Review #12783

The paper applies deep reinforcement learning to the task of player valuation in the National Hockey League (NHL). Using data that provides information about game events and player actions, the paper seeks to assign values to the cumulative marginal contribution of each player. A LSTM is used to integrate over the sequence of game events/player actions that lead to a reinforcement signal. This allows for the network to incorporate game context into each of their prediction. Finally, a new metric - game Impact Metric (GIM) that integrates a player's evaluation is introduced and is demonstrated to correlate highly with other measures of success and player salary.

In general, the writing is concise and the use of descriptive figures (Figure 3), alongside the detailed explanation is helpful in communicating the core concepts. The motivation behind the work is very well established, and provides relevant information to attract the reader's attention. This paper applies deep reinforcement learning techniques to improve solutions to real world problems and is of great relevance to the conference.

The related work and backgorund touches on the major concepts that this paper builds upon. However, the description is too brief and the explanations too shallow. This leaves the reader with a lot of additional reading to properly interpret this work. I understand that the page limitations are very restrictive but a more in depth description would greatly improve the readability of this paper, and help it stand on its own.

It is not clear from the paper why a reinforcement learning approach is required, apart from the interpretability property of the Q-values. The entire dataset is available a priori, and is static such that different exploration strategies does not lead to the procurement of different datasets. Why not use a fully supervised technique to model the relationship between actions, observations and rewards? If the temporal credit assignment properties of Q-learning is an intended feature, use of possessive based LSTM should address the same issue?

Additionally, why is a deep network (Figure 3) required in this application? The dataset constructed already uses computer vision techniques and compiles the raw stream to features nicely structured and bucketed (Table 1). The features further derived from these seem to be manually done as well. If there is no more feature extraction to be done from raw data sources, what utility does a deep network provide that a shallow network cannot provide? Ignoring the LSTM that compiles across time, the input vector that defines a state at time t seems to be composed of 10 features as noted in Section 3. This is a fairly small number of features. Could this be handled by a shallow LSTM network? If there isn't a clear reason for a deep network, an argument for a shallow network would be lower probability of overfitting and faster training times.

The goal impact metric in Section 6.1 (no Eqn label) is an interesting concept. But it focuses on the impact of the team by looking at the delta of the Q function. The player impact only comes through from weighting that delta by the player's action choice. This seems well-suited to problems where (i) there are small number of players, and (ii) one player is likely to take the "impact" action. It would be great to have discussion on how this approach would generalize beyond such games.

Review #12783

This paper proposes an approach for the evaluation of the performance of ice hockey players with Deep Reinforcement Learning. The method consists in learning an action-value Q function from a very large dataset of over three millions play events. Data are taken from the National Hockey League (NHL). The proposed deep network exploits both continuous context signals and game history, using an LSTM that takes into account ball possession and positioning. Such learned Q-function can be used to evaluate the performance of players in each action, whereas to evaluate a player's overall performance, the paper introduces a novel metric, named Game Impact Metric (GIM), that aggregates values of the single actions where that player is involved.

From the experimental evaluation, GIM is shown to be consistent throughout a play season, and it is shown also to highly correlate with other metrics, success measures, and the salary of players in current and especially future seasons.

The paper is interesting and well presented. It is a very well-grounded application of sabermetrics (originally designed for baseball) to ice hockey, with (deep) reinforcement learning techniques. The experimental evaluation is quite convincing, and in particular the correlation of the novel proposed metric with salaries is very interesting. The publication of the source code for replication of the experiments is another plus of the paper.

- In Section 4, "specifies a value" --> "that specifies a value"

Review #12783

This paper is a very interesting and original use of Deep Reinforcement learning for the evaluation of plays in the National Hockey League.

Whilst not being knowledgeable in the area of player evaluation, I feel this paper's claims seems well supported empirically.

Review #29299

The paper presents a method to derive the probability of success of some play in the game of ice hockey given contextual information about game events. The approach involves applying deep RL to a large dataset including key events in all games of NHL season of 2015-16. The dataset was augmented to include state/context information that could be generalized, ie independently of the specific team and players. Together with the event player's action and a reward function indicating when a goal was scored/conceded, the authors applied deep RL so that the Q values can model the probability of scoring in the future. The authors present a new player evaluation metric GIM that takes into account frequency and impact of each player's action. GIM was then tested for its correlation with standard performance measures and results show strong correlation, above other standard methods. The authors also found strong correlations with raises in contracts in subsequent seasons.

Overall, the paper is clear, well written and easy to follow. The background provides sufficient information on the used techniques and the use of the new metric is well motivated. The results are also clear and do show the usefulness of using the proposed approach as a quantitative measure to evaluate player's performance. The paper uses standard deep RL techniques but uses them to produce this novel metric which shows the impact of AI-based evaluation techniques which is  significant in the current context of big data availability. I particularly liked how the authors could evaluate the importance of player actions that are not directly related, but can potentially lead to, scoring goals.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Some comments / considerations:

Regarding related work, the approach in [Hausknecht and Stone, 2015] should be discussed in greater detail as, according to the authors, is the most similar work. Specifically, how are the architectures similar and different?

The reward structure is a bit confusing. As far as I understood, the authors compute a separate Q function (and therefore a different probability) for each scoring type (home, away, none). Does this mean that the rewards are sparse and only occur when some team scores? Also the notion of Q\_team is confusing in this context: does this equate to Q\_home or is that a generic definition for all 3 possibilities (in which case, Q\_neither is Q\_no\_team)? Consider revising this terminology. Finally, are weights learned/optimized in separate for each Q function or together?

Regarding the player evaluation metric, I did not understand why the \*change\* in Q value is used. Shouldn't the impact of an action be inferred directly from the Q value as it already gives the probability for success of executing the action in the given game context? Why is the change needed then? This should be clarified in the text.

In the results section, why the low correlation (of all metrics) for some statistics? And why couldn't GIM overcome the associated problems? This could be shortly discussed in the paper as such low correlation pop-up in comparison to the other results.

Finally, the authors found high correlations between individual events (dictated by how likely they will lead to a goal) and player performance. This seems to suggest that the context in isolation (ie without considering positions of teammates and corresponding actions) is sufficient to predict the chance of a goal occurring. If this is correct, does this mean that the teammates do not have a role in individual performance? Is there some avenue in the future of using the proposed approach and the GIM score to determine correlations between the teammates and their importance for individual performance? Because to me it seems strange that individual events where the puck is located are that informative in isolation to everything else in the game.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Minor aspects:

- Pg 1 "[Routley and Schulte, 2015] used states" -> Routley and Schulte [2015]

- Pg 2, "There are two agents Home resp. Away representing the home resp. away team." -> I find these sentences (there are more like this in the text) harder to read than eg: 'There are two agents: Home and Away representing the home and away team, respectively'

- I could not understand Fig 4. From the description, this represents the 3rd period of \*one\* specific game. As such, if the reward function only has a value (of 1) when a team scores or game ends, then shouldn't Q/scoring probability be 0 everywhere except in the play just before the goal? I do not understand the osculations in Q values throughout the period. Maybe I am missing something but a better explanation of this could improve the interpretation of this Fig.

- In pg 5, what is TPG, it seems there is no definition for this and does not appear in Table 5. Maybe TOI?

- The information about the maximum trace length of the LSTM (10) is only given in parenthesis in sect 7.1. Maybe this information can be included earlier in section 5 when explaining the whole learning architecture.

- The authors claim that they model the problem as a Markov game, but this would imply the notion of different agents (eg different player positions) working and learning together. However what is shown seems more like an individual agent problem, where the procedure is agnostic to who is doing the action. In this sense there is no multiagent learning problem but a single agent one.