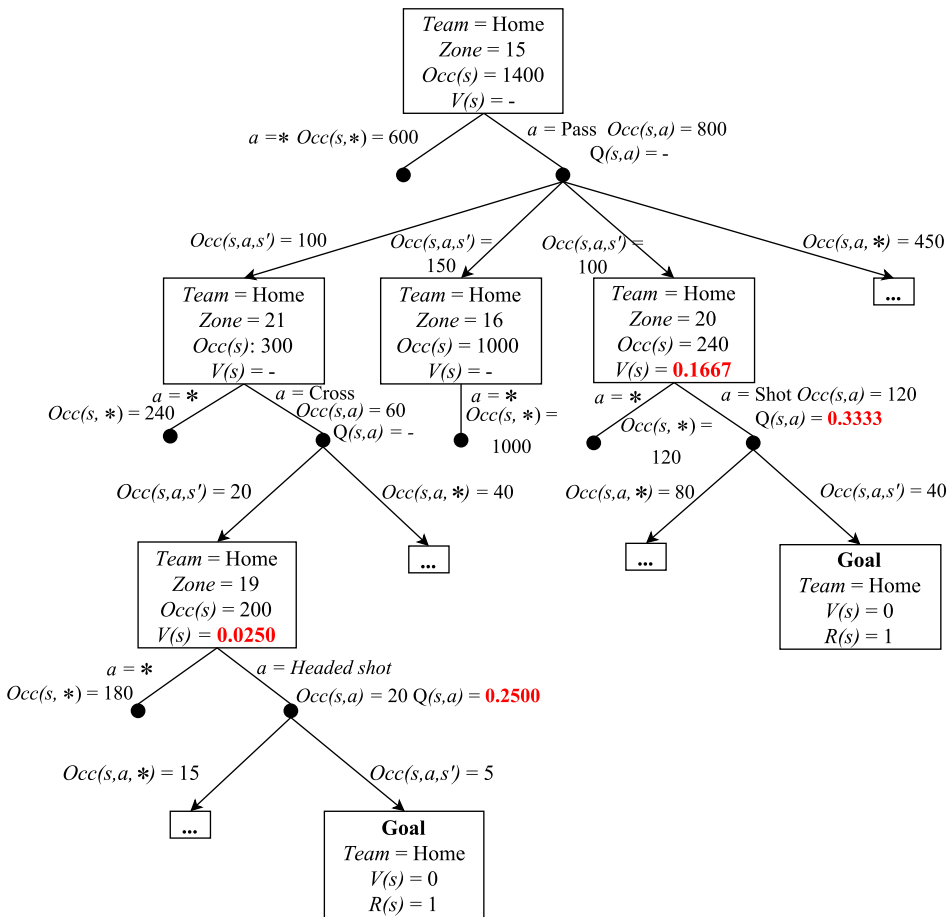


(f) After iteration 5. The values of most states shown in the figure have been updated. The state value updated in the current iteration is highlighted in red, while state values in bold have been updated in previous iterations. As subsequent states not illustrated explicitly in the figure, denoted *, are yet to be updated, the state values in the figure will be updated in later iterations. The algorithm converges when the state values changes by less than the predetermined convergence criterion.

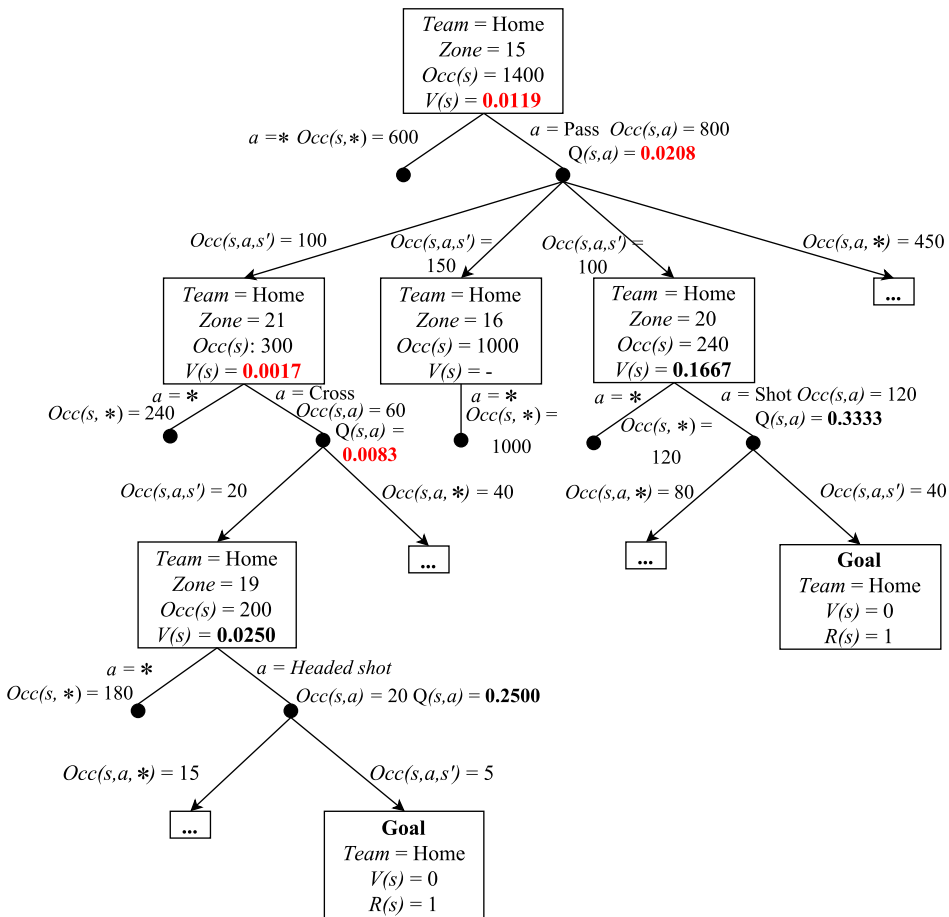
Figure A.2: Model 2 value iteration example, illustrating Algorithm 2.



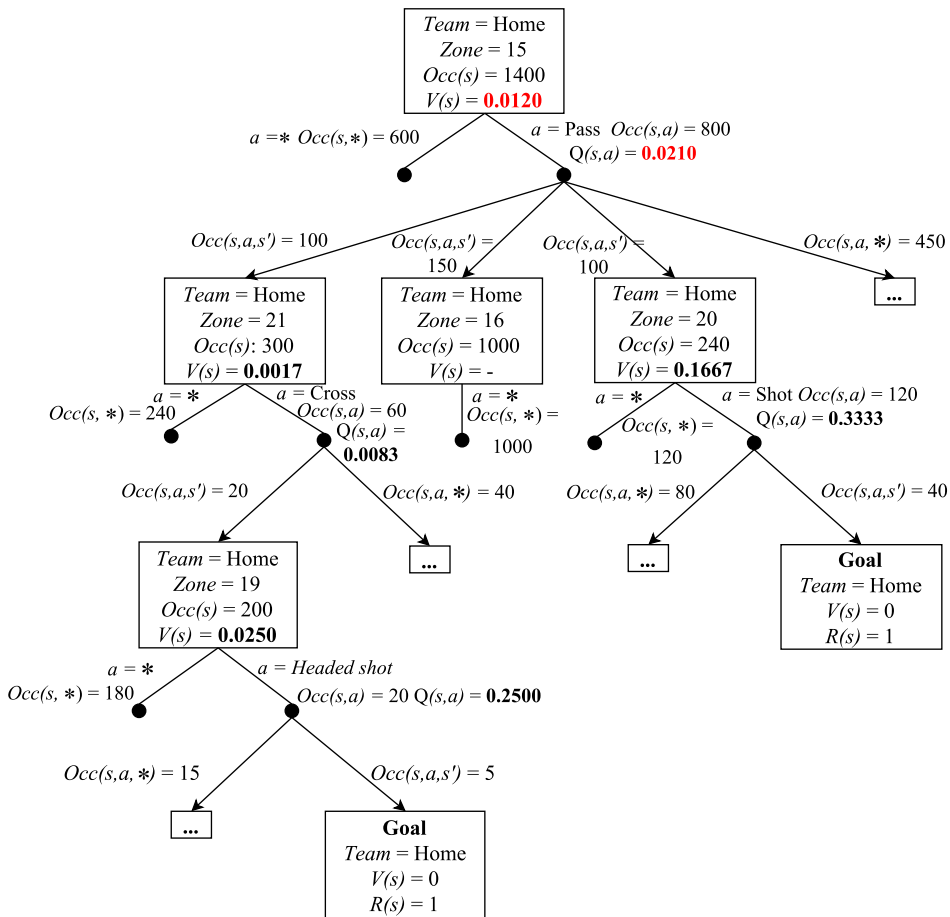
(a) Before running the algorithm. No state values and Q -values are set, except the goal states, which have a state value of 0 and reward of 1 or -1 . Subsequent states and actions not illustrated explicitly in the figure are denoted $*$.



(b) After iteration 1. Two state values and two Q -values are updated in the first iteration in this example. The leftmost Q -value highlighted in red, representing a headed shot in zone 19, is valued 0.2500 as the state-action pair lead to a goal in one out of four occurrences. As headed shot is performed in one tenth of the occurrences of the state, the state is valued 0.0250, until updated again in later iterations. Following the same line of reasoning, the rightmost state and Q -value highlighted in red, are valued 0.1667 and 0.3333 respectively.



(c) After iteration 2. Two state values and two Q -values are updated in the second iteration, according to the occurrences and values of subsequent states and state-action pairs. The values updated in the current iteration are highlighted in red, while values updated in previous iterations are highlighted in bold.



(d) After iteration 3. Most state values and Q -values shown in the figure are updated. As the values of subsequent states and action-action pairs not illustrated explicitly in the figure are yet to be updated, the illustrated values will be updated again in later iterations.

Appendix B

k -fold Cross-Validation Regression Results

The complete regression results from the k -fold cross-validation are presented in this appendix. Ten sets of regression coefficients are computed for each player rating, represented by the underlying impact function. All regression coefficients and the associated p -values are listed in the tables below. One table is presented for each impact function. The results are discussed in Section 6.1.1.

Table B.1: $I_{1,1}$ k -fold cross-validation regression results.

Fold	$\alpha^{(1)}$	$p_{\alpha^{(1)}}$	$\alpha^{(2)}$	$p_{\alpha^{(2)}}$	β	p_{β}
1	-0.983	0.000	0.204	0.016	-3.474	0.000
2	-0.933	0.000	0.289	0.001	-3.540	0.000
3	-0.910	0.000	0.237	0.005	-3.363	0.000
4	-0.967	0.000	0.225	0.008	-3.555	0.000
5	-0.923	0.000	0.220	0.009	-3.463	0.000
6	-0.923	0.000	0.220	0.009	-3.463	0.000
7	-0.986	0.000	0.207	0.015	-3.645	0.000
8	-0.961	0.000	0.213	0.012	-3.588	0.000
9	-0.885	0.000	0.273	0.001	-3.444	0.000
10	-0.940	0.000	0.243	0.004	-3.810	0.000

Table B.2: $I_{1,2}$ k -fold cross-validation regression results.

Fold	$\alpha^{(1)}$	$p_{\alpha^{(1)}}$	$\alpha^{(2)}$	$p_{\alpha^{(2)}}$	β	p_{β}
1	-0.908	0.000	0.103	0.193	3.787	0.223
2	-0.862	0.000	0.172	0.030	2.968	0.330
3	-0.840	0.000	0.141	0.074	4.036	0.189
4	-0.892	0.000	0.127	0.108	2.043	0.505
5	-0.824	0.000	0.172	0.030	2.369	0.436
6	-0.841	0.000	0.138	0.080	1.479	0.624
7	-0.894	0.000	0.112	0.155	0.666	0.831
8	-0.894	0.000	0.108	0.171	3.627	0.242
9	-0.834	0.000	0.163	0.039	1.237	0.672
10	-0.862	0.000	0.133	0.091	-2.666	0.400

Table B.3: $I_{2,1}$ k -fold cross-validation regression results.

Fold	$\alpha^{(1)}$	$p_{\alpha^{(1)}}$	$\alpha^{(2)}$	$p_{\alpha^{(2)}}$	β	p_{β}
1	-0.949	0.000	0.239	0.005	-6.350	0.000
2	-0.897	0.000	0.322	0.000	-6.266	0.000
3	-0.872	0.000	0.266	0.002	-6.044	0.000
4	-0.926	0.000	0.259	0.002	-6.438	0.000
5	-0.858	0.000	0.313	0.000	-7.058	0.000
6	-0.883	0.000	0.252	0.003	-6.519	0.000
7	-0.938	0.000	0.237	0.005	-6.556	0.000
8	-0.916	0.000	0.243	0.004	-6.112	0.000
9	-0.844	0.000	0.310	0.000	-6.285	0.000
10	-0.892	0.000	0.289	0.001	-7.255	0.000

Table B.4: $I_{2,2}$ k -fold cross-validation regression results.

Fold	$\alpha^{(1)}$	$p_{\alpha^{(1)}}$	$\alpha^{(2)}$	$p_{\alpha^{(2)}}$	β	p_{β}
1	-1.002	0.000	0.126	0.127	-16.219	0
2	-0.928	0.000	0.222	0.007	-16.198	0
3	-0.920	0.000	0.169	0.041	-14.913	0
4	-0.981	0.000	0.153	0.064	-16.328	0
5	-0.923	0.000	0.189	0.023	-16.232	0
6	-0.925	0.000	0.166	0.045	-16.183	0
7	-0.981	0.000	0.147	0.076	-16.585	0
8	-0.981	0.000	0.138	0.096	-16.367	0
9	-0.901	0.000	0.202	0.014	-15.440	0
10	-0.942	0.000	0.169	0.041	-16.519	0

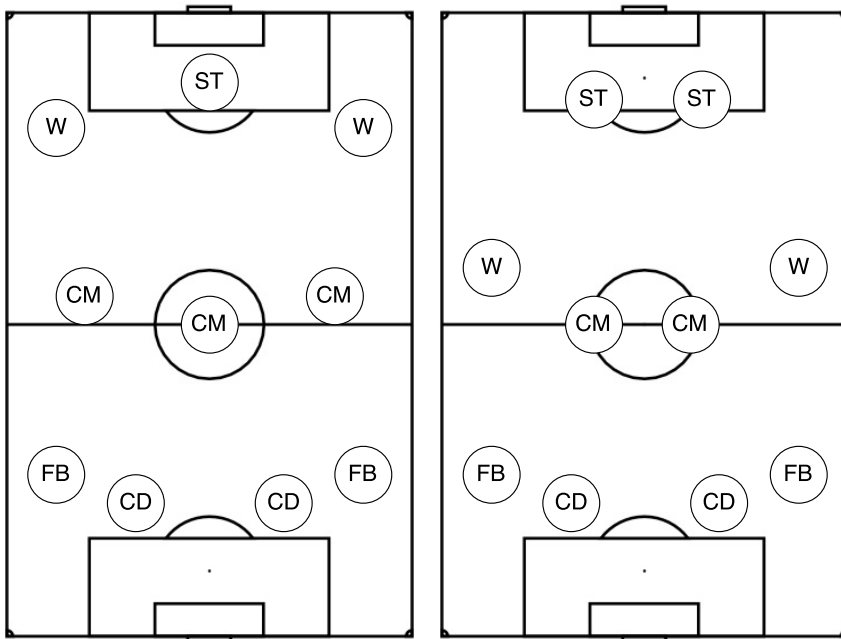
Formations and Positions

This appendix clarifies relevant formation and position terminology. Figure C.1 illustrates the positions of players in the two most common football formations, namely 4-3-3 and 4-4-2. The position abbreviations are explained in table C.1. As argued in Section 6.3, the two wider central midfielders in 4-3-3 resemble the wingers in the 4-4-2 formation. These two roles are primarily occupied by Midtsjø and Jensen in RBK, and help explain why they resemble both central midfielders and wingers. The most central midfielder in 4-3-3, occupied by Konradsen in RBK, resemble the role of the two central midfielders in 4-4-2.

Table C.1: Position abbreviations.

Abbreviation	Position
CD	Central Defender
FB	Full Back
CM	Central Midfielder
W	Winger
ST	Striker

Figure C.1: Player positions in different formations.



(a) 4-3-3 Formation

(b) 4-4-2 Formation