A Markov Game model for valuing actions, locations, and team performance in ice hockey

Article *in* Data Mining and Knowledge Discovery · November 2017

DOI: 10.1007/s10618-017-0496-z

6 authors, including:

Oliver Schulte
Simon Fraser University

Sajjad Gholami
Simon Fraser University

Mehrsan Javan Roshtkhari
McGill University

Data Min Knowl Disc
DOI 10.1007/s10618-017-0496-z

A Markov Game model for valuing actions, locations, and team performance in ice hockey

Oliver Schulte[1,2] · **Mahmoud Khademi**[1,2] · **Sajjad Gholami**[1,2] · **Zeyu Zhao**[1,2] · **Mehrsan Javan**[1,2] · **Philippe Desaulniers**[1,2]

Received: 7 March 2016 / Accepted: 8 February 2017

Abstract We apply the Markov Game formalism to develop a context-aware approach to valuing player actions, locations, and team performance in ice hockey. The Markov Game formalism uses machine learning and AI techniques to incorporate context and look-ahead. Dynamic programming is applied to learn value functions that quantify the impact of actions on goal scoring. Learning is based on a massive new dataset, from SportLogiq, that contains over 1.3M events in the National Hockey League. The SportLogiq data include the location of an action, which has previously been unavailable in hockey analytics. We give examples showing how the model assigns context and location aware values to a large set of 13 action types. Team performance can be assessed as the aggregate value of actions performed by the team's players, or the aggregate value of states reached by the team. Model validation shows that the total team action and state value both provide a strong indicator predictor of team success, as measured by the team's average goal ratio.

Keywords Markov Game · Sports analytics · National Hockey League · Q-learning

1 Introduction

A fundamental goal of sports statistics is to understand which actions contribute to winning in what situation. As sports have entered the world of big data, there is increasing opportunity for large-scale machine learning to model complex sports dynamics. The research described in this paper applies AI techniques to model the dynamics

✉ Oliver Schulte
oschulte@cs.sfu.ca

1 School of Computing Science, Simon Fraser University, Vancouver-Burnaby, Canada

2 SportLoqig, Montreal, Canada

Published online: 24 March 2017

of ice hockey; specifically the Markov Game model formalism (Littman 1994), and related computational techniques such as the dynamic programming value iteration algorithm. We make use of a new dataset for about 446 matches in the National Hockey League (NHL), provided by SportLogiq. This dataset comprises play-by-play events from October to December 2015, for a total of over 1.3M events/actions. The Markov Game model features almost 10,000 states and 148 actions (actions comprise an action type and parameters, see below).

Whereas most previous works on Markov Game models aim to compute optimal strategies or policies (Littman 1994) (i.e., minimax or equilibrium strategies), we learn a model of how hockey is actually played, and do not aim to compute optimal strategies. In reinforcement learning (RL) terminology, we use dynamic programming to compute a *value function* in the *on policy* setting (Sutton and Barto 1998). In RL notation, the expression $V(s)$ denotes the expected reward when the game starts in state s.

1.1 Motivation

Motivation for learning a value function for NHL hockey dynamics includes the following.

Knowledge Discovery The Markov Game model provides information about the likely consequences of actions. The basic model and algorithms can easily be adapted to study different outcomes of interest, such as goals and penalties.

Valuing Actions One of the main tasks for sports statistics is evaluating the performance of teams and players (Schumaker et al. 2010). A common approach is to assign action values, and sum the corresponding values each time a player takes the respective action. A simple and widely used example in ice hockey is the $+/-$ score: for each goal scored by (against) a player's team when he is on the ice, add +1 (-1) point. Researchers have developed several extensions of $+/-$ for hockey (Macdonald 2011; Spagnola 2013; Schuckers and Curro 2013). The NHL has started publishing advanced player statistics such as the Corsi (Shot Attempts) and Fenwick (Unblocked Shot Attempts) ratings. The Markov model approach has the following advantages over the previous action count approaches used in ice hockey.

- Context-Awareness: Action counts are unaware of the *context* of actions within a game. For example, a goal is more valuable in a tied-game situation than when the scorer's team is already four goals ahead (Pettigrew 2015). In the Markov Game model, *context* = *state*. Richer state spaces therefore capture more of the context of an action.
- Look-Ahead: Previous action scores are based on immediate positive consequences of an action (e.g. goals following a shot). However, an action may have medium-term and/or ripple effects rather than immediate consequences in terms of visible rewards like goals. Therefore evaluating the impact of an action requires *look-ahead*. Long-term look-ahead is especially important in ice hockey because evident rewards like goals occur infrequently (Schuckers and Curro 2013). For example, if a player receives a penalty, this leads to a manpower disadvantage for his team,

Journal: 10618-DAMI Article No.: 0496 TYPESET DISK LE CP Disp.:2017/3/23 Pages: 23 Layout: Small-X

known as a powerplay for the other team. It is easier to score a goal during a powerplay, but this does not mean that a goal will be scored immediately after the penalty. The dynamic programming value iteration algorithm of Markov Decision Processes provides a computationally efficient way to perform unbounded look-ahead.

- All Events: Analysis Without looking ahead to the medium-term effects of an action, it is difficult to assign an exact value to actions other than shots and goals. In contrast, the Markov Game model provides a value for *all* actions, including defensive ones. This means that the model utilizes all the information in our rich dataset with many different action types.

While this paper applies the Markov Game approach for valuing actions to ice hockey only, Cervone et al. (2014) have applied it successfully to basketball as well; we discuss this further under related work.

1.2 Evaluation

Our evaluation learns a value function for the chance of scoring the next goal from the SportLogiq data. We provide examples of the context dependence and the look-ahead analysis. We introduce two complementary methods for using the Markov Game model to evaluate team performance. (1) Use the value function to quantify the value of a team's *action* in a context. The action values are then aggregated over games to get an estimate of the team strength. This estimate is a strong indicator of the team's success: the correlation between average action values and average goal ratios is 0.7. (2) Use the value function to quantify the value of a game *state* for a team, where value represents the team's chance of winning the game. The state values are then aggregated over games to get an estimate of the team strength. The intuition is that strong teams often manage to reach states with higher winning chances than their opponent. This estimate is an even stronger indicator of the team's success: the correlation between average state values and average goal ratios is 0.82.

1.3 Contributions

We make our code available on-line (Routley et al. 2015). The main contributions of this paper may be summarized as follows:

1. The first Markov Game model for a large ice hockey state and action space (over 10,000 states and 148 actions) that includes location information.
2. Learning a context and location aware value function that models play dynamics in the National Hockey League from a large data set (1.3M events).
3. Applying the Markov Game value function to evaluate and predict team performance.

We review related work in measuring player contributions and machine learning in sports in Sect. 2. We then give some background information on the ice hockey domain and our play-by-play sequences data. Our Markov Game model translates the hockey

Journal: 10618-DAMI Article No.: 0496 TYPESET DISK LE CP Disp.:2017/3/23 Pages: 23 Layout: Small-X

domain features into the Markov formalism. We describe the dynamic programming method for computing the values of states and actions. We give examples of how actions are evaluated, including shots, dump-ins, carries, and loose puck recoveries. We apply the model to rank the aggregate performance of teams. Past aggregate performance correlates very well with success in terms of goals. We conclude with a number of potential extensions and open problems for future work.

2 Related work

2.1 Extensions of the +/− score

Several papers aim to improve the basic +/− score with statistical techniques (Macdonald 2011; Gramacy et al. 2013; Spagnola 2013). A common approach is to use regression techniques where an indicator variable for each player is used as a regressor for a goal-related quantity (e.g., log-odds of a goal for the player's team vs. the opposing team). The regression weight measures the extent to which the presence of a player contributes to goals for his team or prevents goals for the other team. These approaches look at only goals, no other actions. The only context they take into account is which players are on the ice when a goal is scored. Regression could be combined with our Markov Game model to capture how team impact scores depend on the presence or absence of individual players.

2.2 All event analyses

The total hockey rating (THoR) (Schuckers and Curro 2013) assigns a value to all actions, not only goals. Actions were evaluated based on whether or not a goal occurred in the following 20 s after an action. For penalties, the duration of the penalty was used as the look-ahead window. This work used data from the 2006/2007 NHL season only. THoR assumes a fixed value for every action and does not account for the context in which an action takes place. Furthermore, the window of 20 s restricts the look-ahead value of each action. Our Markov Game learning method is not restricted to any particular time window for look-ahead.

Cervone et al. (2014) use spatial–temporal tracking data for basketball to build the POINTWISE model for valuing player decisions and player actions. Conceptually, their approach to defining action values is the closest predecessor to ours: The counterpart to the value of a state in a Markov Game is called expected possession value (EPV). The counterpart to the impact of an action on this value is called EPV-added (EPVA). Cervone et al. emphasize the broad potential of the context-based impact definitions: "we assert that most questions that coaches, players, and fans have about basketball, particularly those that involve the offense, can be phrased and answered in terms of EPV" (i.e., the state values that we computed in this paper).

The current paper adapts and extends (Routley and Schulte 2015). Both papers use a Markov Game model for evaluating actions. The details of the model and the evaluation are both different. (1) Different state/action space: the previous paper included the recent match trajectory in the model, whereas the current paper includes location

Journal: 10618-DAMI Article No.: 0496 TYPESET DISK LE CP Disp.:2017/3/23 Pages: 23 Layout: Small-X

information instead. (2) The previous paper used the Markov Game model to evaluate player performance. The current paper evaluates team performance instead. Team performance provides an independent ground-truth metric in terms of goals scored for and against a team.

2.3 Markov decision process models for other sports

MDP-type models have been applied in a number of sports settings, such as baseball, soccer and football. For review, please see Cervone et al. (2014). Our work is similar in that our method estimates a value function on a Markovian state space, however, previous Markov models in sports use a much smaller state space. The goal of these models is to find an optimal policy for a critical situation in a sport or game. In contrast, we learn in the on-policy setting whose aim is to model hockey dynamics as it is actually played.

3 Domain description: hockey rules and hockey data

We outline the rules of hockey and describe the dataset available from the NHL in Sect. 3.1 then we will discuss about the data format in Sect. 3.2.

3.1 Hockey rules

We give a brief overview of rules of play in the NHL (National Hockey League 2014). NHL games consist of three periods, each 20 min in duration. A team has to score more goals than their opponent within three periods in order to win the game. If the game is still tied after three periods, the teams will enter a fourth overtime period, where the first team to score a goal wins the game. If the game is still tied after overtime during the regular season, a shootout will commence. During the playoffs, overtime periods are repeated until a team scores a goal to win the game. Teams have five skaters and one goalie on the ice during even strength situations. Penalties result in a player sitting in the penalty box for 2, 4, 5 or 10 min and the penalized team will be shorthanded, creating a manpower differential between the two teams. The period where one team is penalized is called a powerplay for the opposing team with a manpower advantage. A shorthanded goal is a goal scored by the penalized team, and a powerplay goal is a goal scored by the team on the powerplay.

3.2 Data format

The dataset was constructed by SportLogiq using video analysis. A breakdown of this dataset is shown in Table 1. It comprises 446 matches from the fall of 2015. The play-by-play event data records the 13 action types as Table 2 shows. In the full dataset, the action types are classified further for a total of 43 types. For example, for each dump-in, the data distinguish a chip-in from an actual dump-in. We used only the 13 main level types, to reduce the number of parameters of the Markov Game model. Fewer

Journal: 10618-DAMI Article No.: 0496 TYPESET DISK LE CP Disp.:2017/3/23 Pages: 23 Layout: Small-X

Table 1 Dataset statistics

Number of teams	30
Number of players	2233
Number of games	446
Number of events	1,048,576

Table 2 Action types definition and their number of clusters using our clustering algorithm (Sect. 4)

Action	Description	#Clusters	#Occurrences
Block	A block attempt on the puck's trajectory	5	69,269
Carry	Controlled carry over a blue line or the red center line	8	75,731
Check	When a player attempts to use his body to remove possession from an opponent	7	33,993
Dump in	When a player sends the puck into the offensive zone	3	26,559
Dump out	When a defending player dumps the puck up the boards without targeting a teammate for a pass	3	29,870
Goal	The player has scored a goal	1	1804
lpr	Loose puck recovery. The player recovered the puck as it was out of possession of any player	6	214,033
Offside	When a player is caught over the offensive blue line before their teammate brings the puck in	3	2198
Pass	The player attempts a pass to a teammate	7	275,686
Puck protection	When a player uses their body to protect the puck along the boards	7	34,918
Reception	When a player receives a pass from a teammate	6	213,879
Shot	A player shoots on goal	4	42,095
Shot against	A shot was taken by the opposing team	3	28,541

parameters reduce the computational complexity, and can be more reliably estimated given limited data.

Every event is marked with a continuous time stamp, an x–y location, and the team that carries out the action of the event.

We store sequence data in SQL tables (see Table 3). SQL provides fast retrieval, and native support for the necessary COUNT operations.

4 Spatial discretization

Figure 1 shows a schematic layout of the ice hockey rink. The units are feet. Adjusted Y-coordinates run from −42.5 at the bottom to 42.5. The goal line is at X = 89. While the data provide continuous locations, we took the approach of discretizing the continuous space into a discrete set. The disadvantage to this approach is that it loses some information about the exact location of an action event. The computational advantage

Journal: 10618-DAMI Article No.: 0496 TYPESET DISK LE CP Disp.:2017/3/23 Pages: 23 Layout: Small-X

Table 3 Sample play-by-play data in tabular format

GameId	PlayerId	Period	TeamId	xCoord	yCoord	Manpower	Action type
849	402	1	15	−9.5	1.5	Even	lpr
849	402	1	15	−24.5	−17	Even	Carry
849	417	1	16	−75.5	−21.5	Even	Check
849	402	1	15	−79	−19.5	Even	Puckprot
849	413	1	16	−92	−32.5	Even	lpr
849	413	1	16	−92	−32.5	Even	Pass
849	389	1	15	−70	42	Even	Block
849	389	1	15	−70	42	Even	lpr
849	389	1	15	−70	42	Even	Pass
849	425	1	16	−91	34	Even	Block
849	395	1	15	−97	23.5	Even	Reception

is that we can employ algorithms for discrete Markov models. The statistical advantage is that discretization requires neither parametric assumptions (e.g. Gaussian or Poisson distribution), nor stationarity assumptions that treat different locations as the same. To minimize the loss of information, we learned the discrete location bins by clustering occurrences of each action type. There is therefore a separate clustering for each action type. Learning the clusters has the advantage over a fixed discretization that clusters are allocated to areas of high density; some action types occur very rarely in some parts of the rink. For example, shots are usually taken from the offensive zone.

The clustering method we used was affinity propagation (AP) (Frey and Dueck 2007). AP is a clustering algorithm based on the idea of message passing between data points. It uses a set of real-valued pair-wise data point similarities as input. We used the negative of Euclidean distance between the data points as the measure of similarity. Unlike clustering algorithms such as k-means, AP does not need the number of clusters to be determined. AP automatically determines the number of clusters, based on a

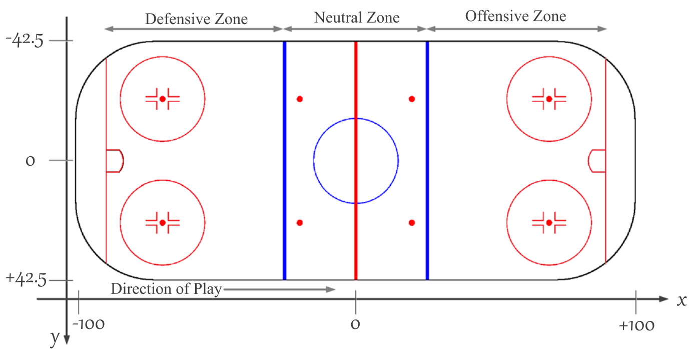

Fig. 1 Rink layout with adjusted coordinates. Coordinates are adjusted so that for the team performing an action, its offensive zone is on the right

Journal: 10618-DAMI Article No.: 0496 TYPESET DISK LE CP Disp.:2017/3/23 Pages: 23 Layout: Small-X

preference hyperparameter $p(i)$; data points with higher preferences are more likely to be selected as cluster centers. The number of identified clusters can be increased or decreased by changing this value accordingly. A recommended default setting is to assign equal preference to all data points, and choose the shared preference value to be the median of the similarity values (Frey and Dueck 2007). However, we found that this resulted in too many clusters to keep the Markov model state space tractable. We tried various multipliers, and found that setting p to four times the median of the similarity values for all data points led to a tractable number of clusters (typically five, no more than nine per action).

For each action type, we found the set of x–y points where the action type was performed, and we applied affinity propagation to produce a custom clustering for the action type. This means that the locations maximize the information about where actions of the given type are performed. Figure 2 illustrates the resulting clusterings for shots and the two actions that occur most frequently in the SportLogiq dataset, passes (Fig. 2b) and loose puck recoveries (Fig. 2c). The occurrence count is the number of times that the action is performed. We incorporated background knowledge by manually dividing the offensive zone into three clusters for shots (Fig. 2a). Cluster 1 is an area known as the *slot*, Cluster 2 the area above and around the slot, Cluster 3 the area below and around the slot. In fact, affinity propagation discovered a similar clustering from the data.

5 Markov Games

A Markov Game (Littman 1994), sometimes called a stochastic game, is defined by a set of states, S, and a collection of action sets, one for each agent in the environment. State transitions are controlled by the current state and a list of actions, one action from each agent. For each agent, there is an associated reward function mapping a state transition to a reward. An overview of how our Markov Game model fills in this schema is as follows. There are two players, the Home Team H and the Away Team A. In each state, only one team performs an action, although not in a turn-based sequence. This reflects the way the data record actions. Thus at each state of the Markov Game, exactly one player chooses no-operation. We introduce the following generic notation for all states, following (Russell and Norvig 2010; Littman 1994).

- $Occ(s, a)$ is the number of times that action a occurs in state s as observed in the play-by-play data.
- $Occ(s, a, s')$ is the number of occurrences of action a in state s being immediately followed by state s' as observed in the play-by-play data. (s, s') forms an edge with label a in the transition graph of the Markov Game model.
- The transition probability function TP is a mapping of $S \times A \times S \to (0, 1]$. We estimate it using the observed transition frequency $\dfrac{Occ(s, a, s')}{Occ(s, a)}$.

We begin by defining the state space, then the action space.

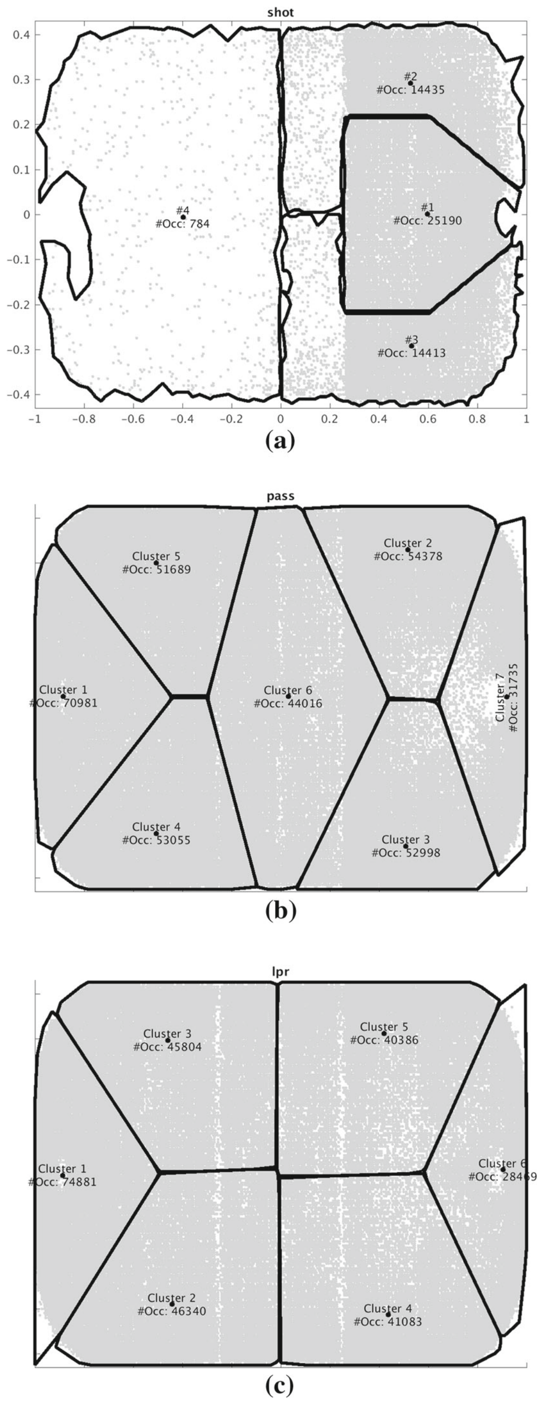

Fig. 2 Locations are discretized separately for each action by clustering (affinity propagation). *Gray dots* indicate occurrences. The cluster mean is shown, with labels that indicate (i) the action occurrence count, (ii) the value of the action at different locations. **a** Clustering of locations for shots. **b** Clustering of locations for pass. **c** Clustering of locations for loose puck recovery

Journal: **10618-DAMI** Article No.: **0496** TYPESET DISK LE CP Disp.:**2017/3/23** Pages: **23** Layout: **Small-X**

Table 4 Context features

Notation	Name	Range
GD	Goal differential	$[-8, 8]$
MD	Manpower differential	{EV, SH, PP}
P	Period	$[1, 3]$

5.1 State space

Previous work on Markov process models for ice hockey (Thomas et al. 2013) defined states in terms of hand-selected features that are intuitively relevant for the game dynamics, such as the goal differential and penalties. We refer to such features as *context features*.

A state lists the values of relevant features at a point in the game. These features are shown in Table 4, together with the range of integer values observed. Goal Differential *GD* is calculated as Number of Home Goals—Number of Away Goals. A positive (negative) goal differential means the home team is leading (trailing). Manpower Differential MD specifies whether the teams are at even strength (EV), the acting team is short-handed (SH) or in a powerplay (PP).[1] Period P represents the current period number the play sequence occurs in, typically ranging in value from 1 to 5. Periods 1–3 are the regular play of an ice hockey game. Our model uses only periods from 1 to 3. Potentially, there are $(17 \times 3 \times 3) = 153$ context states. In our NHL dataset, 116 context states occur at least once.

Most previous research on Markov process models of hockey has used only context features. A Markov process model with context features can answer questions such as how goal scoring or penalty rates depend on the game context (Thomas et al. 2013). However, it cannot answer questions about actions other than goals scored or penalties drawn. We next introduce the action space for our Markov Game model that incorporates actions.

5.2 Action space

The basic set of 13 action types was listed in Sect. 3.2. Each of these action types has two parameters: which team T performs the action and the location L where the action takes place. We defined a discrete space of locations by discretizing continuous x–y coordinates in a data-driven way; we present the details in Sect. 4 below. Using action description language notation (Levesque et al. 1998), we write action events in the form $a(T, L)$. For example, $block(Home, L_1)$ denotes that the home team blocks a shot at location 1. We usually omit the action parameters from generic notation and write a for a generic action event. There are 63 action-location pairs (sum of the number of clusters in Table 2), so considering we have two teams, home and away team, we have $63 \times 2 = 126$ possible action events that define transitions in the Markov Game model.

[1] Pulling the goalie can also result in a skater manpower advantage.

Journal: 10618-DAMI Article No.: 0496 TYPESET DISK LE CP Disp.:2017/3/23 Pages: 23 Layout: Small-X

The total number of potential transitions is therefore $153 \times 126 \times 153 = 2,949,534$. The number of transitions we actually observe in the data is $112,590$. Our model assigns 0 probability to transitions that never occur, so the total number of parameters to estimate is $112,590$.

5.3 Reward functions

A strength of Markov Game modelling is that it can be applied with many reward functions, depending on what results are of interest (Routley and Schulte 2015). Our evaluation below utilizes two reward functions: (1) goal scoring, specifically scoring the next goal, and (2) winning. These can be defined as follows.

1. For any state s where a Home resp. Away goal is scored, we set $R_T(s, a) := 1$ where $T = H, A$. For other states the reward is 0.
2. After a goal is scored, the next state is an absorbing state (no transitions from this state).
3. For any state s where the game ends with a win for Home resp. Away team, we set $R_T(s, a) := 1$ where $T = H, A$. For other states the reward is 0. All such states are absorbing states by definition.

5.4 State values

Given the specification of a Markov Game model, the expected reward for a team represents the expected total reward that a team obtains from a random walk through the state space of unbounded length. The expected reward from a starting state s is the *value* of a state, denoted $V_T(s)$ where $T = H, A$. We will consider the expected total reward from random walks whose length is bounded by a constant ℓ. The bound ℓ can be interpreted as the *look-ahead horizon*. Strictly speaking, the value of a state depends on the transition probabilities, and on the look-ahead horizon. These will be implicitly fixed by context, so we do not show the dependency explicitly in our value function notation. For a given reward function, we learn an appropriate look-ahead horizon from the data (see Sect. 6 below).

We also consider the *conditional state value*. The conditional value for the home team is the value for the home team performing an action, divided by the sum of the value for the home team and the value for the away team. Similarly for the away team. The conditional value corresponds to conditioning on the event that at least one of the team scores a goal within the next 14 time steps (our look-ahead horizon). It measures the expected reward of a team relative to its opponents, in other words, the advantage that a team enjoys over its opponent in a given match state. For example, with reward = scoring the next goal, a conditional value of p for the home team means that, given that one of the teams will score the next goal within the look-ahead horizon, the chance is p that the home team manages the next goal.

Discussion Winning the game is the ultimate aim of a team. (Pettigrew 2015) estimates a Markov model value function with winning as reward function. With winning as the

Journal: 10618-DAMI Article No.: 0496 TYPESET DISK LE CP Disp.:2017/3/23 Pages: 23 Layout: Small-X

reward function, the value of a state for a team is the chance that the team wins the game if play starts from that state (within the look-ahead horizon).

Goals are the events with the most impact on winning chances. With scoring the next goal as the reward function, the value of a state for a team is the chance that the team scores the next goal (within the look-ahead horizon). This value function generalizes action values used previously in hockey analytics, such as Expected Goals (from a shot) (Tegen 2015; Hockey Graphs 2015). It is also comparable to the expected possession value (EPV) that (Cervone et al. 2014) define for their basketball Markov model, in that both measure the expected score in the current local play, rather than the global chance of winning the game overall.

6 Dynamic programming

We describe an algorithm for computing the value of a state. If each state appeared at most once between goals, we could compute the next goal value function for, say the home team, simply as follows: count the number of state occurrences that were followed by a home goal, and divide by the total number of all state occurrences. However, a state may appear any number of times between goals. That is, the transition graph may contain loops. In that case the value function can be computed by an iterative procedure known as dynamic programming (DP). The key observation is that any value function satisfies the *Bellman equation*:

$$V_T(s) = \sum_{s',a} P(s, a, s') \times V_T(s') \tag{1}$$

where $T = H, A$. Value iteration starts from an initial value assignment—typically 0 for each state—then applies the Bellman equation as an update equation. We compute the state values for both home and away teams. The iteration stops when the change in state value differential between the home and the away team falls below a convergence threshold. In our experiments, we use a relative convergence of 0.1% as the convergence threshold. Algorithm 1 provides pseudo code.

Discussion We use an undiscounted reward function for our value iteration (Schwartz 1993). In game terms, the convergence criterion can be interpreted as follows. The quantity $V^{\ell}_{H}(s) - V^{\ell}_{T}(s)$ measures the home team advantage in state s, given a look-ahead horizon ℓ. (A negative advantage represents a disadvantage.) We stop value iteration when the algorithm's estimation of the home team advantage has converged, relative to our tolerance. This is appropriate for a zero-sum game because the aim of the game is to achieve an advantage over the opponent.

With this convergence criterion, DP converges within 14 steps for the next goal reward function. This means that looking ahead more than 14 steps changes the estimate of the how much more likely the home team is to score the next goal than the away team, by less than 0.1%. For the Win reward function, convergence requires 1,950 iterations until the win probability differential falls below 0.1%. This is because the algorithm looks ahead to the complete end of the game rather than just to the next goal.

Journal: 10618-DAMI Article No.: 0496 TYPESET DISK LE CP Disp.:2017/3/23 Pages: 23 Layout: Small-X

Algorithm 1 Dynamic Programming for Value Iteration for the reward differential function R.

Require: Markov Game model, convergence criterion c

$InitialValue = 0$
for $\ell = 1; \ell \leftarrow \ell + 1$ **do**
  for all states s in the Markov Game model **do**
    for all teams $T = H, A$ **do**
      $V_T^{\ell+1}(s) :=$
      $\sum_a R_T(s,a) + \sum_{s'} \frac{Occ(s,a,s')}{Occ(s,a)} \times V_T^{\ell}(s')$
    end for
  end for
  if $|[V_H^{\ell+1}(s) - V_T^{\ell+1}(s)] - [V_H^{\ell}(s) - V_T^{\ell}(s)]| < c$ for all states s **then**
    EXIT
  end if
end for

7 Examples

We provide some examples of how our Markov Game model evaluates actions in context. This illustrates our results for the hockey domain. For readers less familiar with dynamic models, the examples illustrate the concept of look-ahead.

7.1 High-value trajectory through the state space

The model can be used for a planning task: to find successor states that are most likely to lead to a goal. Figure 3 shows a high-value trajectory where a state is connected to its highest value successor (for the Home team).

7.2 Action values

We can evaluate an action in a context-aware way by considering its expected reward after executing it in a given state. This is known in reinforcement learning as the action value, or *Q-value* (Sutton and Barto 1998):

$$Q_T(s,a) \equiv \sum_{s'} P(s,a,s') \times V_T(s) \tag{2}$$

where $T = H, A$. As with state values, we can also consider *conditional Q-values*. The conditional Q-value for the home team is the Q-value for the home team performing an action, divided by the sum of the Q-value for the home team and the Q-value for the away team. Similarly for the away team.

To obtain a single value of an action at a location, we average over some or all of the context features. For example, to evaluate the chance of scoring from a shot location, we average over all shots at that location in all periods, and at all manpower levels. In the following we provide 2D gray scale plots of the value surface for a given

Journal: 10618-DAMI Article No.: 0496 TYPESET DISK LE CP Disp.:2017/3/23 Pages: 23 Layout: Small-X

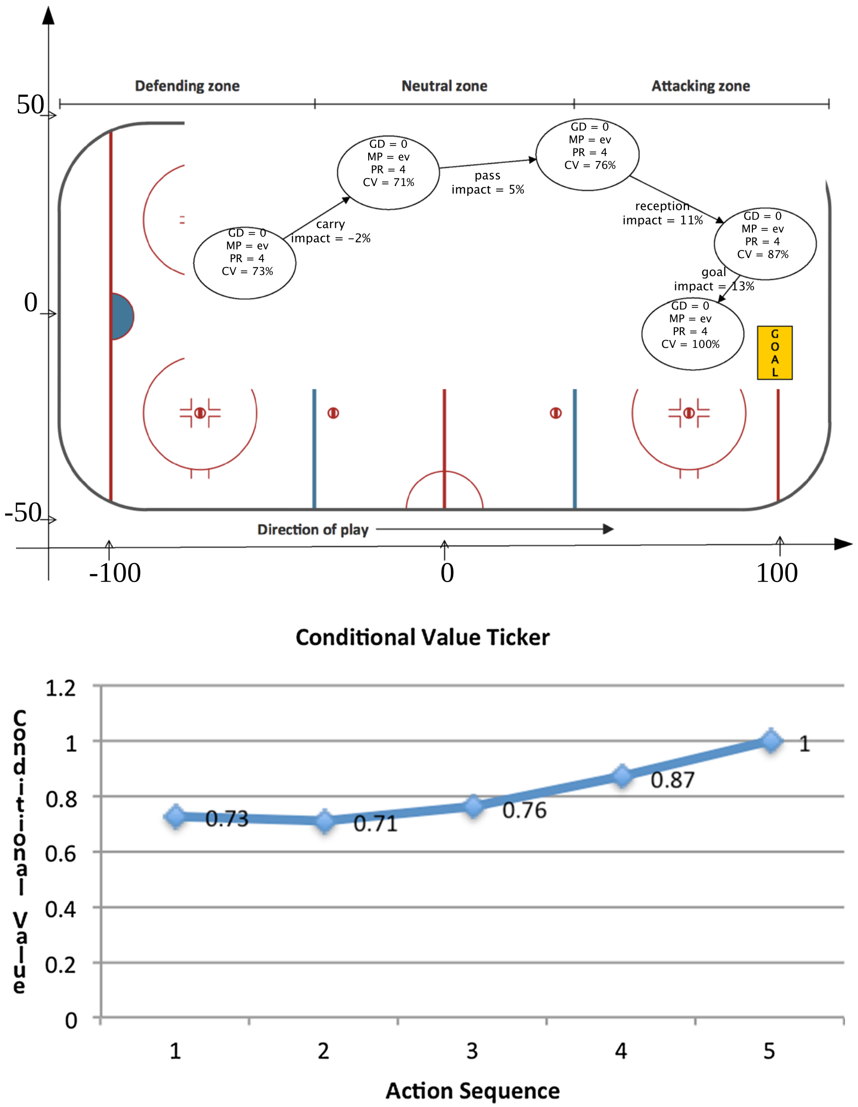

Fig. 3 A high-value trajectory that represents one generic successful play by the home team. *Top* we show a rink outline for orientation. Location coordinates are rescaled to bring nodes closer for legibility. A node represents a game state. Nodes are labelled as GD (goal difference), *MP* (manpower), PR (period), and CV (conditional value of a state for the home team, i.e., probability that the home team scores the next goal, see below). *Edges* are labelled with actions, and with the impact of the action, which is the difference in conditional probabilities (see Sect. 8). *Bottom* the same trajectory using a value-ticker format (Cervone et al. 2014), where the conditional value of the current state is plotted against events in temporal sequence. The horizontal distance between points is the impact of the corresponding action

Journal: 10618-DAMI Article No.: 0496 TYPESET DISK LE CP Disp.:2017/3/23 Pages: 23 Layout: Small-X

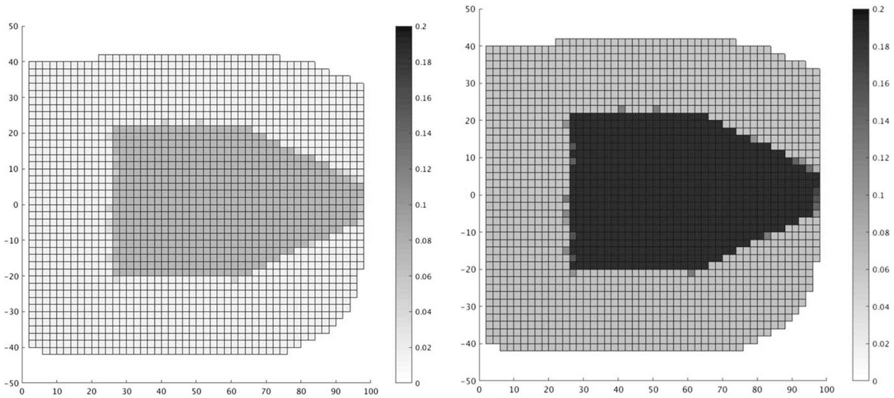

Fig. 4 Average values of shots depending on location. The value indicates the probability that a shot will be followed by a goal within look-ahead $\ell = 1$ or 14 steps. It is averaged over the context features (goal differential, manpower, period) and the shooting team (Home or Away). Locations in the same cluster are treated as equivalent by the Markov Game model, and therefore are assigned the same value. *Left* the highest chance of scoring immediately on a shot is from the slot (6%). The clusters around the slot carry a chance of about 1.5%. *Right* the chance of scoring after a shot within our look-ahead horizon of 14 steps. This includes immediate goals as well as rebounds. The medium-term probability is greater, around 16% for shots from the slot, and around 6% for the other clusters

location. The *x*-*y* coordinates in these plots are *adjusted coordinates*, so the defense zone locations are assigned negative x-values.

7.2.1 Shots

Figure 4 shows the values of shots from different locations, with look-ahead $\ell =$ 1, i.e., considering immediately following goals only, and $\ell = 14$, our maximum look-ahead.

7.2.2 Loose puck recoveries

Figure 5 shows the average location values for *loose puck recoveries*. Loose puck recovery is a frequent event in the Sportlogiq dataset. It is an interesting event to consider, because it is fairly neutral with respect to goal scoring, so approaches that consider immediate goal consequences only find it difficult to assign a value to it. In our model, with look-ahead $\ell = 1$, the expected reward from loose puck recovery vanishes almost everywhere; see Fig. 5 (top right) (zero except for two clusters with value 0.0002 and 0.0001). Figure 5 (left) shows that with look-ahead $\ell = 14$, there is considerable variation in the values for different location. We see how the action value changes depending on whether the defending or the attacking team recovers the loose puck: The conditional chance of scoring the next goal for the attacking team is around 73% in the right bottom of the offensive zone (Cluster 4 in Fig. 2c). For the defending team, in the same physical location, it is around 53%. So by beating the attacking team to a loose puck recovery, the defending team increases its chance of scoring the next goal by 20%. This illustrates an important way in which the model

Journal: **10618-DAMI** Article No.: **0496** ☐ TYPESET ☐ DISK ☐ LE ☐ CP Disp.:**2017/3/23** Pages: **23** Layout: **Small-X**

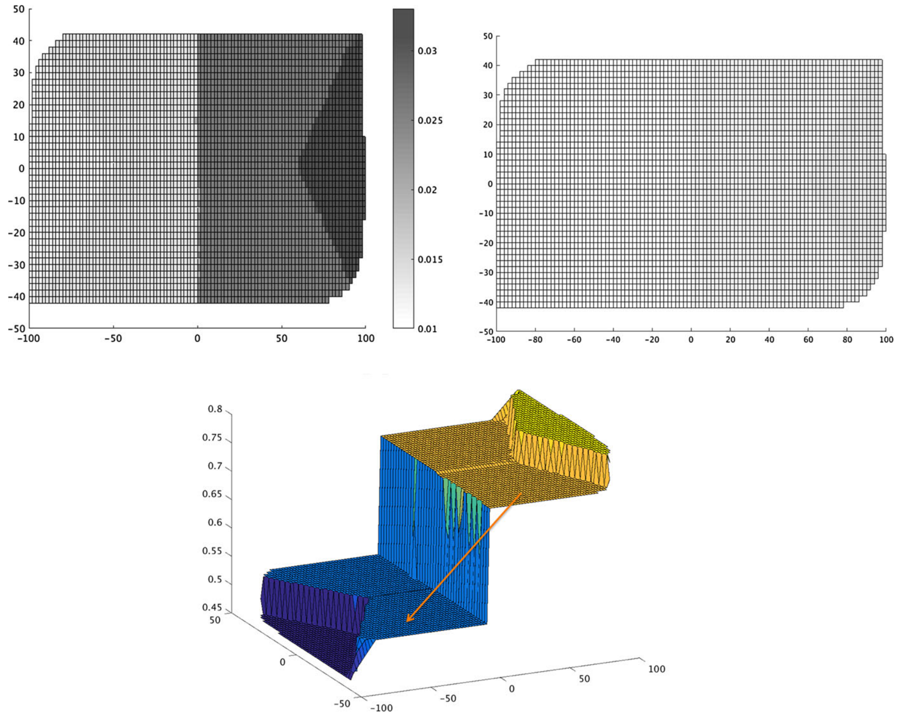

Fig. 5 *Left* average value of loose puck recovery depending on location with look-ahead $\ell = 14$. *Right* average value of loose puck recovery depending on location with $\ell = 1$. *Bottom* average *conditional* value of loose puck recovery depending on location with $\ell = 14$. The *arrow* connects the centre of Cluster 4 with the centre of Cluster 2. Using adjusted coordinates, it connects the same physical *x*–*y* location, but with different values depending on whether the defending team or the attacking team manages the loose puck recovery

captures the value of defensive actions: The chance that the defending team scores the next goal increases by decreasing the chance that the attacking team scores the next goal.

7.2.3 Dumping in the puck versus carrying the puck

Figure 6 shows two important offensive strategies: dumping the puck in (towards the goal) versus carrying it (towards the goal).

The comparative merits of each strategy are debated by hockey analysts (Tulsky 2013). Although these are two different actions, the expected reward provides a common scale on which they can be meaningfully compared; see Fig. 7. Generally carrying is more likely to lead to a successful attack: the conditional value for carrying ranges from 64 to 76%, whereas for dumping the puck it is fairly constant around 56%. This agrees qualitatively with previous analyses that observe a higher correlation between carrying and success than between dumping in and success (Tulsky 2013). Some analysts conclude that players are too conservative: they prefer dumping in the puck

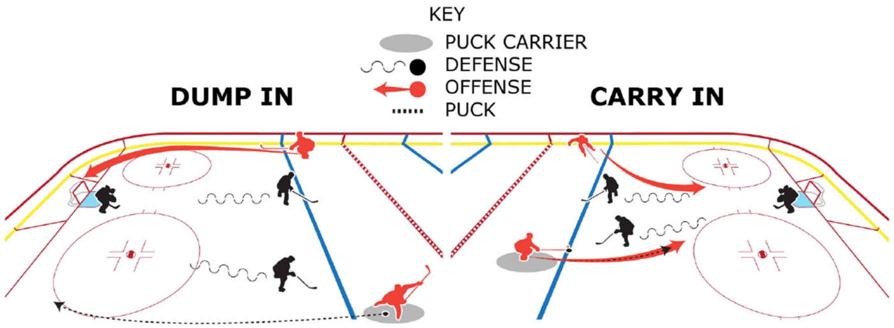

Fig. 6 Getting into the offensive zone: dumping versus carrying the puck. Image by Shaun Kreider, Kreider Designs

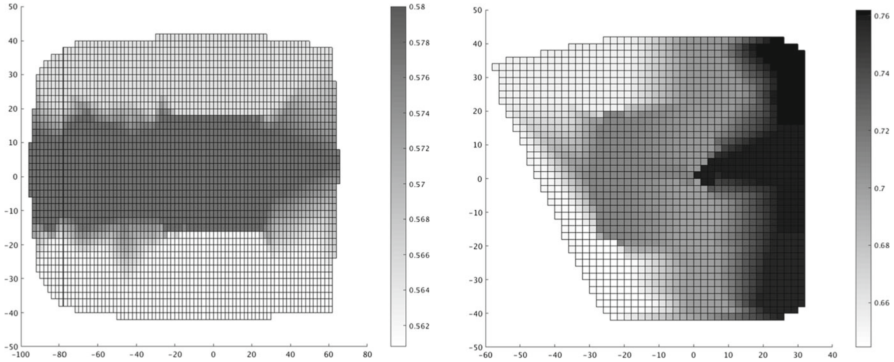

Fig. 7 Average conditional value of dumping in the puck (*left*) versus carrying it (*right*)

towards the goal, rather than carrying it and losing possession. Another hypothesis is that players carry the puck when the opposing defence-men are not well positioned to stop them. Investigating this hypothesis requires tracking data that record the position of all players at a given point.

8 Team performance

We validate our model by showing that it can be used to evaluate team performance. We examine two approaches, first, team performance as an aggregate of action values, and second, team performance as an aggregate of state values.

8.1 Team performance and action values

The general idea is that team strength can be estimated as an aggregate of individual action values, and then correlated with ground truth metrics such as numbers of goals scored. For example, the Expected Goal metric scores each shot according to its chance of leading to a goal (immediately) (Tegen 2015). The sum of goal probabilities for each

Journal: 10618-DAMI Article No.: 0496 TYPESET DISK LE CP Disp.:2017/3/23 Pages: 23 Layout: Small-X

shot is the team's Expected Goal metric. This metric was recently used by the Financial Times to analyse the reasons for Jose Mourinho's departure from the Chelsea Premier League team. The Expected Goal metric is a special case of our value function, where look-ahead $\ell = 1$ and the only action type taken into consideration is a shot.

Schulte and Routley used the *goal impact metric* to rank players (Routley and Schulte 2015). The impact of an action is defined by the equation

$$impact_T(s, a) \equiv Q_T(s, a) - V_T(s) \tag{3}$$

where $T = H, A$. When the reward function is defined by scoring the next goal, the goal impact measures how much a player's action changes the chance that his team scores the next goal, given the current state of the match. The *team goal impact* in a game is the sum of goal impacts over all actions in the match by the players of a team. We examine correlations between the following quantities associated with teams.

Average goal ratio For each game, the goal ratio for a team is #goals by team/#total number of goals by either side. For each team, we compute the average goal ratio over all games. Following (Tegen 2015), we use this as our main metric for team results.

Average team impact The average goal impact for a team, over all games.

Average team impact—look-ahead = 1 Average team impact using a value function with look-ahead = 1 step (rather than 14).

Average team impact—no location Average team impact where all locations are treated the same.

Average team value difference—look-ahead = 1 For each team, we compute the sum of action values $Q(s, a)$ for each action taken by the team in a game, minus the sum of action values taken by their opponent in the same game. Then we use the average of the value differences over all games.

In these computations, we removed the goals as actions, because our aim is to estimate how other actions predict the chance of a goal. We include the average team value difference—look-ahead = 1 because it is very similar to the Expected Goals Ratio metric recommended in (Tegen 2015). With look-ahead = 1, the action value sums are dominated by probability of a shot being successful, which is the basis of the Expected Goals Ratio metric.

8.1.1 Correlations on complete dataset

Figure 8 shows the scatter plots and correlation coefficients ρ between goal ratio and the Impact metrics. *The full team impact metric shows an impressive correlation of 0.7*. Figure 8 plots the datapoints with a linear fit trend line. Reducing the look-ahead to only 1 step drastically reduces the information that the impact metric provides about a team's result, down to a mere 0.09 correlation. Without the location information, the impact metric still gives meaningful results, but is much less informative with a correlation of 0.21. The value difference is in fact more informative with single-step look-ahead, at a correlation of 0.34. The magnitude of this correlation is similar to that found for the Expected Goals Ratio in soccer (Tegen 2015; Hockey Graphs 2015).

Journal: 10618-DAMI Article No.: 0496 TYPESET DISK LE CP Disp.:2017/3/23 Pages: 23 Layout: Small-X

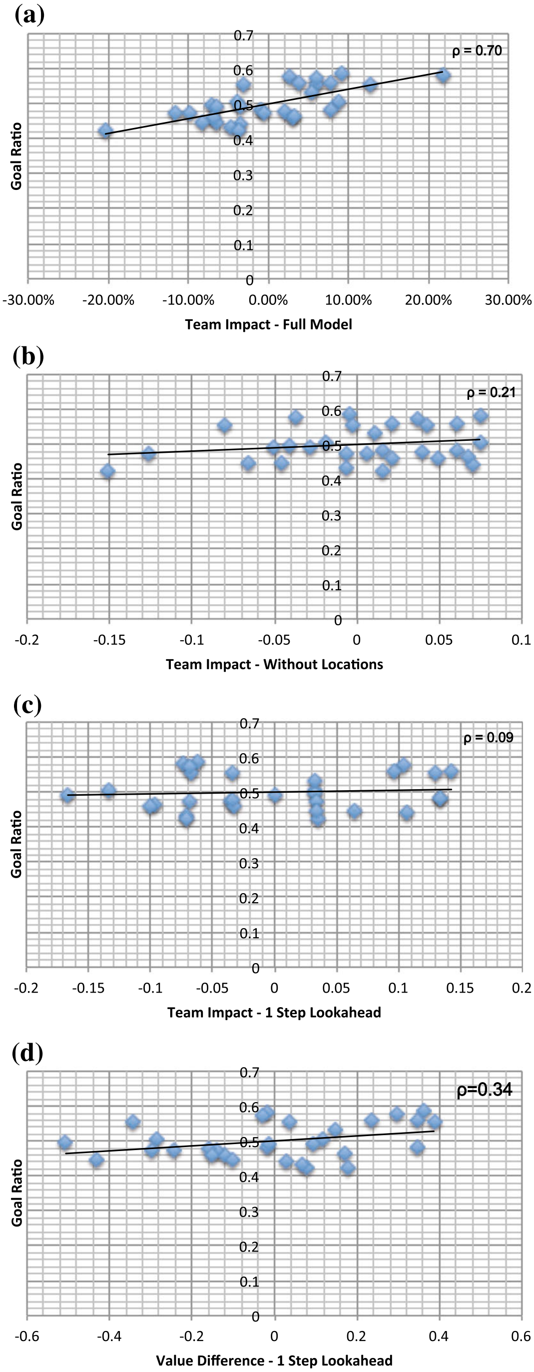

Fig. 8 Average team impact and average team value difference versus average goal ratio. Each datapoint represents a team. Action values are computed using the next goal reward function. **a** Full model: look-ahead = 14, location information. **b** Full look-ahead without location information. **c** 1-step look-ahead. **d** Average value difference

Journal: **10618-DAMI** Article No.: **0496** TYPESET DISK LE CP Disp.:**2017/3/23** Pages: **23** Layout: **Small-X**

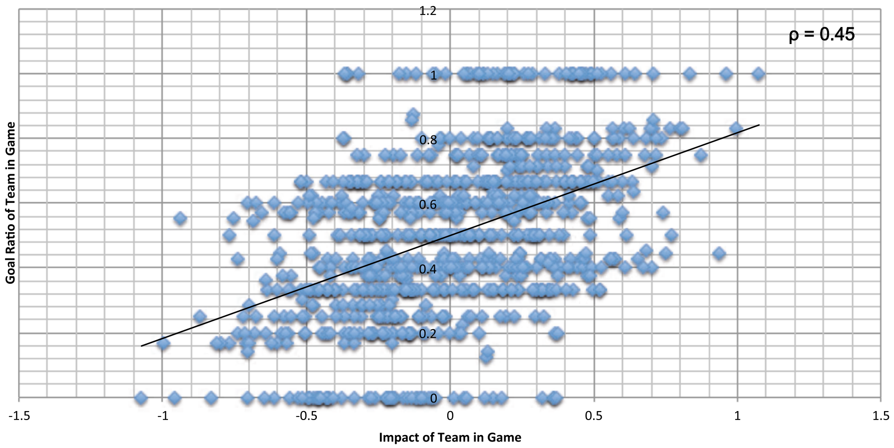

Fig. 9 A team's total impact values in a single match versus the goal ratio result in that match. Each datapoint represents a pair (team-match)

Overall, our conclusion is that the full team impact metric manages to extract by far the most information relevant to predicting a team's performance.

8.1.2 Correlations on incomplete datasets

Single game outcomes Figure 9 shows the scatter plot for the correlation between goal ratio and the Team Impact in an individual match. The correlation coefficient is $\rho = 0.45$. This shows that impact values carry substantial information about the outcome of a match. An application may be to live betting during a match: The impact value total after, say, the first two periods may be good predictors of the final match outcome.

8.1.3 Final goal ratio from initial observations

We also consider correlations computed not from the entire dataset, but from initial observations. Figure 10 gives the correlation coefficients between the final goal ratio, averaged over all matches, and a moving average of the Team Impact Total, averaged over the first *k* matches. The correlation is close to 0.5 after 10 matches, which is less than half the number of total matches for all teams in our dataset. This means that after a relatively small number of observed matches, the average team impact carries substantive information about the final goal ratio performance of a team.

8.2 Team performance and state values

Another plausible approach to measuring team performance using a Markov model is to aggregate the team values for the states reached during the match. The intuition is that good teams manage maintain an advantage, that is, a higher chance of winning,

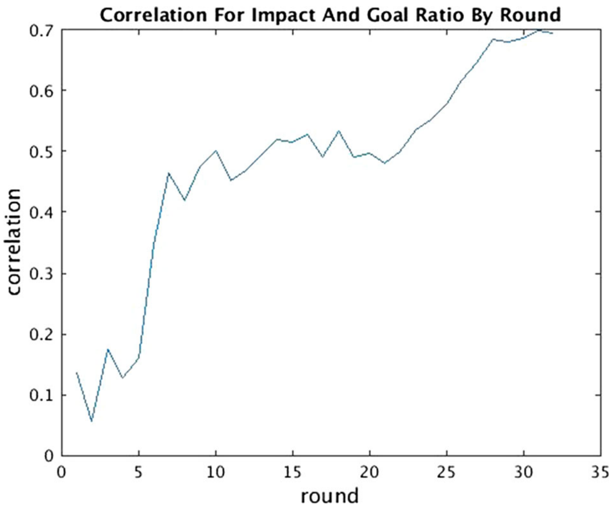

Fig. 10 The correlation between goal ratio (averaged over all matches) and team impact (averaged over the first k matches)

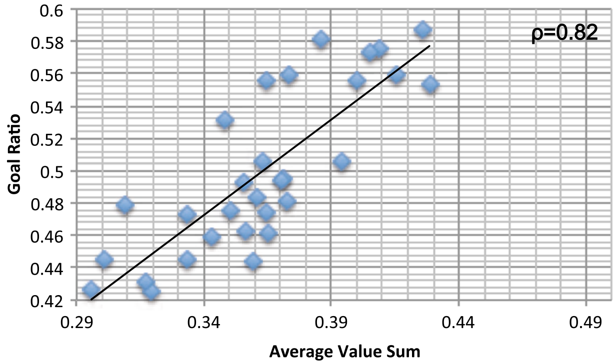

Fig. 11 The correlation between goal ratio (averaged over all matches) and the average winning chance metric, averaged over all matches. Each datapoint represents a team

for longer stretches than inferior teams. This motivates the *average team value sum* metric: For each game, sum the team values of the states reached during the match. Then average the summed values to obtain a team number for the whole dataset. With next goal as the reward function, the correlation with goal ratio is fairly weak, only 0.27. This is consistent with our previous result for the average team value difference (0.34): while achieving a relatively high likelihood of scoring is of course positively correlated with game outcome, it is a weak signal, compared with the ability to impact the scoring likelihood. However, this changes when we use Win as the reward function rather than next goal. We refer to this special case of the average team value sum as the average team winning chance metric. For this metric, a high state value reflects the global likelihood of success in the overall game, not mainly in the current play only.

Journal: **10618-DAMI** Article No.: **0496** TYPESET DISK LE CP Disp.:**2017/3/23** Pages: **23** Layout: **Small-X**

Figure 11 shows that the ability to achieve many states with higher winning chance correlates strongly with goal ratio, with a correlation coefficient of 0.82.

9 Conclusion

We have built a large Markov Game Model for a large set of NHL play-by-play events with a rich state space. A novel aspect of our dataset is that it includes location information about where an action took place.

Value iteration computes the values of each action—the value function of the Markov Game model. Compared to previous work that assigns a single value to actions, the Q-function incorporates two powerful sources of information for valuing hockey actions: (1) It takes into account the context of the action, represented by the Markov Game state. (2) It models the medium-term impact of an action by propagating its effect to future states. The value function provides knowledge about hockey dynamics by quantifying how much which action matters where. We apply our model to evaluate the performance of teams in terms of their actions' total impact on which team scores the next goal. The average impact score of a team for the next goal correlates highly with the team's average goal ratio ($\rho = 0.7$). An even higher correlation is achieved by the average team winning chance (0.82). In sum, the value function is a powerful AI concept that captures much information about hockey dynamics as the game is played in the NHL.

Future Work The NHL data provides a rich dataset for real-world event modelling. A number of further AI techniques can be applied to utilize even more of the available information than our Markov Game model does, especially for utilizing the spatio-temporal information. We mention two specific directions:

1. An *alternative discretization of locations* is to apply matrix factorization to a location transition matrix (Cervone et al. 2014). A cell in the transition matrix specifies the number of transitions from one location to another. (Given an initial fine-grained uniform discretization of locations to define rows and columns). Matrix factorization then produces a clustering of locations that models the information about which transitions are likely to occur: If two states are in the same cluster, their transition probabilities are the same. Our clustering based on location occurrence counts rather than transition counts is simpler and faster to compute, but does not utilize dynamic transition information.
2. We can utilize *process models with continuous time and space points*, rather than discretizing continuous quantities. Continuous quantities include (i) the time between events, (ii) the absolute game time of the events, (iii) unclustered location of actions [the location of shots has been considered in previous work (Krzywicki 2005)]. A promising model class are Piecewise Constant Conditional Intensity Models for continuous time event sequences (Gunawardana et al. 2011; Parikh et al. 2012). These models are especially well suited for sequences with a large set of possible events, such as our action events.

Our use of reinforcement learning techniques has been mainly for finding patterns in a rich data set, in the spirit of descriptive statistics and data mining. Another goal

Journal: 10618-DAMI Article No.: 0496 TYPESET DISK LE CP Disp.:2017/3/23 Pages: 23 Layout: Small-X

is to *predict* a player or team's future performance based on past performance using machine learning techniques. For example, can we predict a team's performance in the next match based on their performance in the previous ones? Or predict their final result in a match based on the first half of the match?

Acknowledgements This work was supported by an Engage Grant from the National Sciences and Engineering Council of Canada. We are grateful for constructive discussions in SFU's Sports Analytics Research Group.

References

Cervone D, D'Amour A, Bornn L, Goldsberry K (2014) Pointwise: predicting points and valuing decisions in real time with NBA optical tracking data. In: 8th Annual MIT sloan sports analytics conference, February, vol 28

Frey BJ, Dueck D (2007) Clustering by passing messages between data points. Science 315(5814):972–976

Gramacy R, Jensen S, Taddy M (2013) Estimating player contribution in hockey with regularized logistic regression. J Quant Anal Sports 9:97–111

Gunawardana A, Meek C, Xu P (2011) A model for temporal dependencies in event streams. In: Advances in neural information processing systems, pp 1962–1970

Hockey Graphs (2015) https://hockey-graphs.com/2015/10/01/expected-goals-are-a-better-predictor-of-future-scoring-than-corsi-goals/

Krzywicki K (2005) Shot quality model: a logistic regression approach to assessing NHL shots on goal. http://www.hockeyanalytics.com/Research_files/Shot_Quality_Krzywicki.pdf

Levesque H, Pirri F, Reiter R (1998) Foundations for the situation calculus. Linköping Electron Artic Comput Inf Sci 3(18):1–18

Littman ML (1994) Markov games as a framework for multi-agent reinforcement learning. In: Proceedings of the eleventh international conference on machine learning, vol 157, pp 157–163

Macdonald B (2011) An improved adjusted plus-minus statistic for NHL players. In: MIT sloan sports analytics conference 2011

National Hockey League (2014) National hockey league official rules 2014–2015

Parikh AP, Gunawardana A, Meek C (2012) Conjoint modeling of temporal dependencies in event streams. In: UAI Bayesian modelling applications workshop

Pettigrew S (2015) Assessing the offensive productivity of NHL players using in-game win probabilities. In: 9th annual MIT sloan sports analytics conference

Routley K, Schulte O (2015a) A Markov game model for valuing player actions in ice hockey. In: Uncertainty in artificial intelligence (UAI). pp 782–791

Routley K, Schulte O, Zhao Z (2015) Sports data mining. Springer, New York. Q-learning for the NHL. http://www.cs.sfu.ca/~oschulte/sports/

Russell S, Norvig P (2010) Artificial intelligence: a modern approach. Prentice Hall, Upper Saddle River, NJ

Schuckers M, Curro J (2013) Total hockey rating (THOR): a comprehensive statistical rating of national hockey league forwards and defensemen based upon all on-ice events. In: 7th annual MIT sloan sports analytics conference

Schumaker RP, Solieman OK, Chen H (2010) Sports data mining. Springer, New York

Schwartz A (1993) A reinforcement learning method for maximizing undiscounted rewards. In: Proceedings of the tenth international conference on machine learning, vol 298, pp 298–305

Spagnola N (2013) The complete plus-minus: a case study of the columbus blue jackets. Master's thesis, University of South Carolina

Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. MIT Press, Cambridge

Tegen E (2015) The best predictor for future performance is expected goals. http://11tegen11.net/2015/01/05/the-best-predictor-for-future-performance-is-expected-goals/

Thomas A, Ventura S, Jensen S, Ma S (2013) Competing process hazard function models for player ratings in ice hockey. Ann Appl Stat 7(3):1497–1524

Tulsky E (2013) Zone entries. http://nhlnumbers.com/2013/2/20/zone-entries-and-the-sloan-sports-analytics-conference

Journal: 10618-DAMI Article No.: 0496 TYPESET DISK LE CP Disp.:2017/3/23 Pages: 23 Layout: Small-X