Towards a real-time possession value framework in ice hockey

Frans Murto

Tampere University
frans.murto@tuni.fi

Abstract. Measuring the individual performance of players is an important task in sports analytics. Traditional statistics-based approaches for evaluating hockey players fail to account for context and long-term impact. Recent advances in data gathering have enabled valuing possessions and actions directly to address these issues. This talk describes the implementation of the first real-time possession value framework for ice hockey.

Keywords: ice hockey · possession value · tracking data

1 Introduction

Being able to objectively quantify performance in ice hockey has important implications for player evaluation and acquisition. Traditional goal- and shot-based statistics are problematic in this regard as they ignore the impact of preceding plays and do not differentiate between situational contexts. Advanced metrics derived from expected goals address the latter issue by incorporating information on how dangerous the shots a team takes or faces, but are unable to directly measure the value of non-shot actions and fail to consider situations where no shot is taken.

With the increasing availability of high-granularity event data streams and tracking data in recent years, various approaches in different team sports have been proposed for directly modeling the value of individual actions or possessions [1, 2, 4]. Related work in ice hockey has been scarce [5, 6], however, owing to a lack of the aforementioned data at least in the public domain. Our work intends to bridge this gap by proposing and implementing the first hockey-specific framework for deriving the total value of any given possession in real time, as well as valuing the risk and reward of individual actions separately.

2 Methodology

Estimating the value of possessions can be framed as a Markov decision process (MDP) where the possible actions a player in possession of the puck can take is represented by the discrete set A for all possible match states S. The actions we consider to be part of this action space are shots, passes, moving with the puck, dump-ins and dump-outs. Each action can be further separated depending on their outcome: whether a shot results in a goal, a pass reaches its intended target, a player maintains possession while moving with the puck, a dump-out is recovered by a teammate or a dump-out successfully exits the zone.

Players can be assumed to perform actions that intend to increase the probability of scoring for their team and decrease the probability of conceding a goal. Some actions that are valuable offensively, however, are inherently more risky despite their greater reward. To represent this trade-off we model the return of an action separately for both outcomes with success states yielding the probability of a team in possession scoring and failure states the probability of a team in possession conceding. We assume that only successful actions have positive returns and unsuccessful actions have negative returns, as even though some failed actions may lead to a positive outcome (e.g. a missed pass is received by another teammate in a relatively good position), from a modeling standpoint we want to consider only the intended target.

A common approach for representing the immediate and future impact of an action has been to use either a time window or a fixed number of future actions to assign positive labels for actions that end up affecting the score and negative labels for actions that do not [3, 4]. We experiment with different approaches and observe in our data that using a fixed window of eight seconds from the start of an action provides the best balance between short- and long-term return. When modeling reward actions we label actions that lead to the team in possession scoring a goal in the next eight seconds as positive, and when modeling risk we label actions that lead to conceding a goal in the next eight seconds as positive.

To formalize this we follow the definition of Fernández et al., where the value of a possession P_t is taken to be the total expectation of all actions in a given state [4]. The probability to take action a and its expected value are learned from $\mathbf{X}_t$, which is the feature vector representation of state s derived from a tracking data snapshot at time t.

$$\mathbb{E}[P_t] = \sum_{a \in A} \Big[\mathbb{E}[A = a | \mathbf{X}_t] \quad \mathbb{P}(A = a | \mathbf{X}_t) \Big] \tag{1}$$

As the outcome of successful and unsuccessful actions is modeled separately for all actions except shots (which we assign a fixed value of zero risk due to lack of a true failure condition), the expectation of an action can be generally decomposed as the difference between its expected reward and risk and how likely it is to succeed or fail. Because we assume a single end location for moves, the expected value of a move action follows this formulation exactly.

$$\begin{aligned} \mathbb{E}[A = \text{Move} | \mathbf{X}_t] = \mathbb{E}[A = \text{Move}_{Success} | \mathbf{X}_t] \quad \mathbb{P}(A = \text{Move}_{Success} | \mathbf{X}_t) \\ - \mathbb{E}[A = \text{Move}_{Failure} | \mathbf{X}_t] \quad \mathbb{P}(A = \text{Move}_{Failure} | \mathbf{X}_t) \end{aligned} \tag{2}$$

Because any teammate excluding goalies can be considered as the possible receiver of a pass, we take the expected value of a pass action to be the total expectation of all possible passes. We define an additional transition probability

$\mathbb{P}(R_t|\mathbf{X}_t)$ for all receivers R_t to represent how likely player r becomes the receiver of a pass.

$$\mathbb{E}[A = \text{Pass}|\mathbf{X}_t] = \sum_{r \in R} \Big[\mathbb{E}[A = \text{Pass}, R_t = r|\mathbf{X}_t] \quad \mathbb{P}(R_t = r|\mathbf{X}_t) \Big] \tag{3}$$

For dump-ins we consider all players except the goalie of the team in possession to be able to recover the puck. As dump-ins do not always have an intended receiver, but a general location instead, we model their success probability as $\mathbb{P}(A, R_t|\mathbf{X}_t)$ to represent how likely player r is to recover the dumped-in puck first. We take the expected value of a dump-in action to be the difference between the total expectation of all teammates T and all opponents O that can recover a dump-in.

$$\begin{aligned} &\mathbb{E}[A = \text{DumpIn}|\mathbf{X}_t] = \\ &\quad \sum_{r \in T} \Big[\mathbb{E}[A = \text{DumpIn}, R_t = r|\mathbf{X}_t] \quad \mathbb{P}(A = \text{DumpIn}, R_t = r|\mathbf{X}_t) \Big] \\ &\quad - \sum_{r \in O} \Big[\mathbb{E}[A = \text{DumpIn}, R_t = r|\mathbf{X}_t] \quad \mathbb{P}(A = \text{DumpIn}, R_t = r|\mathbf{X}_t) \Big] \end{aligned} \tag{4}$$

For dump-outs we consider all opponents except the goalie to be able to intercept the puck in the offensive zone. We model $\mathbb{P}(A, I_t|\mathbf{X}_t)$ to represent how likely opponent i is to intercept the dump-out and use the total interception probability to determine the success probability. The expected value of a dump-out action is taken to be the difference between expected reward from successfully exiting the zone and the total expectation of risk for each possible interceptor.

$$\mathbb{P}(A = \text{DumpOut}_{Success}|\mathbf{X}_t) = 1 - \sum_{i \in I} \mathbb{P}(A = \text{DumpOut}, I_t = i|\mathbf{X}_t) \tag{5}$$

$$\begin{aligned} &\mathbb{E}[A = \text{DumpOut}|\mathbf{X}_t] = \\ &\quad \mathbb{E}[A = \text{DumpOut}_{Success}|\mathbf{X}_t] \quad \mathbb{P}(A = \text{DumpOut}_{Success}|\mathbf{X}_t) \\ &- \sum_{i \in I} \Big[\mathbb{E}[A = \text{DumpOut}, I_t = i|\mathbf{X}_t] \quad \mathbb{P}(A = \text{DumpOut}, I_t = i|\mathbf{X}_t) \Big] \end{aligned} \tag{6}$$

3 Data & Modeling

We use event and tracking data from the Liiga regular season matches 2020-21 and 2021-22 provided by Wisehockey. The player and puck tracking data is gathered using an indoor positioning system and sampled at a frequency of 20 Hz. The tracking snapshots are automatically synchronized by the system to align with the start timestamps of the events. We split 75% of the 872 matches in our data to train our models and use the remaining 25% as the test set. Events where the team whose scoring probability is being modeled faces an empty net are omitted due to the dynamics of scoring changing significantly in such situations. An overview of the events is presented in Table 1.

Table 1. Event data counts for the Liiga seasons 2020-21 and 2021-22.

Event type	Success	Total	Training	Test
Pass	79.87%	535,449	401,583	133,866
Move	95.35%	899,809	674,851	224,958
Shot	5.29%	78,141	58,608	19,533
Dump-in	39.67%	86,579	64,931	21,648
Dump-out	69.88%	48,836	36,630	12,206

Because the system provides situations where players move with the puck as continuous sequences, we split the puck controls into discrete one-second length actions. We assume that moves have one possible end location and set this as the player's position one second into the future based on their velocity at the time of the event. For unsuccessful passes we determine the intended receiver using a nearest-neighbor approach. We take the direction and velocity of the known trajectory of the pass and project it forward from the starting position of the pass at different time steps. The intended receiver is then chosen as the player who is closest to the projected end positions most frequently.

Each model uses positional features engineered from tracking data like location, velocity, direction, distance to goal, angle to goal, distance to puck and angle to puck for the player in possession of the puck and the target of the action (if applicable). To represent the local context of an action these features are calculated for the closest teammate and opponent of the player and the target. We also adapt the pressure model of Andrienko et al. to a hockey context and to consider the velocity of the players [7]. For the global context of an action we derive features using hierarchical clustering with two clusters on the players' locations to represent the tactical structure of both teams.

For the pass and shot expectation models we calculate features relating to the positioning of the goalie and how much of the net the goalie has to cover to make a save against the shooter or pass target. Pass expectation models also include information about the crowdedness and width of the pass lane between the passer and the target. The dump-in models have additional features based on the area where the puck is estimated to be played in for a given target. We use a simple convolutional neural network that takes raw tracking data snapshots as its input to estimate this end location. Finally, we include game and score state features in the action probability model to represent how teams adjust their playing style depending on the overall match situation.

We train our models using XGBoost[8] and optimize hyperparameters for tree depth and various regularization parameters using Bayesian optimization with cross-validation. To prevent overfitting into particular feature combinations we use moderate to high regularization parameter bounds for each model. As many of the learning tasks for our models consist of imbalanced data, we use a low max delta step parameter to ensure well-calibrated posterior probabilities. Based on domain knowledge we enforce monotonicity constraints on some features like distance to goal as this improves predictive performance and helps in dealing with outliers like goals scored from the neutral zone. Model metrics on the test set are presented in Table 2. We summarize the calibration of the models using expected calibration error (ECE), which takes the weighted average of the difference between binned output probabilities and data points.

Table 2. Model performance on the test set.

Model	**Log loss**	**AUC**	**ECE**
Action probability	0.186		
Pass success	0.423	0.803	0.037
Pass receiver	0.422	0.892	0.007
Pass scoring for	0.071	0.823	0.018
Pass scoring against	0.045	0.757	0.008
Move success	0.155	0.856	0.008
Move scoring for	0.053	0.791	0.012
Move scoring against	0.026	0.764	0.005
Dump-in success	0.165	0.873	0.006
Dump-in scoring for	0.029	0.743	0.002
Dump-in scoring against	0.015	0.750	0.003
Dump-out success	0.207	0.858	0.008
Dump-out scoring for	0.019	0.735	0.002
Dump-out scoring against	0.032	0.755	0.005
Shot scoring for	0.174	0.827	0.015

4 Results

As we can determine the total value of any given possession P_t, we can use this to evaluate the impact of each action. The value of an action is then defined by taking the difference between total possession value at the start of the action P_{Start} and the total possession value at end of the action P_{End} for all actions except shots, for which the total value is defined as-is. We define actions that have a positive impact as progressive and divide actions that have a negative impact into three categories: regressive, lost and conceded. Regressive refers to how much value is lost through successful actions that decrease the probability of the team scoring, lost refers to how much value is lost through unsuccessful actions, and conceded refers to the value gained by the opponent through unsuccessful actions. To account for any possible error in the detection of the moment of reception for passes, dump-ins and dump-outs, we take a one-second window after the end timestamp of the action to determine its end value.

The value of move actions is calculated continuously during matches, so we take the cumulative sum of a puck control sequence as the total value of a move action. As the value of move actions is the most sensitive to how other players are positioned, and we do not want to punish the puck carrier for factors outside their control, we limit the amount of regressive and lost value between two consecutive puck controls by the average value of a move action in our data (approximately one goal scored per 100 moves) and divide this limit by the sampling rate used to generate tracking data snapshots.

Because there are no existing ground-truth labels for evaluating the quality of a player's actions, we compare the total estimated possession value of common in-game situations and how often they have led to a goal being scored in the Liiga playoffs 2022-23. A comparison of the best forwards and defencemen that shows how their actions have added and lost value is also presented. We use playoff matches as our out-of-sample prediction because there is a recognized difference in play styles between the regular season and the playoffs, where players are allowed by the referees to play a more physical and disruptive game. This is illustrated by the fact that in our training data the average number of goals scored per match is 5.3, while in the 2022-23 playoffs it was 4.6 goals. We surmise that the features engineered from tracking data provide enough context to help overcome this domain shift.

Table 3 displays the danger level between controlled zone entry types at even-strength. We take the danger of a zone entry to be average of the total possession value in a one-second window following the moment that the offensive blueline has been crossed. We denote a zone entry to have resulted in a goal using the same eight-second window as with our models. Our framework correctly identifies that breakaways and odd-man rushes are likely to result in more dangerous situations than entries where the team in possession has an equal or lower number of players involved compared to the defending team. The estimated danger level generally correlates well with how often the different entry types result in goals, though some types are affected by the natural variance of goal-scoring in relatively small sample sizes. Based on observing individual entries, it is also fairly common for entry types to change a couple seconds after the entry is performed in the favor of the defending team as their forwards hustle to join the backcheck.

Table 3. Even-strength controlled entry danger by type in the Liiga 2022-23 playoffs.

Entry type	Count	Danger	Goal scored
1-on-0	18	12.7%	11.1%
2-on-1	24	7.4%	8.3%
1-on-1	43	5.4%	2.4%
3-on-2	70	4.6%	4.3%
2-on-2	146	4.3%	2.8%
1-on-2	90	3.6%	4.4%
3-on-3	473	3.3%	2.8%
2-on-3	310	3.1%	1.6%
1-on-3	101	2.6%	1.9%

Table 4 displays the danger level of successful even-strength passes by type. We take the danger level of a pass to be its progressive impact as defined above. Passes into the slot, which end in the area between the goal and the first hash marks of the faceoff circle, are intuitively the most dangerous with lateral (east-west) passes and behind-the-net passes inside the offensive zone following suit. We define entry passes as passes that facilitate a zone entry and outlet passes as passes that lead to a zone exit. Low-to-high and high-to-low passes are passes inside the offensive zone that do not cross the center line. It can be seen that the value added by these pass types closely follows the true scoring probabilities.

Table 4. Even-strength pass danger by type in the Liiga 2022-23 playoffs.

Pass type	Count	Danger	Goal scored
Slot	195	9.1%	9.2%
Lateral	994	3.6%	3.4%
Behind the net	226	3.3%	3.1%
Entry	611	2.3%	2.0%
Low-to-high	610	2.1%	1.8%
High-to-low	1142	1.4%	1.8%
Outlet	1631	1.1%	0.9%

Tables 5 and 6 represent the best-performing forwards and defencemen by total value added and lost at even-strength play. Passes, dump-ins and dump-outs have been consolidated together into a single value. The positive impact of an action indicates the progressive value added and the negative impact indicates the sum of regressive, lost and conceded value as defined before. Forwards generate more progressive value on average, but tend to lose more as well through attempting more ambitious and difficult actions. A greater proportion of the negative impact by defencemen is through conceded value, however, as their unsuccessful actions tend to occur closer to their defensive zone.

Table 5. Even-strength forward performance per 60 by possession value type in the Liiga 2022-23 playoffs.

Player	Team	Pass+	Pass−	Move+	Move−	Shot+	Total
Eemeli Suomi	Ilves	0.95	0.53	1.21	0.50	1.00	2.13
Anton Levtchi	Tappara	1.08	0.52	1.16	0.26	0.59	2.05
Joona Ikonen	Ilves	0.88	0.38	0.58	0.15	1.10	2.03
Waltteri Merelä	Tappara	0.48	0.39	0.93	0.34	1.20	1.88
Kristian Tanus	Tappara	0.91	0.49	0.91	0.24	0.73	1.82
Balázs Sebők	Ilves	1.04	0.81	1.31	0.23	0.50	1.81
Santeri Virtanen	Ilves	0.69	0.38	0.66	0.20	0.99	1.76
Matias Mäntykivi	Ilves	1.03	0.45	1.00	0.44	0.59	1.73
Niko Ojamäki	Tappara	0.75	0.35	0.44	0.15	0.97	1.66

Table 6. Even-strength defenceman performance per 60 by possession value type in the Liiga 2022-23 playoffs.

Player	Team	Pass+	Pass−	Move+	Move−	Shot+	Total
Les Lancaster	Ilves	1.06	0.43	0.33	0.10	0.82	1.68
Tarmo Reunanen	Lukko	0.76	0.28	0.68	0.20	0.58	1.54
Colby Sissons	KalPa	0.69	0.31	0.79	0.17	0.40	1.40
Valtteri Kemiläinen	Tappara	0.93	0.22	0.60	0.16	0.22	1.37
Maksim Matushkin	Tappara	0.69	0.20	0.70	0.21	0.33	1.31
Ben Thomas	Tappara	0.58	0.27	0.44	0.09	0.63	1.29
Leo Lööf	Ilves	0.82	0.32	0.62	0.21	0.15	1.06
Casimir Jürgens	Tappara	0.59	0.28	0.42	0.09	0.42	1.06
Thomas Grégoire	Lukko	0.71	0.28	0.43	0.12	0.22	0.96

5 Conclusions

In this talk we have shown that a real-time possession value framework can be implemented in an ice hockey context. The out-of-sample performance of the framework in typical in-game scenarios and events as well as in differentiating the value created by players in different positions matches both domain knowledge and the true underlying scoring probabilities. In the future it would be interesting to extend the framework to model defensive actions and the possibility of shots being indirect passes. Similarly, considering banked and rimmed passes separately would likely improve the performance of pass-related models [9]. Another promising avenue of research would be to use graph-convolutional neural networks with tracking data snapshots, which has been shown to improve model performance over tree-based models and remove the need for advanced feature engineering [10].

References

1. Sicilia, A., Pelechrinis, K., Goldsberry, K: DeepHoops: Evaluating micro-actions in basketball using deep feature representations of spatio-temporal data. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2096–2104. (2019)
2. Burke, B.: DeepQB: Deep learning with player tracking to quantify quarterback decision-making & performance. 2019 MIT Sloan Sports Analytics Conference (2019)
3. Decroos, T., Bransen, L., Van Haaren, J., Davis, J.: Actions speak louder than goals: Valuing player actions in soccer. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1851–1861. (2019)
4. Fernández, J., Bornn, L., Cervone, D.: A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions. Machine Learning, **110**(6), 1389–1427 (2021)

5. Liu, G., Schulte, O.: Deep reinforcement learning in ice hockey for context-aware player evaluation. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, pp. 3442–3448. (2018)
6. Liu, G., Schulte, O., Poupart, P., Rudd, M., Javan, M.: Learning agent representations in ice hockey. In: Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2020 (2020)
7. Andrienko, G., Andrienko, N., Budziak, G: Visual analysis of pressure in football. Data Min Knowl Disc **31**, pp. 1793–1839 (2017)
8. Chen, T., Guestrin, C.: XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 785–794. (2016)
9. Radke, D., Brecht, T., Radke, D.: Identifying completed pass types and improving passing lane models. In: Proceedings of the Linköping Hockey Analytics Conference LINHAC 2022, pp. 71–86. (2022)
10. Stöckl, M., Seidl, T., Marley, D., Power, P.: Making offensive play predictable - using a graph convolutional neural network to understand defensive performance in soccer. 2021 MIT Sloan Sports Analytics Conference (2021)