Valuing Actions and Players

- Sports analytics provides professional methodology 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 rate players, which reduces players evaluation to actions evaluation.
- We propose a RL with domain knowledge method to learn reward function for game state 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, define advantage value as action impact:
  \[ V(s, a) = Q(s, a) - Q(s, \pi(s)) \]
- 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:
  \[ V(s, a) = Q(s, a) - Q(s, \pi(s)) \]
where \( H \) 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 emphasize on-goals and related actions (shots, assists).
- Player rankings bias towards offensive players (score more goals than defensive players).

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 sports Home team and Away team. The state space includes the following features:
  - MaxPower: Even Strength, Shorthanded, PowerPlay
  - Goal Difference: difference between home and away goals
  - Period: range 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:
  \[ T(s', a) = P(s', a, s) = \frac{n(s', a, s)}{\sum_{a'} n(s', a', s)} \]
where \( n(s', a, s) \) counts the occurrence number in our dataset.

Alternating Learning for Multi-agent IRL

We alternate the policies of two professional teams in Markov Game satisfy Nash Equilibrium, as each team optimizes against the opponent (WARR), 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 player ranking. We calculate the correlation with successful measures (Asstais, Game Play, Goals..), provided by NHL website, which is generally regarded an important measure of a player's ability to impact a game.

MaxEnt IRL with Domain Knowledge

Maximum Entropy (MaxEnt) IRL
- Reward is a linear function of state features, with weights \( \theta \) in \( R^S \).
- The reward for a trajectory \( \gamma \) is the sum reward of visited states: \( R(\gamma) = \sum_{s \in \gamma} \theta' = \sum_{s \in \gamma} \theta(s) \)
- Maximize the likelihood of trajectories (data) given rewards \( \theta \)
Goals are such rare events in low scoring sports, e.g. ice hockey. Directly applying single agent IRL fail 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:
  \[
  \text{argmax}_{\theta} \sum_{\gamma} K(\gamma)\mathbb{E}[\theta(\gamma)] + \alpha + \frac{1}{2}\|\theta\|_2^2,
  \]
where \( \alpha \) is the regularization on reward knowledge, namely observed goals.

Round by Round Correlation

Good player ranking metric should be temporally consistent.
- Player's performance is usually stable across the season.
- Predict player's future performance from the past.
Correlation between first n round values with Assistais and Points.

Player ranking Comparison

We compare our method with plausibility (+) Valuing Actions by Estimating Probabilistic Wins-Above-Replacement (WARR), Scoring Impact (SI), Goal Impact Metric (GIM). We name our method as IRL-DK, and we also adopt IRL as a baseline.

Our ranking does not show 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 defense men are ranked among the top 50.