Uncertainty-Aware Reinforcement Learning for Risk-Sensitive Player Evaluation in Sports Game

Guiliang Liu, Yudong Luo, Oliver Schulte, Pascal Poupart

University of Waterloo, The Chinese University of Hong Kong, Shenzhen, Vector Institute, and Simon Fraser University

Neurips 2022 Presentation

Problem Definition

Player Evaluation:

- **Definition:** Evaluate the contribution of players by quantifying their action impacts.

- **Challenges:**

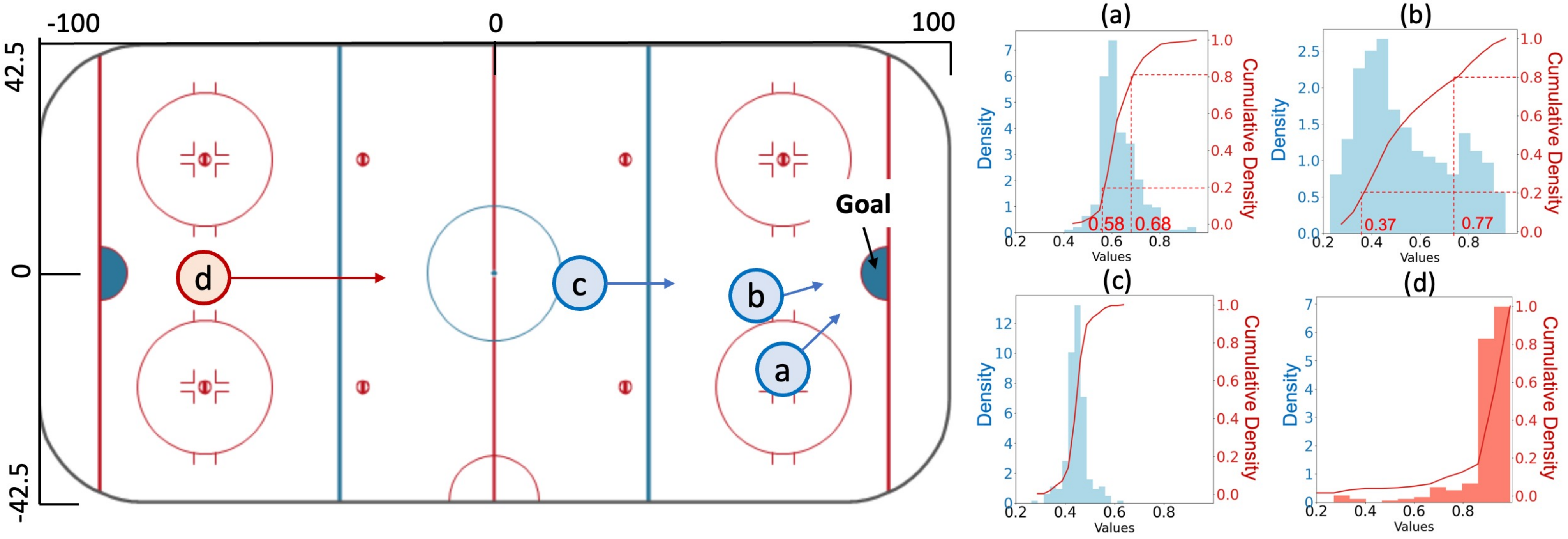
1) Previous methods are expectation-based, which cannot differentiate the risk-seeking actions from the risk-averse ones.
2) How to distinguish these actions and assign proper credits to the players remains a fundamental challenge in sports analytics.

- **Our solution:** Risk-Sensitive Player Evaluation with Post-hoc Calibration

Motivation

Example: The predicted distribution of future goals for the shots made at positions (a to d).

- **Risk-Sensitive Evaluation**: Distributions (a) and (b) have the same expectation, but the first shot has a larger risk-averse estimate and a smaller risk-seeking estimate.
- **Post-hoc Calibration**: shot made at the position (d) is rare in an ice hockey game, and thus this event is likely to be OoD, leading to a biased prediction.



Uncertainty-Aware Reinforcement Learning

Estimate the aleatoric and epistemic uncertainty for a risk-sensitive player evaluation.

- Distributional RL for **Aleatoric Uncertainty**

1) Learn the distribution of $Z_k(s_t, a_t)$, i.e., number of goals when a player performs action a_t in state s_t.

2) Represent $Z_k(s_t, a_t)$ by a uniform mixture of N supporting quantiles.

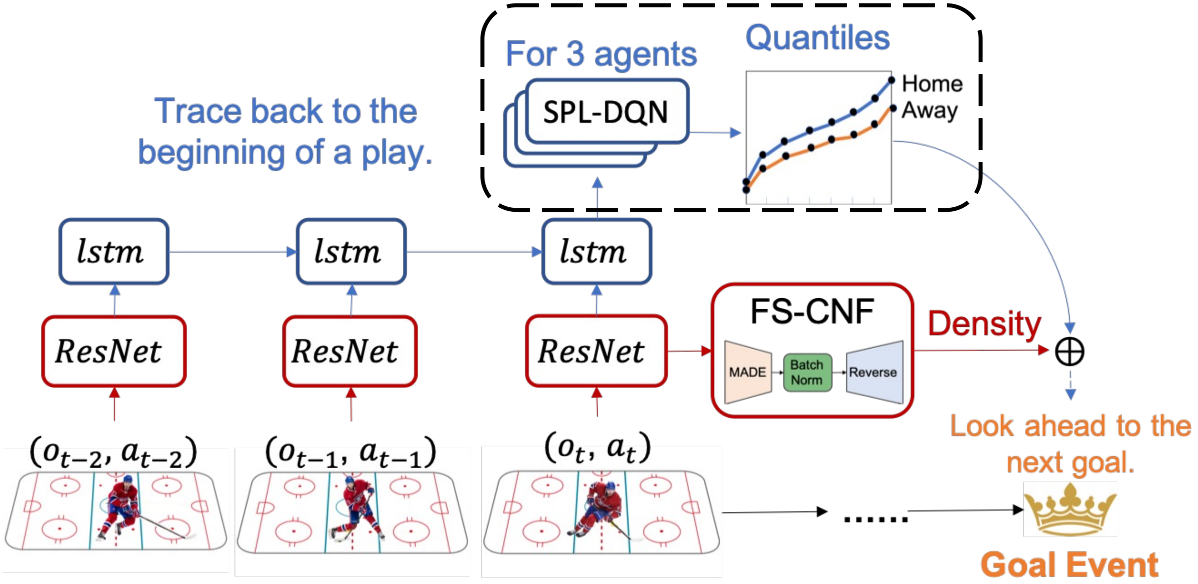
$$\hat{Z}_k(s_t, a_t) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\theta_{k,i}(s_t,a_t)}$$

($\theta_{k,i}$ estimates the ith quantile)

3) Distributional Bellman Operator

$$\mathcal{T}^{\pi} Z_k(s_t, a_t) \triangleq R_k(s_t, a_t) + \gamma Z_k(S_{t+1}, A_{t+1})$$

Where $s_{t+1} \sim P_{\mathcal{T}}(S_{t+1} | s_t, a_t)$ and $a_{t+1} \sim \pi(A_{t+1} | S_{t+1})$



Uncertainty-Aware Reinforcement Learning

Estimate the aleatoric and epistemic uncertainty for a risk-sensitive player evaluation.

- Density Estimator for **Epistemic Uncertainty**

Feature Space Conditional Normalizing Flow (FS-CNF)

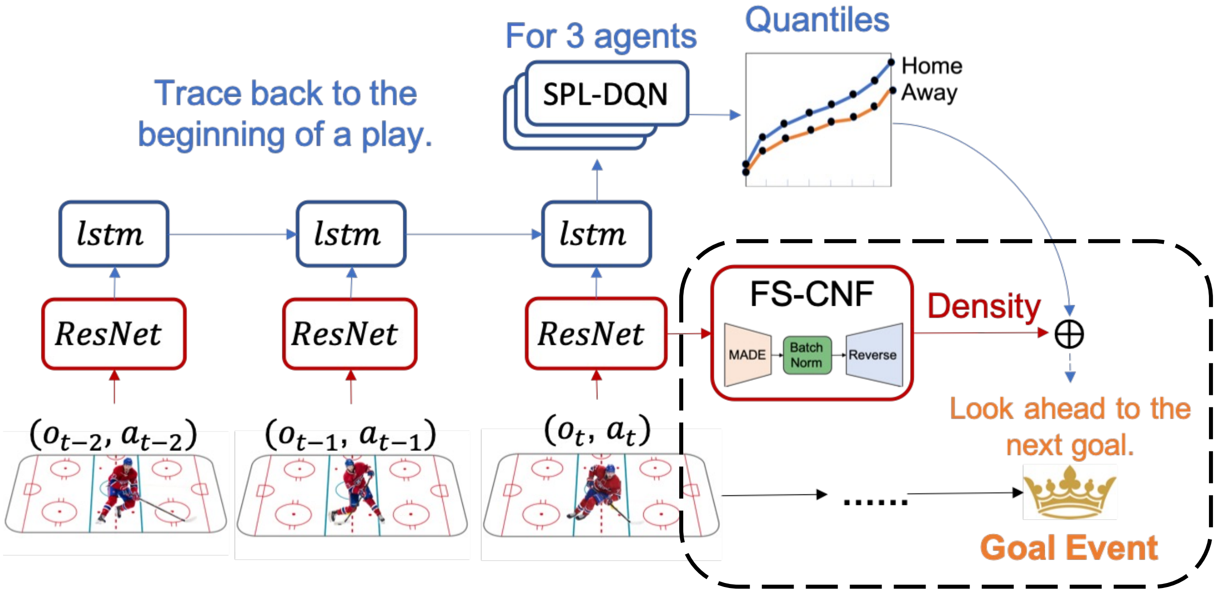
1) Feature Extractor.

To prevent feature collapse, the extractor is subjected to a bi-Lipschitz constraint:

$$\boxed{\beta_1 \|x_1 - x_2\|_I \geq} \|f_\theta(x_1) - f_\theta(x_2)\|_F \boxed{\geq \beta_2 \|x_1 - x_2\|_I}$$

Upper bound ensures smoothness

Lower bound ensures sensitivity to distance



2) Density Estimator.

Based on the extracted features, FS-CNF utilizes the Masked Auto-regressive Flow (MAF).

Player Evaluation

- **Risk-sensitive Impact Metric**

To understand how players respond to risk, we propose a Risk-sensitive Game Impact Metric (RiGIM)

$$RiGIM_l(c) = \sum_{(s,a)\in\mathcal{D}'} n(s,a,l) \times \phi_k(s,a,c) \quad \text{where} \quad \phi_k(s_{t+1}, a_{t+1}, c) = \left[\hat{Z}_k^c(s_{t+1}, a_{t+1}) - \hat{Z}_k^c(s_t, a_t)\right]\mathbb{I}_{p(\cdot|\boldsymbol{z}_E)\geq\epsilon}$$

- **Case Study: Player Ranking in Testing Games**

We rank players according to their RiGIM scores in the NHL testing games.

Table 1: Top 10 players with confidence 0.2.

Player Name	Position	Team	P	A	G	RiGIM
Jonathan Toews	C	CHI	10	5	5	14.72
Anze Kopitar	C	LAK	12	9	3	14.55
Vincent Trocheck	C	FLA	8	5	3	14.02
Tomas Hertl	C	SJS	12	8	4	13.97
John Tavares	C	TOR	12	3	9	13.92
Tyler Seguin	C	DAL	18	12	6	13.71
Leon Draisaitl	C	EDM	16	8	8	13.16
Aleksander Barkov	C	FLA	19	14	5	12.63
Sean Couturier	C	PHI	11	6	5	12.62
Nathan MacKinnon	C	COL	12	6	6	12.48

Table 2: Top 10 players with confidence 0.8.

Player Name	Position	Team	P	A	G	RiGIM
Radek Faksa	C	DAL	6	3	3	2.74
Leon Draisaitl	C	EDM	16	8	8	2.51
John Klingberg	D	DAL	10	9	1	2.46
Esa Lindell	D	DAL	3	1	2	2.29
Connor McDavid	C	EDM	18	11	7	2.23
Tomas Hertl	C	SJS	12	8	4	1.93
Miro Heiskanen	D	DAL	5	3	2	1.86
Elias Pettersson	C	VAN	8	6	2	1.79
Tyler Seguin	C	DAL	18	12	6	1.78
Roope Hintz	LW	DAL	11	7	4	1.77

Question and Answering (Q&A)