Inverse Reinforcement Learning for Team Sports: Valuing Actions and Players

Yudong Luo[1,2], Oliver Schulte[2,3], Pascal Poupart[2]
[1]Simon Fraser University, [2]University of Waterloo, [3]SportLogiq

Valuing Actions and Players

- Sports analytics provides professional method for analyzing sports data to facilitate decision making before and during sports events.
- A major task of sports statistics is player evaluation, which supports drafting, coaching, and trading decisions.
- Most approaches use the total value of player's actions to rank players, which reduces players evaluation to actions evaluation.
- We propose a IRL with domain knowledge method to learn reward function for game states and recover action values for NHL ice hockey players.

Reward Sparsity Problem

Previous state-of-the-art methods use RL to learn action value Q function, but use sparse reward signals. 1 for goal and 0 for other actions.

- Scoring Impact (SI): uses Markov model to model game dynamics, dynamic program to compute value function, defines advantage value as action impact:
  $\mathrm{impact}(s,a) = Q_{H/A}(s,a) - V_{H/A}(s)$
- Goal Impact Metric (GIM): uses deep recurrent Q-network, Q represents the probability of scoring the next goal, defines the differences between two consecutive Q-values as action impact:
  $\mathrm{impact}(s,a) = Q_{H/A}(s_t,a_t) - Q_{H/A}(s_{t-1},a_{t-1})$, where H/A represents home or away team.

In low scoring games, like ice hockey and soccer, RL with sparse reward leads to the fact that explicit values are only attached to rare goal events.

- Action values emphasis on goals and related actions (shots, assists).
- Player rankings bias towards offensive players (score more goals than defensive players).

We use IRL to recover a dense reward function.

Markov Game Model for Ice Hockey

We use a play-by-play dataset provided by SportLogiq and our Markov Game for Ice Hockey follows SI.
The Markov Game has two agents Home team and Away team. The state space includes the following features:

- ManPower: Even Strength, Shorthanded, PowerPlay
- Goal Difference: difference between home and away goals
- Period: ranges from 1 to 3 (not consider overtime play)
- Team identity: Home or Away
- Location: cluster into 6 regions

The dataset records 27 different action types, and Home and Away teams share the same action space. Transition function is calculated by observed frequency

$$\mathrm{T}(s,a,s') = P(s'|s,a) = O(s,a,s')/O(s,a), \quad (1)$$

where $O(\cdot)$ counts the occurrence number in our dataset.

Alternating Learning for Multi-agent IRL

We assume the polices of two professional teams in Markov Game satisfy Nash Equilibrium, as each team optimizes against the observed polices of another team

- In sports, teams have direct access only to the observed behaviour of the other team.
- When an opponent's observed behaviour falls shorts of their optimal strategy, successful teams take advantage of it.

Transform multi-agent to single-agent problem.

- Treat B as A's environment, and learn reward for A using single-agent IRL.
- Repeat the procedure with the role of teams A and B reversed.

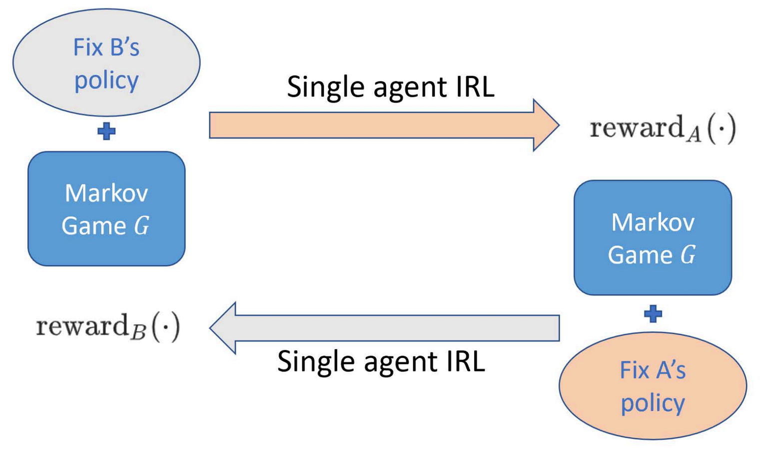

Fig. 1: Alternating IRL for two agents Markov Game.

MaxEnt IRL with Domain Knowledge

Maximum Entropy (MaxEnt) IRL

- Reward is a linear function of state features, with weights $\theta \in \mathbb{R}^d$.
- The reward for a trajectory ζ is the sum reward of visited states.
- The likelihood of a trajectory ζ is proportional to exponential reward $P(\zeta) \propto \exp(r_\theta(\zeta))$
- Maximize the likelihood of trajectories (data) given rewards (r_θ)

Goals are such rare events in low scoring sports, e.g. ice hockey. Directly applying single agent IRL fails to learn the importance of goals.

- Choose a kernel function k to measure the similarity between learned rewards and observed goals.
- Maximize the regularized likelihood function.

$$\operatorname{argmax} L(\{\zeta\}|r_\theta) + \lambda k(r_\theta, r_K), \quad (2)$$

where r_K is the rewards from domain knowledge, namely observed goals.

Player ranking system flow.

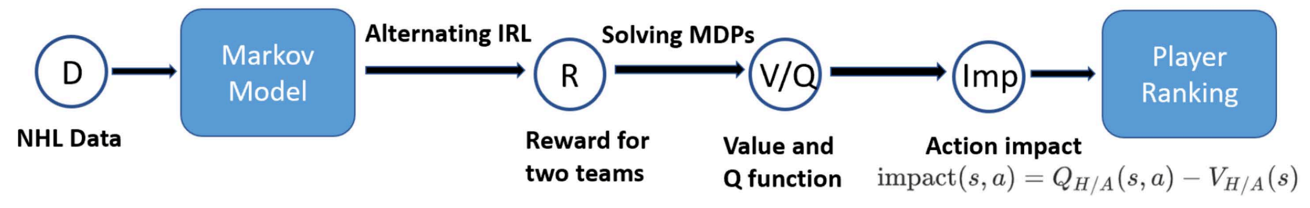

Fig. 2: System flow of player ranking

Player ranking Comparison

We compare our method with plus/minus (+/-), Valuing Actions by Estimating Probabilities (VAEP), Win-Above-Replacement (WAR), Scoring Impact (SI), Goal Impact Metric (GIM). We name our method as IRL-DK, and we also adopt IRL as a baseline.

On ground truth for play ranking. We calculate the correlation with successful measures (Assists, Game Play, Goals...), provided by NHL website, which is generally regarded as important measures of a player's ability to impact a game.

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	0.872	0.812	0.587	0.123	0.513	0.901
IRL-DK	**0.882**	**0.887**	0.824	0.607	0.125	0.537	**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	0.891	0.207	0.696	0.294	0.741	0.818	0.437
IRL-DK	**0.908**	**0.213**	**0.734**	**0.298**	**0.769**	**0.820**	**0.446**

Fig. 3: Correlation with success measures (offensive player)

Our ranking does not have obvious bias towards player positions. E.g. For the top-50 players, SI rankings are all offensive players and GIM rankings only contain one defence man. In our ranking, 32 defence men are ranked among the top 50.

Round by Round Correlation

Good player ranking metric should be temporal consistent.

- Player's performance is usually stable across the season.
- Predict player's future performance from the past.

Correlation between **first n round** value with Assists and Points.

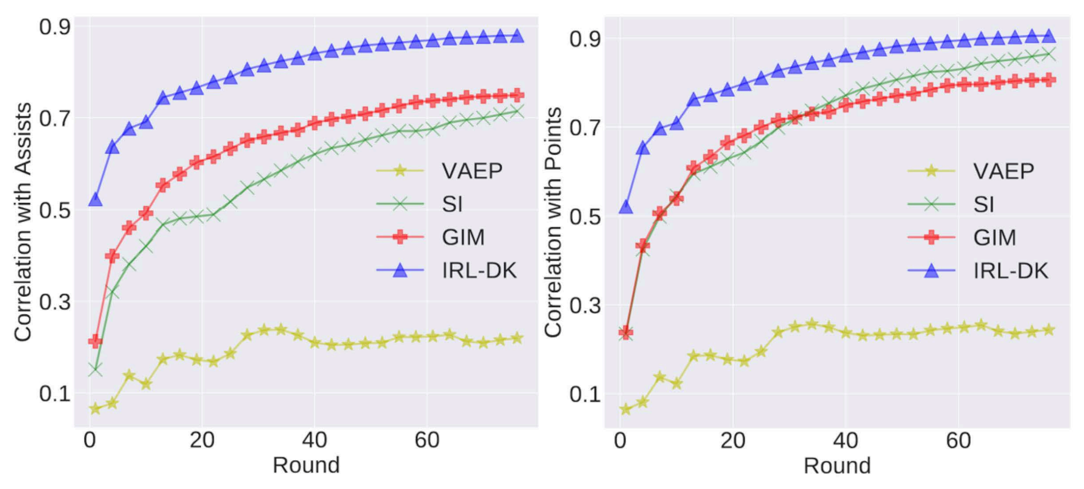

Fig. 4: Round by Round Correlation with Assists and Points (offensive player)