Valuing Sports Actions and Players with Inverse Reinforcement Learning

Yudong Luo and **Oliver Schulte**
Simon Fraser University, Burnaby, Canada
yudong_luo@sfu.ca, oschulte@cs.sfu.ca

Abstract

A major problem of sports analytics is to rank players based on the impact of their actions. Recent player ranking methods have applied reinforcement learning (RL) to assess the value of action from a learned action-value or Q-function. This paper combines Q-function learning with inverse reinforcement learning (IRL) to provide a novel RL-based player ranking method that is especially effective for low-scoring games such as soccer and ice hockey. We propose to treat professional play as expert demonstrations for learning an implicit reward function. Domain knowledge about the rules of the game is represented by regularizing learned rewards with goal rewards. Learning is based on 4.5M play-by-play events in the National Hockey League (NHL). Empirical Evaluation indicates that player ranking based on learned rewards achieves high correlations with standard success measures and temporal consistency throughout a season.

Introduction: Valuing Actions and Players

A major task of sports statistics is player evaluation, which supports drafting, coaching, and trading decisions. The most common approach is to quantify the impact values for players' actions (Schuckers and Curro 2013; Routley and Schulte 2015; Schulte et al. 2017; Liu and Schulte 2018; Decroos et al. 2019). Whereas actions with immediate impact on goals, such as shots, are relatively easy to value, valuing actions with medium-term effects is challenging. Several RL models have been proposed to tackle this issue (Routley and Schulte 2015; Schulte et al. 2017; Liu and Schulte 2018). These RL models explicitly use goals as the reward signals. However, for sports with sparse goals, it is still the case that goals and actions closely connected to goals are assigned the largest impact values. Therefore, the performance evaluation is biased towards offensive players.

To tackle the sparse reward issue, we propose an inverse reinforcement learning method with domain knowledge (IRL-DK) to recover reward function for game dynamics. In IRL (Ng, Russell, and others 2000), agents are assumed to act by optimizing an unobserved internal reward function. The learning task is to estimate the agent's reward function from their observed behavior (demonstrations). Sports are different from the general IRL setting, because some aspects of a player's reward function can be inferred from domain knowledge. For instance, scoring a goal should have a relatively high reward because it helps the team to win a game. To benefit from both IRL and domain knowledge, we adopt transfer learning methods to combine the reward inferred from demonstrations and the one inferred from our domain knowledge. The final aggregated reward is used to calculate a Q-function, which measures the expected total reward from an action given a match state. As in previous RL work, the Q-function can be used to value actions and rank players. We apply IRL-DK to the 2018-19 play-by-play data in NHL. The resulting distribution of top players is mixed among offensive and defensive players rather than concentrated among offensive players. Empirical comparison among 7 player evaluation metrics shows the high correlations with standard success measures and temporal consistency of our method.

Markov Game Model

As an extension of game theory (Von Neumann and Morgenstern 1947) to Markov decision process (MDP), a Markov Game (Littman 1994) is defined by a set of states and a collection of action sets, one for each agent in the environment. Our Markov game model for ice hockey follows previous work (Routley and Schulte 2015). We treat home team H and away team A as two players in the game. At each timestamp, only one player performs an action, and the player not controlling the puck chooses no operation. Each ice hockey game is modeled as a semi-episodic task (Sutton and Barto 1998), where games switch from episode to episode. Each episode starts at the beginning of the game or right after a goal, and ends up with a goal or the end of the game. The transition function is calculated using the observed frequency $T(s,a,s') = p(s'|s,a) = O(s,a,s')/O(s,a)$, where $O(\cdot)$ counts the occurrence number in our dataset.

Similar to previous Markov models for ice hockey (Thomas et al. 2013; Routley and Schulte 2015; Schulte et al. 2017), we choose defining features for states, including game context, team identity (H/A) and location (L). Game context consists of Goal Difference (GD), ManPower (MP), and Period (P). GD is calculated as number of home goals minus number of away goals. MP specifies shorthanded, even strength, and powerplay. P represents the current period, ranging from 1 to 3. (We do not consider overtime play.) We divide hockey rink into 6 regions indexed by L based on the two blue lines to divide the X axis. We add an absorbing goal state for each team, with no transition out of it. There are total 27 actions recorded in our dataset, and home and away teams share the same action space.

IRL with Domain Knowledge

We first describe the MaxEnt IRL method before adding domain knowledge to it.

Maximum Entropy IRL

In the maximum entropy (MaxEnt) IRL (Ziebart et al. 2008), each state s is assigned a feature vector $\boldsymbol{f}_s \in \mathbb{R}^k$ and the reward function is parameterized as a linear function with reward weights $\boldsymbol{\theta} \in \mathbb{R}^k$ as $r_s = \boldsymbol{\theta}^T \boldsymbol{f}_s$. The reward value for a trajectory ζ_i is simply the cumulative reward of state,

$$r_{\zeta_i} = \sum_{s_j \in \zeta_i} \boldsymbol{\theta}^T \boldsymbol{f}_{s_j} = \boldsymbol{\theta}^T \boldsymbol{f}_{\zeta_i},$$

where $\boldsymbol{f}_{\zeta_i} = \sum_{s_j \in \zeta_i} \boldsymbol{f}_{s_j}$ is called the feature count of the trajectory. The observed feature counts are calculated as $\tilde{\boldsymbol{f}} = \frac{1}{n} \sum_i \boldsymbol{f}_{\zeta_i}$ where n is the number of trajectories.

Assume that agents act under a maximum entropy (Jaynes 1957) policy, the probability of a demonstrated trajectory ζ_i increases exponentially with higher rewards, so we have

$$P(\zeta_i|\boldsymbol{\theta}, T) = \frac{e^{r_{\zeta_i}}}{Z(\boldsymbol{\theta}, T)} \prod_{s_{t+1}, a_t, s_t \in \zeta_i} P_T(s_{t+1}|a_t, s_t) \quad (1)$$

where $Z(\boldsymbol{\theta})$ is the partition function and T is the state transition distribution. Fixing T, the optimal $\boldsymbol{\theta}^*$ maximizes the log-likelihood $L(\boldsymbol{\theta})$ of the demonstrations

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, L(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{\zeta} \log P(\zeta|\boldsymbol{\theta}, T). \quad (2)$$

The maximum is obtained using gradient ascent; the gradient of log-likelihood is the difference between observed and expected feature counts, which can be expressed in terms of state visitation frequencies D_s. The frequency of visiting a state given a policy can be computed with an iterative algorithm

$$\nabla L(\boldsymbol{\theta}) = \tilde{\boldsymbol{f}} - \sum_{\zeta} P(\zeta|\boldsymbol{\theta}, T) \boldsymbol{f}_{\zeta} = \tilde{\boldsymbol{f}} - \sum_{s_i} D_{s_i} \boldsymbol{f}_{s_i}. \quad (3)$$

Domain Knowledge with MMD

Directly using IRL algorithm to recover reward function from game dynamics models what situations professional players want to be in, that is, their internal reward function. But the MaxEnt approach fails to learn the importance of goals in a game, mainly because goals are such rare events in ice hockey. Previous RL methods define the reward function explicitly in terms of goals. The **rule reward function** assigns reward 1 for scoring a goal (i.e., getting the puck into the net) and 0 for other actions.

Motivated by knowledge transfer between reward functions (Mendez, Shivkumar, and Eaton 2018), we propose a new solution concept that allows us to combine IRL with domain knowledge during training. Here we introduce the *maximum mean discrepancy* (MMD) (Gretton et al. 2012) to transfer knowledge between these two reward functions and bridge their disparity.

Denote by X a random variable with distribution p. Denote by Y a random variable with distribution q. Formally, MMD defines the following difference measure

$$d_{\mathcal{H}}(X, Y) = \sup_{f \in \mathcal{H}} (\mathbb{E}_X[f(X)] - \mathbb{E}_Y[f(Y)]), \quad (4)$$

where $\mathcal{H}$ is a class of functions, known as a reproducing kernel Hilbert space (RKHS). An unbiased estimation of squared MMD is given by (Long et al. 2017):

$$\begin{aligned} d^2_{\mathcal{H}_k}(X, Y) = &\frac{1}{n_x^2} \sum_{i=1}^{n_x} \sum_{j=1}^{n_x} k(x_i, x_j) + \frac{1}{n_y^2} \sum_{i=1}^{n_y} \sum_{j=1}^{n_y} k(y_i, y_j) \\ &- \frac{2}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} k(x_i, y_j). \end{aligned} \quad (5)$$

Combining MaxEnt IRL with MMD, the learning process is expressed as follows. At each training step, our model aims to maximize the log-likelihood of demonstrations as well as to minimize the MMD between two reward functions. The optimal $\boldsymbol{\theta}^*$ is then derived by

$$\begin{aligned} \boldsymbol{\theta}^* &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, L(\boldsymbol{\theta}) - \lambda d^2_{\mathcal{H}_k}(R, \hat{R}) \\ &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, L(\boldsymbol{\theta}) + 2\lambda k(r, \hat{r}), \end{aligned} \quad (6)$$

where r is the reward inferred from demonstrations, $\hat{r}$ is the reward from our domain knowledge, and λ is a trade-off parameter. The kernel function k is a Gaussian kernel $k(x_i, x_j) = e^{-||x_i - x_j||^2/2\sigma^2}$ as in most knowledge transfer frameworks (Long et al. 2017). Following (Wulfmeier, Rao, and Posner 2016), we pretrain a $\hat{\boldsymbol{\theta}}$ to match our domain knowledge $\hat{r}$ and initialize $\boldsymbol{\theta}$ with this pretrained parameter.

Policy Learning Performance

To evaluate how well the reward function recovered by our model approximates players' behavior, we compare the demonstrated trajectories with the probabilistic distribution over trajectories generated by our algorithm using two common metrics: negative log-likelihood (NLL) and modified Hausdorff Distance (MHD) (Kitani et al. 2012).

$$NLL(\zeta) = \mathbb{E}_{\pi(a|s)}[-\log \prod_t P(s_{t+1}|s_t, a_t)] \quad (7)$$

$$\begin{aligned} MHD(\{\zeta_d\}, \{\zeta_g\}) &= \max(h(\{\zeta_d\}, \{\zeta_g\}), h(\{\zeta_g\}, \{\zeta_d\})) \\ h(\{\zeta\}, \{\hat{\zeta}\}) &= \frac{1}{|\{\zeta\}|} \sum_{\zeta_i \in \{\zeta\}} \min_{\hat{\zeta}_j \in \{\hat{\zeta}\}} ||\zeta_i - \hat{\zeta}_j|| \end{aligned} \quad (8)$$

NLL calculates how likely the demonstrations are under policy π, and MHD is a spatial measure of the distance between demonstrated and generated trajectories. The (optimal) policy π is first got via solving MDP given reward. Table 1 shows the results. Using rule reward only cannot predict any demonstrated trajectories because too many states are assigned visitation probability 0. The reward recovered by IRL with domain knowledge outperforms its counterparts in both comparisons, where lower numbers represent models approximating expert behaviour with higher precision.

Methods	NLL	HMD
Rule reward function	-	13.37
IRL recovered reward function	57.3	9.71
IRL+Rule recovered reward function	**52.7**	**7.77**

Table 1: Evaluation of trajectory likelihoods under optimal policies derived from different reward functions. Likelihood metrics used are NLL and HMD.

Player Evaluation

We first define the action impact values and then give examples of player ranking.

Action Impact Values

Action impact, which quantifies the difference made by an action, has been used for player evaluation (Routley and Schulte 2015; Schulte et al. 2017; Liu and Schulte 2018). We adopt action impact values as a function of game context (Markov state) defined by (Routley and Schulte 2015)

$$impact(s, a) \equiv Q_T(s, a) - V_T(s), \tag{9}$$

where T is the team executing the action a, $Q(\cdot)$ is the Q-function, and $V(\cdot)$ is the value function. This action impact function measures how much an action improves over the average action. The value of a state is defined as the expected total reward given a policy, and the Q-function and value function can be calculated using the expected Bellman equation (Sutton and Barto 1998).

Player Rankings

Following (Liu and Schulte 2018), the ranking score for a player is the sum of this player's total action impact values, which is expressed as

$$Score_i = \sum_{s,a} n^i_{\mathcal{D}}(s, a) \times impact(s, a), \tag{10}$$

where $\mathcal{D}$ denotes the dataset we use, i is the playerId, and $n^i_{\mathcal{D}}(s, a)$ is the occurrence number that player i performed action a at state s observed from $\mathcal{D}$. The total impact is not normalized for time-on-ice (TOI), because TOI correlates with player strength. Dividing the ranking score by TOI therefore reduces the score differences among players. Note that impact values can be both positive and negative, so a high total impact reflects the net value of a player's actions, rather than the total number of the actions.

Different from (Routley and Schulte 2015; Liu and Schulte 2018) where all the players are evaluated together, we evaluate offensive players (Center, Left Wing, Right Wing) and defensive players (Defenceman, Goalie) separately with the following considerations. First, previous RL methods with sparse reward rank offensive players higher than defensive players in most cases. Second, these two types of players play different roles in a team under diverse strategies leading to distinct behavior.

Name	Assists	Goals	Points	Team	Salary
Anze Kopitar	38	22	60	LA	11,000,000
Aleksander Barkov	61	35	96	FLA	6,900,000
Dylan Larkin	41	32	73	DET	7,000,000
Mark Scheifele	46	38	84	WPG	6,750,000
Jack Eichel	54	28	82	BUF	10,000,000
Jonthan Toews	46	35	81	CHI	9,800,000
Leon Draisaitl	55	50	105	EDM	9,000,000
Nathan Mackinnon	58	41	99	COL	6,750,000
Mika Zibanejad	44	30	74	NYR	5,350,000
Sebastian Aho	53	30	83	CAR	12,000,000

Table 2: 2018-19 Top-10 offensive players

Name	Assists	Goals	Points	Team	Salary
Drew Doughty	37	8	45	LA	12,000,000
Miro Heiskane	21	12	33	DAL	925,000
Duncan Keith	34	6	40	CHI	3,500,000
Brent Burns	67	16	83	SJ	10,000,000
Roman Josi	41	15	56	NSH	4,000,000
Mattias Ekholm	36	8	44	NSH	4,000,000
Morgan Rielly	52	20	72	TOR	5,000,000
Ryan Suter	40	7	47	MIN	9,000,000
Ivan Provorov	19	7	26	PHI	6,750,000
Esa Lindell	21	11	32	DAL	7,000,000

Table 3: 2018-19 Top-10 defensive players

Tables 2 and 3 list the top-10 highest impacts offensive and defensive players by our algorithm. All these players are fantasy NHL stars according to recent NHL news (Jensen 2019; Reese 2019). Our ranking can be used to identify promising players. For instance, Miro Heiskane just began his career in 2017 and drew salaries below other top ranking players but is nominated as top-50 Defenseman by NHL (Reese 2019). Our ranking does not have apparent bias to player position compared with two recent RL methods, Score Impact (SI) (Routley and Schulte 2015) and Goal Imapct Metric (GIM) (Liu and Schulte 2018). For instance, the top-50 players given by SI are all offensive players, and top-50 by GIM only contains one Defenceman, while ours contains 34 Defencemen.

Empirical Evaluation

To access player evaluation metrics, we follow previous work (Routley and Schulte 2015; Schulte et al. 2017; Liu and Schulte 2018) to compute their correlation with commonly used statistic measurements like Assists, Goals, Points, as these statistics are generally regarded as important measures of player strength.

We compare our method with the following player evaluation metrics. Plus-minus (**+/-**) is a commonly used basic metric to measure the influence of player presence to the goals (Macdonald 2011). Valuing Actions by Estimating Probabilities (**VAEP**) defines the impact of an action as its

Methods	Assists	GP	Goals	GWG	SHG	PPG	S
+/-	0.269	0.086	0.282	0.278	0.118	0.124	0.156
VAEP	0.215	0.185	0.215	0.089	-0.074	0.160	0.239
WAR	0.591	0.322	0.742	0.571	0.179	0.610	0.576
EG	0.656	0.629	0.633	0.489	0.099	0.391	0.737
SI	0.717	0.633	**0.975**	**0.665**	**0.249**	**0.770**	0.860
GIM	0.757	0.772	0.781	0.518	0.147	0.477	0.795
IRL	0.855	0.881	0.810	0.587	0.123	0.511	0.901
IRL-DK	**0.874**	**0.890**	0.820	0.601	0.125	0.528	**0.907**
Methods	Points	SHP	PPP	FOW	P/GP	SFT/GP	PIM
+/-	0.285	0.179	0.157	0.012	0.306	0.109	0.100
VAEP	0.235	-0.076	0.185	0.021	0.204	0.129	0.172
WAR	0.692	0.147	0.605	0.040	0.699	0.396	0.145
EG	0.694	0.183	0.508	0.254	0.644	0.713	0.355
SI	0.869	0.204	0.708	0.135	0.728	0.639	0.361
GIM	0.818	0.151	0.561	0.289	0.705	0.751	0.372
IRL	0.887	0.207	0.696	0.295	0.741	0.818	0.439
IRL-DK	**0.902**	**0.210**	**0.723**	**0.298**	**0.760**	**0.820**	**0.445**

Table 4: Correlation with success measures (offensive)

Methods	Assists	GP	Goals	GWG	SHG	PPG	S
+/-	0.173	0.132	0.144	0.177	0.235	-0.116	0.113
VAEP	0.054	-0.045	0.005	0.010	0.384	0.071	-0.016
WAR	0.204	0.028	0.365	0.275	0.097	0.246	0.186
EG	0.589	0.688	0.507	0.321	0.327	0.306	0.679
SI	0.607	0.488	**0.934**	**0.449**	**0.491**	**0.457**	0.709
GIM	0.702	0.862	0.596	0.263	0.130	0.170	0.764
IRL	0.809	0.943	0.656	0.410	0.267	0.326	0.897
IRL-DK	**0.839**	**0.950**	0.685	0.429	0.281	0.346	**0.913**
Methods	Points	SHP	PPP	FOW	P/GP	SFT/GP	PIM
+/-	0.175	0.107	-0.05	0.095	0.169	0.067	0.072
VAEP	0.042	0.065	-0.003	0.101	0.064	-0.036	-0.031
WAR	0.252	0.128	0.266	0.174	0.279	0.006	-0.089
EG	0.611	0.278	0.399	0.118	0.503	0.694	0.360
SI	0.720	0.174	0.488	0.103	0.521	0.499	0.272
GIM	0.730	0.085	0.358	0.140	0.471	0.706	0.438
IRL	0.841	0.281	0.549	0.184	0.557	0.776	0.559
IRL-DK	**0.853**	**0.289**	**0.553**	**0.185**	**0.643**	**0.778**	**0.570**

Table 5: Correlation with success measures (defensive)

offensive score plus defensive score (Decroos et al. 2019). Because our dataset was too large to be processed by the VAEP authors' code, we replaced the gradient-boosted tree of the original implementation by a neural network classifier. Win-Above-Replacement (**WAR**) estimates the difference of team's wining chance if a target player is replaced by an average player (Gerstenberg et al. 2014). Expected Goal (**EG**) weights each shot by its chance of leading to a goal. Scoring Impact (**SI**) is most related to our method, but uses a sparse reward (Routley and Schulte 2015; Schulte et al. 2017). Goal Impact Metric (**GIM**) uses deep Q-network with sparse reward to predict Q values and defines the difference between two consecutive Q values as action impact (Liu and Schulte 2018). We also adopt the IRL method without domain knowledge as a baseline.

Season Totals: Correlations with Standard Success Measures

The following experiment computes the correlations with success measures over the entire season. The NHL official website provides 14 standard success measures (www.nhl.com/stats/player), including Assists, Goals, Points, Game Play (GP), Game Wining Goal (GWG), Short-handed Goal (SHG), Power-play Goal (PPG), Shots (S), Short-handed Point (SHP), Power-play Point (PPP), Face-off Win Percentage (FOW), Points per game (P/GP), Shifts per game (SFT/GP), and Penalty Minute (PIM). The results for offensive and defensive players are shown in Tables 4 and 5. Our method achieves the highest correlation in 10 out of 14 success measures except for goal and three goal related items (GWG, SHG, and PPG). For GWG, our results are comparable to the highest for both offensive and defensive player measures. For SHG and PPG, it achieves the second best results or comparable to the second best.

Round-by-Round Correlations: Predicting Future Performance from Past Performance

A sport season normally consists of several rounds. A team or player will finish n competitions at the end of round n. We compute the correlation between player values at the end of round n and three main success measures, including Assists,

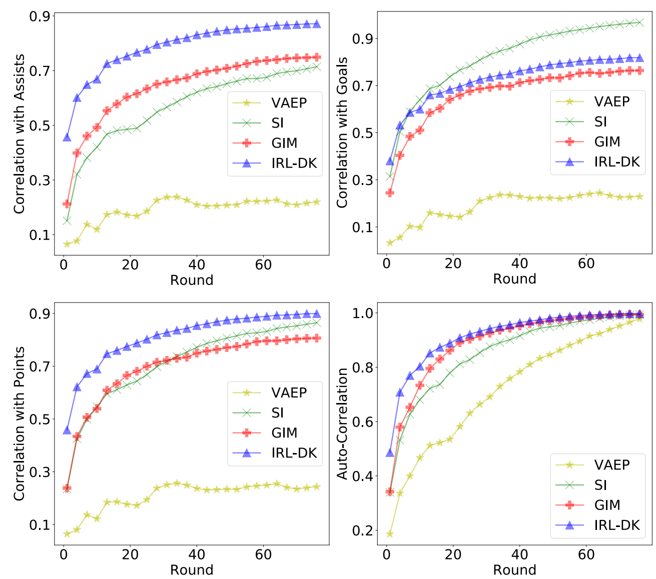

Figure 1: Correlations between round-by-round metrics and season totals for offensive players

Goals, and Points, over the whole sport season. This experiment assesses the learning ability of different metric, so that the future performance of players can be inferred from the past performance. We also compute the *auto-correlation* for different metrics between players' round values and final season values. Auto-correlation evaluates the temporal consistency of a metric (Pettigrew 2015). Since most players' strength is stable throughout a season, a good player metric should show temporal consistency (Franks et al. 2016).

We focus on four machine learning based methods VAEP, SI, GIM, and IRL-DK. Figure 1 shows round-by-round correlation with Assists, Goals, Points, and the auto-correlation between round values and season total for offensive players. IRL-DK is the most stable model measured by auto-correlation. We also find IRL-DK is able to learn knowledge faster from data, as its performance is better than others even at the very beginning of the season.

Conclusion

We investigate inverse reinforcement learning for professional ice hockey game analytics. We apply IRL with domain knowledge to recover reward for complex game dynamics, which addresses the sparse reward issue for RL models. Based on the recovered reward function and calculated Q-values, we build a context-aware player performance metric that provides a comprehensive evaluation for both offensive and defensive players in NHL by taking all their actions into account. In experiments our method shows no obvious bias for any player position, achieves highest correlation with most standard success measures, and is most temporally consistent. While we have focused on ice hockey for concreteness, the inverse RL method can be applied to a Markov model for any sport.

References

Decroos, T.; Bransen, L.; Haaren, J. V.; and Davis, J. 2019. Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th International Conference on Knowledge Discovery & Data Mining (KDD-19)*, 1851–1861.

Franks, A. M.; D'Amour, A.; Cervone, D.; and Bornn, L. 2016. Meta-analytics: tools for understanding the statistical properties of sports metrics. *Journal of Quantitative Analysis in Sports* 12(4):151–165.

Gerstenberg, T.; Ullman, T.; Kleiman-Weiner, M.; Lagnado, D.; and Tenenbaum, J. 2014. Wins above replacement: Responsibility attributions as counterfactual replacements. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 36.

Gretton, A.; Borgwardt, K. M.; Rasch, M. J.; Schölkopf, B.; and Smola, A. 2012. A kernel two-sample test. *Journal of Machine Learning Research* 13(Mar):723–773.

Jaynes, E. T. 1957. Information theory and statistical mechanics. *Physical review* 106(4):620.

Jensen, P. 2019. Fantasy center top 50 rankings for 2019-20. https://www.nhl.com/news/2019-20-nhl-fantasy-hockey-center-top-50-rankings/c-299921028. [Online; accessed 15-October-2019].

Kitani, K. M.; Ziebart, B. D.; Bagnell, J. A.; and Hebert, M. 2012. Activity forecasting. In *Proceedings of the 12nd European Conference on Computer Vision (ECCV-12)*, 201–214. Springer.

Littman, M. L. 1994. Markov games as a framework for multi-agent reinforcement learning. In *Proceedings of the 11st International Conference on Machine learning (ICML-94)*, volume 157. 157–163.

Liu, G., and Schulte, O. 2018. Deep reinforcement learning in ice hockey for context-aware player evaluation. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI-18)*, 3442–3448.

Long, M.; Zhu, H.; Wang, J.; and Jordan, M. I. 2017. Deep transfer learning with joint adaptation networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML-17)*, volume 70, 2208–2217. JMLR. org.

Macdonald, B. 2011. An improved adjusted plus-minus statistic for nhl players. In *Proceedings of the 5th annual MIT Sloan Sports Analytics Conference*, volume 3, 1–8.

Mendez, J. A.; Shivkumar, S.; and Eaton, E. 2018. Lifelong inverse reinforcement learning. In *Proceedings of the 32nd Conference on Neural Information Processing Systems (NIPS-18)*, 4502–4513.

Ng, A. Y.; Russell, S. J.; et al. 2000. Algorithms for inverse reinforcement learning. In *Proceedings of the 17th International Conference on Machine Learning (ICML-00)*, volume 1, 2.

Pettigrew, S. 2015. Assessing the offensive productivity of nhl players using in-game win probabilities. In *Proceedings of the 9th annual MIT sloan sports analytics conference*, volume 2, 8.

Reese, R. 2019. Fantasy defenseman top 50 rankings for 2019-20. https://www.nhl.com/news/nhl-fantasy-hockey-top-50-defenseman-rankings-2019-20/c-282830728. [Online; accessed 15-October-2019].

Routley, K., and Schulte, O. 2015. A markov game model for valuing player actions in ice hockey. In *Proceedings of the 31st Uncertainty in Artificial Intelligence (UAI-15)*, 782–791.

Schuckers, M., and 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 *Proceedings of the 7th annual MIT sloan sports analytics conference*.

Schulte, O.; Khademi, M.; Gholami, S.; Zhao, Z.; Javan, M.; and Desaulniers, P. 2017. A markov game model for valuing actions, locations, and team performance in ice hockey. *Data Mining and Knowledge Discovery* 31(6):1735–1757.

Sutton, R. S., and Barto, A. G. 1998. *Reinforcement learning: An introduction*. MIT press.

Thomas, A.; Ventura, S. L.; Jensen, S. T.; and Ma, S. 2013. Competing process hazard function models for player ratings in ice hockey. *The Annals of Applied Statistics* 1497–1524.

Von Neumann, J., and Morgenstern, O. 1947. Theory of games and economic behavior, 2nd rev.

Wulfmeier, M.; Rao, D.; and Posner, I. 2016. Incorporating human domain knowledge into large scale cost function learning. *arXiv preprint arXiv:1612.04318*.

Ziebart, B. D.; Maas, A.; Bagnell, J. A.; and Dey, A. K. 2008. Maximum entropy inverse reinforcement learning. In *Proceedings of the 23rd Conference on Artificial Intelligence (AAAI-08)*.

Top players given by other metrics

The top-50 players given by SI (Routley and Schulte 2015) are all offensive players. Tables 6 and 7 list the top-10 highest impacts offensive and defensive players by SI. The ranking is based on the 2018-19 season.

Name	Assists	Goals	Points	Team	Salary
Alex Ovechkin	38	51	89	WSH	10,000,000
John Tavares	41	47	88	TOR	15,900,000
Leon Draisaitl	55	50	105	EDM	9,000,000
Cam Atkinson	28	41	69	CBJ	7,375,000
Alex Debrincat	35	41	76	CHI	800,000
Steven Stamkos	53	45	98	TBL	9,500,000
Jake Guentzel	36	40	76	PIT	7,000,000
Brayden Point	51	41	92	TBL	5,250,000
Patrick Kane	66	44	110	CHI	9,800,000
David Pastrnak	43	38	81	BOS	6,800,000

Table 6: SI Top-10 offensive players

Name	Assists	Goals	Points	Team	Salary
Morgan Rielly	52	20	72	TOR	5,000,000
Dougie Hamilton	21	18	39	CAR	6,000,000
Kris Letang	40	16	56	PIT	7,250,000
Mark Giordano	57	17	74	CGY	6,750,000
Jared Spurgeon	29	14	43	MIN	5,500,000
Matt Dumba	10	12	22	MIN	7,400,000
Shea Weber	19	14	33	MTL	6,000,000
Erik Gustafsson	43	17	60	CHI	1,800,000
Alex Pietrangelo	28	13	41	STL	7,500,000
Roman Josi	41	15	56	NSH	4,000,000

Table 7: SI Top-10 defensive players

The top-50 players given by GIM (Liu and Schulte 2018) contains 49 offensive players and one defensive player. Tables 8 and 9 list the top-10 highest impacts offensive and defensive players by GIM.

Name	Assists	Goals	Points	Team	Salary
Sidney Crosby	65	35	100	PIT	9,000,000
Mark Scheifele	46	38	84	WPG	6,750,000
Leon Draisaitl	55	50	105	EDM	9,000,000
Jonathan Toews	46	35	81	CHI	9,800,000
Anze Kopitar	38	22	60	LA	11,000,000
Aleksander Barkov	61	35	96	FLA	6,900,000
John Tavares	41	47	88	TOR	15,900,000
Sean Couturier	43	33	76	PHI	4,500,000
Nicklas Backstrom	52	22	74	WSH	8,000,000
Connor McDavid	75	41	116	EDM	14,000,000

Table 8: GIM Top-10 offensive players

Name	Assists	Goals	Points	Team	Salary
Drew Doughty	37	8	45	LA	12,000,000
Jaccob Slavin	23	8	31	CAR	5,500,000
Samuel Girard	23	4	27	COL	700,000
T.J. Brodie	25	9	34	CGY	4,837,500
Michael Matheson	19	8	27	FLA	3,500,000
Thomas Chabot	41	14	55	OTT	832,500
Shea Theodore	25	12	37	VGK	5,200,000
Dmitry Orlov	26	3	29	WSH	6,500,000
Ivan Provorov	19	7	26	PHI	6,750,000
Morgan Rielly	52	20	72	TOR	5,000,000

Table 9: GIM Top-10 defensive players

The top-50 players given by VAEP (Decroos et al. 2019) contains 38 offensive players and 12 defensive players. Tables 10 and 11 list the top-10 highest impacts offensive and defensive players by VAEP.

Name	Assists	Goals	Points	Team	Salary
Jack Eichel	54	28	82	BUF	10,000,000
Ryan Getzlaf	34	14	48	ANA	8,275,000
Mika Zibanejad	44	30	74	NYR	5,350,000
Sidney Crosby	65	35	100	PIT	9,000,000
Brock Nelson	28	25	53	NYI	8,000,000
Lars Eller	23	13	36	WSH	4,000,000
Zach Aston-Reese	9	8	17	PIT	1,000,000
Chris Kreider	24	28	52	NYR	4,000,000
Nikita Kucherov	87	41	128	TBL	12,000,000
Leon Draisaitl	55	50	105	EDM	9,000,000

Table 10: VAEP Top-10 offensive players

Name	Assists	Goals	Points	Team	Salary
Jonas Brodin	14	4	18	MIN	5,750,000
Jaccob Slavin	23	8	31	CAR	5,500,000
Mark Giordano	57	17	74	CGY	6,750,000
Jake Gardiner	27	3	30	TOR	3,650,000
Jordie Benn	17	5	22	NYR	2,400,000
Anton Stralman	15	2	17	TBL	5,500,000
Ryan Suter	40	7	47	MIN	9,000,000
Trevor Van Riemsdyk	11	3	14	CAR	2,500,000
Esa Lindell	21	11	32	PHI	7,000,000
Duncan Keith	34	6	40	ARI	3,500,000

Table 11: VAEP Top-10 defensive players