Rebuttal

We’d like to thank the reviewers for their time and comments. We especially appreciate the constructive suggestions. In our paper, a predictive Deep Reinforcement Learning(DRL) model is constructed to assess every players' actions under game context. Our work has extended DRL to the field of sport analysis. With specified reward function (e.g. goals, assist or penalty), fine-grained analysis can be performed to every players' action. As the beginning of this story, we evaluate players' performance and rating them by aggregating those values. In future work, more exciting system like game outcome prediction or even in-game gambling will be explored in successive work. As our foundation of DRL in Sport analytic, we do hope this paper can be accepted and we promise to open source code and let everybody play around the DRL in real professional sport.

Despite reviewers’ encouraging comments, several issues need to be clarified

1. Generalization ability of our method

Generalization is a key benefit of applying neural network approximation function to Temporal Difference RL. That's why we emphasize it in our paper. For example, (1) Spatial Projection in section Illustration of Network Predictions, where model generalize players' shot-value function over the entire ice hockey court. (2) Auto-Correlation in section Player Ranking also estimates player's future performance with round by round games data. (3) In our supplementary material, “Temporal auto-correlation of GIM” provides more detail how we applied our GIM to predict player's future performance with data in each rounds of game season. We promise the paper will be reorganized to state it more clearly.

2. Splitting training/testing data

Intuitively, you might suggest split games data to training, testing and validation set to build generalization learning environment. To predict game outcome, you need to generalize value function from training games to future unknown testing game. However, playing ranking is descriptive job. We choose to train our model to fit player's performance within current season well. Like the commonly applied statistic +/- metric (<https://en.wikipedia.org/wiki/Plus-minus>), which is also computed within current season, game or even episode.

3. Comparison with other methods or evaluation metrics

As stated in Lesion Study in section Empirical Evaluation, our comparison methods are designed following Lesion Study. Each time, we remove a part of our method and test its effectiveness. (First remove Dynamic Trace Length, then LSTM Layer, then Windows size and Temporal Difference Learning at last). Following this, we can further remove reinforcement learning as reviewer suggested, but result show that removing Temporal Difference has already hurted the performance significantly, removing RL generates even worse result (Weighted KLD over 10^-1). We can add this small experiment in the final version of our paper. Besides, actually we demonstrate the advantage of using RL on player ranking over other metric like (+/-), goals+assists, in section player ranking of our paper, from the point of identifying undervalued players (in Case Studies) and temporal consistency (in Auto-Correlation).

4. GIM and QAR

To measure players performance with Q values generated by our model, we propose Goal Impact Metric (GIM) to aggregate those values and define it in Player Ranking section. For Q-value above-replacement (QAR) metric, it is inspired by a common player assessment statistic called Value Over Replacement(VOR) (<https://en.wikipedia.org/wiki/Value_over_replacement_player>), which defines how much a player contributes to his team in comparison to a fictitious "replacement player," who is an average-skilled player at his position. By proving those two metrics are essentially computing the same thing w.r.t Q values, we build a theoretical foundation for our GIM metric form the definition of VOR.

5. The benefit of having small effect on the final outcome and numerically close performance metrics:

It's because we want our metric to be consistent over the entire season instead of being drastically influenced by winning or losing a single game. Part Auto-Correlation in Player Ranking section further demonstrate our metric is temporal consistency.

6. Definition of Episode

Episode is defined in section Play Dynamics in the NHL. Episode begins at the beginning of the game, or immediately after a goal, and terminates with a goal or the end of the game. But when we divide episode into plays, it is terminated by changing possession.

7. About Monte Carlo Method

Monte Carlo is exactly the supervised learning approach that reviewers mention. Monte Carlo Action Value Estimation (MCAVE) is comparison method and is briefly defined in Lesion Study. MCAVE applies Monte Carlo learning method to learn Q(a,s) for every action under certain game context.

8. About in-sample vs. out of sample problem

As explained in Point 2 above, playing ranking is descriptive job instead of a generalization learning. We want our model fit the player's performance in the entire season well, so we don't specify training and evaluation dataset.

9. Model Input and architecture

The input data is shown in Table 1. How to encode the input data is discussed in Section Play Dynamics in the NHL. But adding a figure is a nice idea.

10. Time cost with Different Algorithms

In the result of Empirical Evaluation, we report that our DP-LSTM model is trained to converge with 5 million minibatch gradient descent steps which costs 5 days. For comparison methods, to guarantee convergence, c1-MC, c1-Sarsa, c4-Sarsa, FT-LSTM use 12 million steps, 3 million steps, 3 million steps and 7 million steps receptively. We haven't record each running time as it depends on machine. It's a nice idea of adding those running steps.

We respect reviewers’ suggestion and will certain get a well-composed camera-ready version of paper. (I want them feel guilty to refuse us)