

# Deep Reinforcement Learning in Ice Hockey for Context-Aware Player Evaluation

Guiliang Liu and Oliver Schulte  
Simon Fraser University, Burnaby, Canada  
gla68@sfu.ca, oschulte@cs.sfu.ca

## Abstract

A variety of machine learning models have been proposed to assess the performance of players in professional sports. However, they have only a limited ability to model how player performance depends on the game context. This paper proposes a new approach to capturing game context: we apply Deep Reinforcement Learning (DRL) to learn an action-value Q function from 3M play-by-play events in the National Hockey League (NHL). The neural network representation integrates both continuous context signals and game history, using a possession-based LSTM. The learned Q-function is used to value players’ actions under different game contexts. To assess a player’s overall performance, we introduce a novel Game Impact Metric (GIM) that aggregates the values of the player’s actions. Empirical Evaluation shows GIM is consistent throughout a play season, and correlates highly with standard success measures and future salary.

## 1 Introduction: Valuing Actions and Players

With the advancement of high frequency optical tracking and object detection systems, more and larger event stream datasets for sports matches have become available. There is increasing opportunity for large-scale machine learning to model complex sports dynamics. Player evaluation is a major task for sports modeling that draws attention from both fans and team managers, who want to know which players to draft, sign or trade. Many models have been proposed [Buttrey *et al.*, 2011; Macdonald, 2011; Decroos *et al.*, 2018; Kaplan *et al.*, 2014]. The most common approach has been to quantify the value of a player’s action, and to evaluate players by the total value of the actions they took [Schuckers and Curro, 2013; McHale *et al.*, 2012].

However, traditional sports models assess only the actions that have immediate impact on goals (e.g. shots), but not the actions that lead up to them (e.g. pass, reception). And action values are assigned taking into account only a limited context of the action. But in realistic professional sports, the relevant context is very complex, including game time, position of players, score and manpower differential, etc.

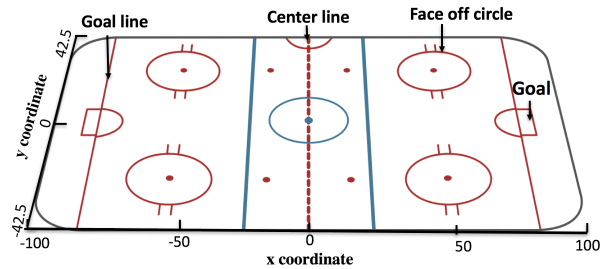


Figure 1: Ice Hockey Rink. **Ice hockey** is a fast-paced team sport, where two teams of skaters must shoot a puck into their opponent’s net to score goals.

Recently, Markov models have been used to address these limitations. [Routley and Schulte, 2015] used states of a Markov Game Model to capture game context and compute a Q function, representing the chance that a team scores the next goal, for *all* actions. [Cervone *et al.*, 2014] applied a competing risk framework with Markov chain to model game context, and developed EPV, a point-wise conditional value similar to a Q function, for each action. The Q-function concept offers two key advantages for assigning values to actions [Schulte *et al.*, 2017a; Decroos *et al.*, 2018]: 1) All actions are scored on the same scale by looking ahead to expected outcomes. 2) Action values reflect the match context in which they occur. For example, a late check near the opponent’s goal generates different scoring chances than a check at other locations and times.

The states in the previous Markov models represent only a partial game context in the real sports match, but nonetheless the models assume full observability. Also, they pre-discretized input features, which leads to loss of information. In this work, we utilize a deep reinforcement learning (DRL) model to learn an action-value Q function for capturing the current match context. The neural network representation can easily incorporate continuous quantities like rink location and game time. To handle partial observability, we introduce a possession-based Long Short Term Memory (LSTM) architecture that takes into account the current play history. Unlike most previous work on active reinforcement learning (RL), which aims to compute *optimal strategies* for complex continuous-flow games [Hausknecht and Stone, 2015;

Mnih *et al.*, 2015], we solve a prediction (not control) problem in the passive learning (on policy) setting [Sutton and Barto, 1998]. We use RL as a behavioral analytics tool for real human agents, not to control artificial agents.

Given a Q-function, the *impact* of an action is the change in Q-value due to the action. Our novel Goal Impact Metric (GIM) aggregates the impact of all actions of a player. To our knowledge, this is the first player evaluation metric based on DRL. The GIM metric measures both players’ offensive and defensive contribution to goal scoring. For player evaluation, similar to clustering, ground truth is not available. A common methodology [Routley and Schulte, 2015; Pettigrew, 2015] is to assess the predictive value of a player evaluation metric for standard measures of success. Empirical comparison between 7 player evaluation metrics finds that 1) given a complete season, GIM correlates the most with 12 standard success measures and is the most temporally consistent metric, 2) given partial game information, GIM generalizes best to future salary and season total success.

## 2 Related Work

We discuss the previous work most related to our approach.

*Deep Reinforcement Learning.* Previous DRL work has focused on *control* in continuous-flow games, not prediction [Mnih *et al.*, 2015]. Among these papers, [Hausknecht and Stone, 2015] use a very similar network architecture to ours, but with a fixed trace length parameter rather than our possession-based method. Hausknecht and Stone find that for partially observable control problems, the LSTM mechanism outperforms a memory window. Our study confirms this finding in an on policy prediction problem.

*Player Evaluation.* Albert *et al.* 2017 provide several up-to-date survey articles about evaluating players. A fundamental difficulty for action value counts in continuous-flow games is that they traditionally have been restricted to goals and actions immediately related to goals (e.g. shots). The Q-function solves this problem by using lookahead to assign values to all actions.

*Player Evaluation with Reinforcement Learning.* Using the Q-function to evaluate players is a recent development [Schulte *et al.*, 2017a; Cervone *et al.*, 2014; Routley and Schulte, 2015]. Schulte *et al.* discretized location and time coordinates and applied dynamic programming to learn a Q-function. Discretization leads to loss of information, undesirable spatio-temporal discontinuities in the Q-function, and generalizes poorly to unobserved parts of the state space. For basketball, Cervone *et al.* defined a player performance metric based on an expected point value model that is equivalent to a Q-function. Their approach assumes complete observability (of all players at all times), while our data provide partial observability only.

## 3 Task Formulation and Approach

Player evaluation (the “Moneyball” problem) is one of the most studied tasks in sports analytics. Players are rated by their observed performance over a set of games. Our approach to evaluating players is illustrated in Figure 2. Given

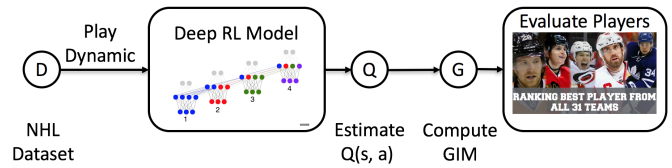


Figure 2: System Flow for Player Evaluation

dynamic game tracking data, we apply Reinforcement Learning to estimate the *action value* function  $Q(s, a)$ , which assigns a value to action  $a$  given game state  $s$ . We define a new player evaluation metric called **Goal Impact Metric (GIM)** to value each player, based on the aggregated impact of their actions, which is defined in Section 6 below. Player evaluation is a descriptive task rather than a predictive generalization problem. As game event data does not provide a ground truth rating of player performance, our experiments assess player evaluation as an unsupervised problem in Section 7.

## 4 Play Dynamic in NHL

We utilize a **dataset** constructed by SPORTLOGiQ using computer vision techniques. The data provide information about *game events* and *player actions* for the entire 2015-2016 NHL (largest professional ice hockey league) season, which contains 3,382,129 events, covering 30 teams, 1140 games and 2,233 players. Table 1 shows an excerpt. The data track events around the puck, and record the identity and actions of the player in possession, with space and time stamps, and features of the game context. The table utilizes adjusted spatial coordinates where negative numbers refer to the defensive zone of the acting player, positive numbers to his offensive zone. Adjusted X-coordinates run from -100 to +100, Y-coordinates from 42.5 to -42.5, and the origin is at the ice center as in Figure 1. We augment the data with derived features in Table 2 and list the complete feature set in Table 3.

We apply the Markov Game framework [Littman, 1994] to learn an action value function for NHL play. Our notation for RL concepts follows [Mnih *et al.*, 2015]. There are two agents *Home* resp. *Away* representing the home resp. away team. The **reward**, represented by goal vector  $g_t$  is a 1-of-3 indicator vector that specifies which team scores (*Home*, *Away*, *Neither*). An **action**  $a_t$  is one of 13 types, including shot, block, assist, etc., together with a mark that specifies the team executing the action, e.g. *Shot(Home)*. An **observation** is a feature vector  $\mathbf{x}_t$  for discrete time step  $t$  that specifies a value for the 10 features listed in Table 3. We use the complete sequence  $s_t \equiv (\mathbf{x}_t, a_{t-1}, \mathbf{x}_{t-1}, \dots, \mathbf{x}_0)$  as the state representation at time step  $t$  [Mnih *et al.*, 2015], which satisfies the Markov property.

We divide NHL games into **goal-scoring episodes**, so that each episode 1) begins at the beginning of the game, or immediately after a goal, and 2) terminates with a goal or the end of the game. A **Q function** represents the conditional probability of the event that the home resp. away team scores the goal at the end of the current episode (denoted  $goal_{Home} = 1$  resp.  $goal_{Away} = 1$ ), or neither team does (denoted  $goal_{Neither} = 1$ ):

GID=GameId, PID=playerId, GT=GameTime, TID=TeamId, MP=Manpower, GD=Goal Difference, OC = Outcome, S=Succeed, F=Fail, P = Team Possess puck, H=Home, A=Away, H/A=Team who performs action, TR = Time Remain, PN = Play Number, D = Duration

GID	PID	GT	TID	X	Y	MP	GD	Action	OC	P	Velocity	TR	D	Angle	H/A	PN
1365	126	14.3	6	-11.0	25.5	Even	0	Lpr	S	A	(-23.4, 1.5)	3585.7	3.4	0.250	A	4
1365	126	17.5	6	-23.5	-36.5	Even	0	Carry	S	A	(-4.0, -3.5)	3582.5	3.1	0.314	A	4
1365	270	17.8	23	14.5	35.5	Even	0	Block	S	A	(-27.0, -3.0)	3582.2	0.3	0.445	H	4
1365	126	17.8	6	-18.5	-37.0	Even	0	Pass	F	A	(0, 0)	3582.2	0.0	0.331	A	4
1365	609	19.3	23	-28.0	25.5	Even	0	Lpr	S	H	(-30.3, -7.5)	3580.6	1.5	0.214	H	5
1365	609	19.3	23	-28.0	25.5	Even	0	Pass	S	H	(0,0)	3580.6	0.0	0.214	H	5

Table 1: Dataset Example

Table 2: Derived Features

Name	Type	Range
X Coordinate of Puck	Continuous	[-100, 100]
Y Coordinate of Puck	Continuous	[-42.5, 42.5]
Velocity of Puck	Continuous	(-inf, +inf)
Game Time Remain	Continuous	[0, 3600]
Score Differential	Discrete	(-inf, +inf)
Manpower Situation	Discrete	{EV, SH, PP}
Event Duration	Continuous	[0, +inf)
Action Outcome	Discrete	{successful, failure}
Angle between puck and goal	Continuous	[-3.14, 3.14]
Home or Away Team	Discrete	{Home, Away}

Table 3: Complete Feature List

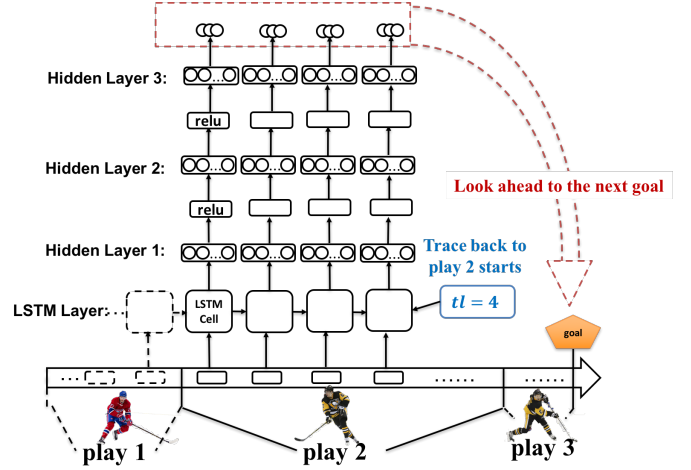


Figure 3: Our design is a 5-layer network with 3 hidden layers. Each hidden layer contains 1000 nodes, which utilize a relu activation function. The first hidden layer is the LSTM layer, the remaining layers are fully connected. Temporal-difference learning looks ahead to the next goal, and the LSTM memory traces back to the beginning of the play (the last possession change).

$$Q_{team}(s, a) = P(goal_{team} = 1 | s_t = s, a_t = a)$$

where *team* is a placeholder for one of *Home, Away, Neither*. This  $Q$ -function represents the probability that a team scores the next goal, given current play dynamics in the NHL (cf. Schulte *et al.*; Routley and Schulte). Different  $Q$ -functions for different expected outcomes have been used to capture different aspects of NHL play dynamics, such as match win [Pettigrew, 2015; Kaplan *et al.*, 2014; Routley and Schulte, 2015] and penalties [Routley and Schulte, 2015]. For player evaluation, the next-goal  $Q$  function has three advantages. 1) The next-goal reward captures what a coach expects from a player. For example, if a team is ahead by two goals with one minute left in the match, a player’s actions have negligible effect on final match outcome. Nonetheless professionals should keep playing as well as they can and maximize the scoring chances for their own team. 2) The  $Q$ -values are easy to interpret, since they model the probability of an event that is a relatively short time away (compared to final match outcome). 3) Increasing the probability that a player’s team scores the next goal captures both offensive and defensive value. For example, a defensive action like blocking a shot decreases the probability that the other team will score the next goal, thereby increasing the probability that the player’s own team will score the next goal.

## 5 Learning Q values with DP-LSTM Sarsa

We take a function approximation approach and learn a neural network that represents the  $Q$ -function ( $Q_{team}(s, a)$ ).

### 5.1 Network Architecture

Figure 3 shows our model structure. Three output nodes represent the estimates  $\hat{Q}_{Home}(s, a)$ ,  $\hat{Q}_{Away}(s, a)$  and  $\hat{Q}_{Neither}(s, a)$ . Output values are normalized to probabilities. The  $\hat{Q}$ -functions for each team share weights. The network architecture is a Dynamic LSTM that takes as inputs a current sequence  $s_t$ , an action  $a_t$  and a dynamic trace length  $tl_t$ .<sup>1</sup>

### 5.2 Weight Training

We apply an on-policy Temporal Difference (TD) prediction method **Sarsa** [Sutton and Barto, 1998, Ch.6.4], to estimate  $Q_{team}(s, a)$  for the NHL play dynamics observed in our dataset. Weights  $\theta$  are optimized by minibatch gradient descent via backpropagation. We used batch size 32 (determined experimentally). The Sarsa gradient descent update at time step  $t$  is based on a squared-error loss function:

<sup>1</sup>We experimented with a single-hidden layer, but weight training failed to converge.

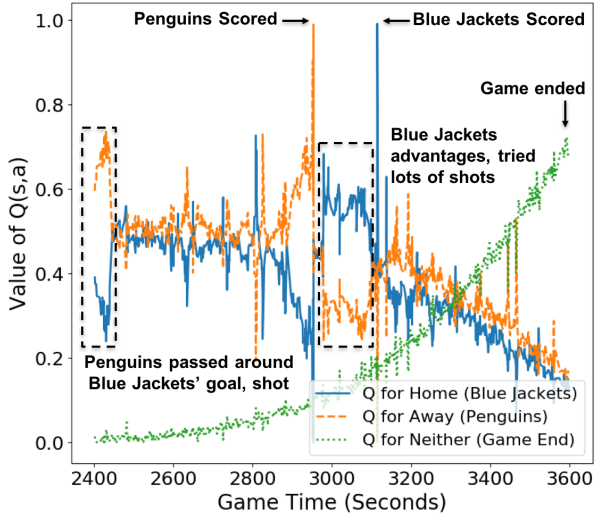


Figure 4: Temporal Projection of the method. For each team, and each game time, the graph shows the chance that that team scores the next goal, as estimated by the model. Major events lead to major changes in scoring chances, as annotated. The network also captures smaller changes associated with *every* action under different game contexts.

$$\mathcal{L}_t(\theta_t) = \mathbb{E}[(g_t + \hat{Q}(s_{t+1}, a_{t+1}, \theta_t) - \hat{Q}(s_t, a_t, \theta_t))^2]$$

$$\theta_{t+1} = \theta_t + \alpha \nabla_{\theta} \mathcal{L}(\theta_t)$$

where  $g$  and  $\hat{Q}$  are for a single team. LSTM training requires setting a *trace length*  $tl_t$  parameter. This key parameter controls how far back in time the LSTM propagates the error signal from the current time at the input history. Team sports like Ice Hockey show a turn-taking aspect where one team is on the offensive and the other defends; one such turn is called a *play*. We set  $tl_t$  to the number of time steps from current time  $t$  to the beginning of the current *play* (with a maximum of 10 steps). A play ends when the possession of puck changes from one team to another. Using possession changes as break points for temporal models is common in several continuous-flow sports, especially basketball [Cervone *et al.*, 2014; Omidiran, 2011]. We apply Tensorflow to implement training; our source code is published on-line.<sup>2</sup>

*Illustration of Temporal Projection.* Figure 4 shows a value ticker [Decroos *et al.*, 2017; Cervone *et al.*, 2014] that represents the evolution of the Q function from the 3<sup>rd</sup> period of a match between the Blue Jackets (Home team) and the Penguins (Away team), Nov. 17, 2015. The figure plots values of the three output nodes. We highlight critical events and match contexts to show the context-sensitivity of the Q function. High scoring probabilities for one team decrease those of its opponent. The probability that neither team scores rises significantly at the end of the match.

<sup>2</sup><https://github.com/Guiliang/DRL-ice-hockey>

## 6 Player Evaluation

In this section, we define our novel Goal Impact Metric and give an example player ranking.

### 6.1 Player Evaluation Metric

Our  $Q$ -function concept provides a novel AI-based definition for assigning a value to an action. Like [Schulte *et al.*, 2017b], we measure the quality of an action by how much it changes the expected return of a player’s team. Whereas the scoring chance at a time measures the value of a state, and therefore depends on the previous efforts of the entire team, the change in value measures directly the impact of an action by a specific player. In terms of the  $Q$ -function, this is the *change in  $Q$ -value* due to a player’s action. This quantity is defined as the action’s **impact**. The impact can be visualized as the difference between successive points in the  $Q$ -value ticker (Figure 4). For our specific choice of Next Goal as the reward function, we refer to **goal impact**. The total impact of a player’s actions is his **Goal Impact Metric (GIM)**. The formal equations are:

$$impact^{team}(s_t, a_t) = Q^{team}(s_t, a_t) - Q^{team}(s_{t-1}, a_{t-1})$$

$$GIM^i(D) = \sum_{s,a} n_D^i(s, a) \times impact^{team_i}(s, a)$$

where  $D$  indicates our dataset,  $team_i$  denotes the team of player  $i$ , and  $n_D^i(s, a)$  is the number of times that player  $i$  was observed to perform action  $a$  at  $s$ . Because it is the sum of differences between subsequent  $Q$  values, the GIM metric inherits context-sensitivity from the  $Q$  function.

### 6.2 Rank Players with GIM

Table 4 lists the top-20 highest impacts players, with basic statistics. All these players are well-known NHL stars. Taylor Hall tops the ranking although he did not score the most goals. This shows how our ranking, while correlated with goals, also *reflects the value of other actions by the player*. For instance, we find that the total number of passes performed by Taylor Hall is exceptionally high at 320. Our metric can be used to *identify undervalued players*. For instance, Johnny Gaudreau and Mark Scheifele drew salaries below what their GIM rank would suggest. Later they received a \$5M+ contract for the 2016-17 season.

## 7 Empirical Evaluation

We describe our comparison methods and evaluation methodology. Similar to clustering problems, there is *no ground truth* for the task of player evaluation. To assess a player evaluation metric, we follow previous work [Routley and Schulte, 2015; Pettigrew, 2015] and compute its correlation with statistics that directly measure success like Goals, Assists, Points, Play Time (Section 7.2). There are two justifications for comparing with *success measures*. (1) These statistics are generally recognized as important measures of a player’s strength, because they indicate the player’s ability to contribute to game-changing events. So a comprehensive performance metric ought to be related to them. (2) The success measures are



Name	GIM	Assists	Goals	Points	Team	Salary
Taylor Hall	96.40	39	26	65	EDM	\$6,000,000
Joe Pavelski	94.56	40	38	78	SJS	\$6,000,000
Johnny Gaudreau	94.51	48	30	78	CGY	\$925,000
Anze Kopitar	94.10	49	25	74	LAK	\$7,700,000
Erik Karlsson	92.41	66	16	82	OTT	\$7,000,000
Patrice Bergeron	92.06	36	32	68	BOS	\$8,750,000
Mark Scheifele	90.67	32	29	61	WPG	\$832,500
Sidney Crosby	90.21	49	36	85	PIT	\$12,000,000
Claude Giroux	89.64	45	22	67	PHI	\$9,000,000
Dustin Byfuglien	89.46	34	19	53	WPG	\$6,000,000
Jamie Benn	88.38	48	41	89	DAL	\$5,750,000
Patrick Kane	87.81	60	46	106	CHI	\$13,800,000
Mark Stone	86.42	38	23	61	OTT	\$2,250,000
Blake Wheeler	85.83	52	26	78	WPG	\$5,800,000
Tyler Toffoli	83.25	27	31	58	DAL	\$2,600,000
Charlie Coyle	81.50	21	21	42	MIN	\$1,900,000
Tyson Barrie	81.46	36	13	49	COL	\$3,200,000
Jonathan Toews	80.92	30	28	58	CHI	\$13,800,000
Sean Monahan	80.92	36	27	63	CGY	\$925,000
Vladimir Tarasenko	80.68	34	40	74	STL	\$8,000,000

Table 4: 2015-2016 Top-20 Player Impact Scores

often forecasting targets for hockey stakeholders, so a good player evaluation metric should have predictive value for them. For example, teams would want to know how many points an offensive player will contribute. To evaluate the ability of the GIM metric for generalizing from past performance to future success, we report two measurements: How well the GIM metric predicts a total season success measure from a sample of matches only (Section 7.3), and how well the GIM metric predicts the future salary of a player in subsequent seasons (Section 7.4). Mapping performance to salaries is a practically important task because it provides an objective standard to guide players and teams in salary negotiations [Idson and Kahane, 2000].

### 7.1 Comparison Player Evaluation Metrics

We compare GIM with the following player evaluation metrics to show the advantage of 1) modeling game context 2) incorporating continuous context signal 3) including history.

Our first baseline method **Plus-Minus (+/-)** is a commonly used metric that measures how the presence of a player influences the goals of his team [Macdonald, 2011]. The second baseline method **Goal-Above-Replacement (GAR)** estimates the difference of team’s scoring chances when the target player plays, vs. replacing him or her with an average player [Gerstenberg *et al.*, 2014]. **Win-Above-Replacement (WAR)**, our third baseline method, is the same as GAR but for winning chances [Gerstenberg *et al.*, 2014]. Our fourth baseline method **Expected Goal (EG)** weights each shot by the chance of it leading to a goal. These four methods consider only very limited game context. The last baseline method **Scoring Impact (SI)** is the most similar method to GIM based on Q-values. But Q-values are learned with pre-discretized spatial regions and game time [Schulte *et al.*, 2017a]. As a lesion method, we include **GIM-T1**, where we set the maximum trace length of LSTM to 1 (instead of 10) in computing GIM. This comparison assesses the importance of including enough history information.

*Computing Cost.* Compared to traditional metrics like +/-, learning a Q-function is computationally demanding (over 5 million gradient descent steps on our dataset). However, after

the model has been trained off-line, the GIM metric can be computed quickly with a single pass over the data.

*Significance Test.* To assess whether GIM is significantly different from the other player evaluation metrics, we perform paired t-tests over all players. The null hypothesis is rejected with respective p-values:  $1.1 * 10^{-186}$ ,  $7.6 * 10^{-204}$ ,  $8 * 10^{-218}$ ,  $3.9 * 10^{-181}$ ,  $4.7 * 10^{-201}$  and  $1.3 * 10^{-05}$  for PlusMinus, GAR, WAR, EG, SI and GIM-T1, which shows that GIM values are very different from other metrics’ values.

### 7.2 Season Totals: Correlations with standard Success Measures

In the following experiment, we compute the correlation between player ranking metrics and success measures over the entire season. Table 5 shows the correlation coefficients of the comparison methods with 14 standard success measures: Assist, Goal, Game Wining Goal (GWG), Overtime Goal (OTG), Short-handed Goal (SHG), Power-play Goal (PPG), Shots (S), Point, Short-handed Point (SHP), Power-play Point (PPP), Face-off Win Percentage (FOW), Points Per Game (P/GP), Time On Ice (TOI) and Penalty Minute (PIM). These are all commonly used measures available from the NHL official website ([www.nhl.com/stats/player](http://www.nhl.com/stats/player)). *GIM achieves the highest correlation in 12 out of 14 success measures.* For the remaining two (TOI and PIM), GIM is comparable to the highest. Together, the Q-based metrics GIM, GIM-1 and SI show the highest correlations with success measures. EG is only the fourth best metric, because it considers only the expected value of shots without look-ahead. The traditional sports analytics metrics correlate poorly with almost all success measures. This is evidence that AI techniques that provide fine-grained expected action value estimates lead to better performance metrics. With the neural network model, GIM can handle continuous input without pre-discretization. This prevents the loss of game context information and explains why both GIM and GIM-T1 performs better than SI in most success measures. And the higher correlation of GIM compared to GIM-T1 also demonstrates the value of game history. In terms of absolute correlations, GIM achieves high values, except for the very rare events OTG, SHG, SHP and FOW. Another exception is Penalty Minutes (PIM), which interestingly, show positive correlation with all player evaluation metrics, although penalties are undesirable. We hypothesize that better players are more likely to receive penalties, because they play more often and more aggressively.

### 7.3 Round-by-Round Correlations: Predicting Future Performance From Past Performance

A sports season is commonly divided into **rounds**. In round  $n$ , a team or player has finished  $n$  games in a season. For a given performance metric, we measure the correlation between (i) its value computed *over the first  $n$  rounds*, and (ii) the value of the three main success measures, assists, goals, and points, computed *over the entire season*. This allows us to assess how quickly different metrics acquire predictive power for the final season total, so that future performance can be predicted from past performance. We also evaluate the *auto-correlation* of a metric’s round-by-round total with its own season total. The auto-correlation is a measure of

methods	Assist	Goal	GWG	OTG	SHG	PPG	S
+/-	0.236	0.204	0.217	0.16	0.095	0.099	0.118
GAR	0.527	0.633	0.552	0.324	0.191	0.583	0.549
WAR	0.516	0.652	0.551	0.332	0.192	0.564	0.532
EG	0.783	0.834	0.704	0.448	0.249	0.684	0.891
SI	0.869	0.745	0.631	0.411	0.27	0.591	0.898
GIM-T1	0.873	0.752	0.682	0.428	0.291	0.607	0.877
<b>GIM</b>	<b>0.875</b>	<b>0.878</b>	<b>0.751</b>	<b>0.465</b>	<b>0.345</b>	<b>0.71</b>	<b>0.912</b>

methods	Point	SHP	PPP	FOW	P/GP	TOI	PIM
+/-	0.237	0.159	0.089	-0.045	0.238	0.141	0.049
GAR	0.622	0.226	0.532	0.16	0.616	0.323	0.089
WAR	0.612	0.235	0.531	0.153	0.605	0.331	0.078
EG	0.854	0.287	0.729	0.28	0.702	0.722	0.354
SI	0.869	0.37	0.707	0.185	0.655	<b>0.955</b>	<b>0.492</b>
GIM-T1	0.902	0.384	0.736	0.288	0.738	0.777	0.347
<b>GIM</b>	<b>0.93</b>	<b>0.399</b>	<b>0.774</b>	<b>0.295</b>	<b>0.749</b>	0.835	0.405

Table 5: Correlation with standard success measures.

temporal consistency, which is a desirable feature [Pettigrew, 2015], because generally the skill of a player does not change greatly throughout a season. Therefore a good performance metric should show temporal consistency.

We focused on the expected value metrics EG, SI, GIM-T1 and GIM, which had the highest correlations with success in Table 5. Figure 5 shows metrics’ round-by-round correlation coefficients with assists, goals, and points. The bottom right shows the auto-correlation of a metric’s round-by-round total with its own season total. *GIM is the most stable metric* as measured by auto-correlation: after half the season, the correlation between the round-by-round GIM and the final GIM is already above 0.9.

We find both GIM and GIM-T1 eventually dominate the predictive value of the other metrics, which shows the advantages of modeling sports game context without pre-discretization. And possession-based GIM also dominates GIM-T1 after the first season half, which shows the value of including play history in the game context. But how quickly and how much the GIM metrics improve depends on the specific success measure. For instance, in Figure 5, GIM’s round-by-round correlation with Goal (top right graph) dominates by round 10, while others require a longer time.

### 7.4 Future Seasons: Predicting Players’ Salary

In professional sports, a team will give a comprehensive evaluation to players before deciding their contract. The more value players provide, the larger contract they will get. Accordingly, a good performance metric should be positively related to the amount of players’ *future* contract. The NHL regulates when players can renegotiate their contracts, so we focus on players receiving a new contract following the games in our dataset (2015-2016 season).

Table 6 shows the metrics’ correlations with the amount of players’ contract over all the players who obtained a new contract during the 2016-17 and 2017-18 NHL seasons. Our GIM score achieves the highest correlation in both seasons. This means that the metric can serve as an objective basis for contract negotiations. The scatter plots of Figure 6 illustrate GIM’s correlation with amount of players’ future contract. In the 2016-17 season (left), we find many underestimated

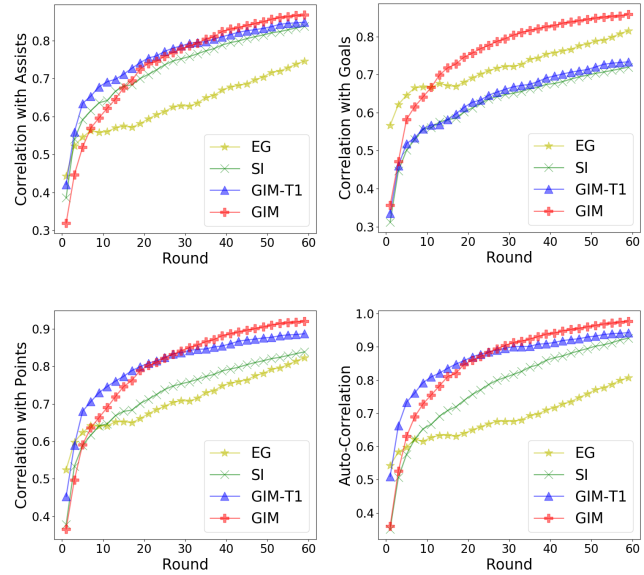


Figure 5: Correlations between round-by-round metrics and season totals.

methods	2016 to 2017 Season	2017 to 2018 Season
Plus Minus	0.177	0.225
GAR	0.328	0.372
WAR	0.328	0.372
EG	0.587	0.6
SI	0.609	0.668
GIM-T1	0.596	0.69
<b>GIM</b>	<b>0.666</b>	<b>0.763</b>

Table 6: Correlation with Players’ Contract

players in the right bottom part, with high GIM but low salary in their new contract. It is interesting that the percentage of players who are undervalued in their new contract decreases in the next season (from 32/258 in 2016-17 season to 8/125 in 2017-2018 season). This suggests that GIM provides an early signal of a player’s value after one season, while it often takes teams an additional season to recognize performance enough to award a higher salary.

## 8 Conclusion and Future Work

We investigated Deep Reinforcement Learning (DRL) for professional sports analytics. We applied DRL to learn complex spatio-temporal NHL dynamics. The trained neural network provides a rich source of knowledge about how a team’s chance of scoring the next goal depends on the match context. Based on the learned action values, we developed an innovative context-aware performance metric GIM that provides a comprehensive evaluation of NHL players, taking into account *all* of their actions. In our experiments, GIM had the highest correlation with most standard success measures, was the most temporally consistent metric, and generalized best to players’ future salary. Our approach applies to similar continuous-flow sports games with rich game contexts, like soccer and basketball. A limitation of our approach is

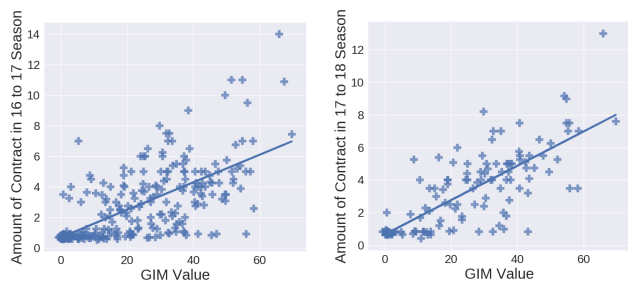


Figure 6: Player GIM vs. Value of new contracts in the 2016-17 (left) and 2017-18 (right) NHL season.

that players get credit only for recorded individual actions. An influential approach to extend credit to all players on the rink has been based on regression [Macdonald, 2011; Thomas *et al.*, 2013]. A promising direction for future work is to combine Q-values with regression.

## Acknowledgements

This work was supported by an Engage Grant from the National Sciences and Engineering Council of Canada, and a GPU donation from NVIDIA Corporation.

## References

[Albert *et al.*, 2017] Jim Albert, Mark E Glickman, Tim B Swartz, and Ruud H Koning. *Handbook of Statistical Methods and Analyses in Sports*. CRC Press, 2017.

[Buttrey *et al.*, 2011] Samuel Buttrey, Alan Washburn, and Wilson Price. Estimating NHL scoring rates. *Journal of Quantitative Analysis in Sports*, 7(3), 2011.

[Cervone *et al.*, 2014] Dan Cervone, Alexander D’Amour, Luke Bornn, and Kirk Goldsberry. Pointwise: Predicting points and valuing decisions in real time with NBA optical tracking data. In *MIT Sloan Sports Analytics Conference*, 2014.

[Decroos *et al.*, 2017] Tom Decroos, Vladimir Dzyuba, Jan Van Haaren, and Jesse Davis. Predicting soccer highlights from spatio-temporal match event streams. In *AAAI 2017*, pages 1302–1308, 2017.

[Decroos *et al.*, 2018] Tom Decroos, Lotte Bransen, Jan Van Haaren, and Jesse Davis. Actions speak louder than goals: Valuing player actions in soccer. *arXiv preprint arXiv:1802.07127*, 2018.

[Gerstenberg *et al.*, 2014] Tobias Gerstenberg, Tomer Ullman, Max Kleiman-Weiner, David Lagnado, and Josh Tenenbaum. Wins above replacement: Responsibility attributions as counterfactual replacements. In *Proceedings of the Cognitive Science Society*, volume 36, 2014.

[Hausknecht and Stone, 2015] Matthew Hausknecht and Peter Stone. Deep recurrent Q-learning for partially observable MDPs. *CoRR, abs/1507.06527*, 2015.

[Idson and Kahane, 2000] Todd L Idson and Leo H Kahane. Team effects on compensation: an application to salary determination in the National Hockey League. *Economic Inquiry*, 38(2):345–357, 2000.

[Kaplan *et al.*, 2014] Edward H Kaplan, Kevin Mongeon, and John T Ryan. A Markov model for hockey: Manpower differential and win probability added. *INFOR: Information Systems and Operational Research*, 52(2):39–50, 2014.

[Littman, 1994] Michael L Littman. Markov games as a framework for multi-agent reinforcement learning. In *Proceedings International Conference on Machine Learning*, volume 157, pages 157–163, 1994.

[Macdonald, 2011] Brian Macdonald. A regression-based adjusted plus-minus statistic for NHL players. *Journal of Quantitative Analysis in Sports*, 7(3):29, 2011.

[McHale *et al.*, 2012] Ian G McHale, Philip A Scarf, and David E Folker. On the development of a soccer player performance rating system for the English Premier League. *Interfaces*, 42(4):339–351, 2012.

[Mnih *et al.*, 2015] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[Omidiran, 2011] Dapo Omidiran. A new look at adjusted plus/minus for basketball analysis. In *MIT Sloan Sports Analytics Conference [online]*, 2011.

[Pettigrew, 2015] Stephen Pettigrew. Assessing the offensive productivity of NHL players using in-game win probabilities. In *MIT Sloan Sports Analytics Conference*, 2015.

[Routley and Schulte, 2015] Kurt Routley and Oliver Schulte. A Markov game model for valuing player actions in ice hockey. In *Proceedings Uncertainty in Artificial Intelligence (UAI)*, pages 782–791, 2015.

[Schuckers and Curro, 2013] Michael Schuckers and James Curro. Total hockey rating (THoR): A comprehensive statistical rating of national hockey league forwards and defensemen based upon all on-ice events. In *MIT Sloan Sports Analytics Conference*, 2013.

[Schulte *et al.*, 2017a] Oliver Schulte, Mahmoud Khademi, Sajjad Gholami, et al. A Markov game model for valuing actions, locations, and team performance in ice hockey. *Data Mining and Knowledge Discovery*, pages 1–23, 2017.

[Schulte *et al.*, 2017b] Oliver Schulte, Zeyu Zhao, Mehrsan Javan, and Philippe Desaulniers. Apples-to-apples: Clustering and ranking NHL players using location information and scoring impact. In *MIT Sloan Sports Analytics Conference*, 2017.

[Sutton and Barto, 1998] Richard S Sutton and Andrew G Barto. *Introduction to reinforcement learning*, volume 135. MIT Press Cambridge, 1998.

[Thomas *et al.*, 2013] A.C. Thomas, S.L. Ventura, S. Jensen, and S. Ma. Competing process hazard function models for player ratings in ice hockey. *The Annals of Applied Statistics*, 7(3):1497–1524, 2013.