Net Expected Threat: a Tool for Scouting in Hockey

Adrian Cadena-Medina[*], Pablo L. Landeros[†], Evan Cheney[‡]

March, 2021

1 Introduction

Imagine a five-play possession that ended in a goal. Box scores would list the scoring player and up to two assisting players; however, how would we credit the other players who helped develop the possession? Which offensive action was the most impactful and significant to put the team in a position to score? We borrowed a soccer analytics concept called Expected Threat (xT) developed by Karun Singh to answer these questions. Besides adapting xT to hockey, we expanded it by adding time-remaining-in-period and turnover probability/opposing threat. We name the new metric: Net Expected Threat (nxT).

nxT recognizes that at any point, a player possessing the puck has two options: to shoot (attempt to score) or to move the puck (zone-entry or pass). If a player chooses to shoot, the score-probability will depend on field location. If a player decides to move, there are multiple probable destinations, each with a different probability of success, depending on field location. The difference between nxT at the start and nxT at the end of the action is called nxT generated (nxTg). nxTg assesses how much a player helped the team place the puck in a position to score, regardless of the outcome of the possession. Another benefit of the xT models is the ability to divide credit. Take our five-play scoring-possession, for example. We can divide the nxTg of each move action over the possession's total nxTg. The result would be credit percentages to calculate which action or player had the highest impact. In this submission, we explain how nxT expands xT to analyze a broader range of outcomes. Then, we describe the methodology behind nxT. Finally, we utilize nxT generated (nxTg) to grade offensive players for scouting purposes.

2 Problem statement

Hockey analysts utilize metrics such as expected goals to grade offensive hockey players. However, there is still a need to quantify how each move action generates scoring-threat. Expected Threat Generated (xTg) is a partial solution, but some limitations exist. xT only considers field position when assessing the potential threat. Also, xT does not account for the probability of losing the puck; therefore, we cannot measure the impact of unsuccessful move actions (turnovers). As a result, xT can only quantify the impact of completed passes and successful zone-entries. There is a glaring opportunity to generate a metric that quantifies the impact of successful and unsuccessful move actions, regardless of possession outcome, while incorporating additional variables. By successfully expanding xT, we would quantify the impact of every offensive move action. The new metric would facilitate the scouting process by quickly analyzing large amounts of data and identifying threat-generating players.

[*]https://adriancadena.com/
[†]https://pablolopez2733.github.io/
[‡]cheney380@gmail.com

3 Solution

The data for this analysis was provided by Stathletes. Thanks to the Erie Otters for allowing Stathletes to publish a sample data set from their 2019/2020 season. Thanks to Karun Singh for the developing the methodology behind Expected Threat.

We expanded xT by adding time-remaining-in-period to the score probability estimation. We then generated score-probability matrices to implement them into our xT model. Accounting for the probability of a turnover required a dynamic programming approach, similar to Singh's process. We split the field into 128 cells: 8 width cells and 16 length cells. Then, we estimated the probability of losing the puck at each cell, each probable turnover cell location, and the hypothetical opposing xT for each turnover cell location. We multiply each cell's probability of turnover by their corresponding xT and sum the entries to obtain Opposing Expected Threat (oxT). oxT is the Expected Threat that the defensive team has at a given moment. By calculating the difference between the possession team's expected threat (xT) and the opposing team's expected threat (oxT), we obtain the Net Expected Threat (nxT). nxT accounts for the risk of turnover and the danger that the opposing team represents based on probable turnover location. Net Expected Threat is our proposed solution.

$$nxT_{x,y,t} = xT_{x,y,t} - oxT_{x,y,t} \tag{1}$$

To quantify each move action's impact (nxTg), we calculate the difference in nxT before and after play i for team e.

$$nxTg_{e,i} = nxT_{e,i+1} - nxT_{e,i} \tag{2}$$

By using oxT, we can measure the impact that a turnover had on nxT for team e during play i

$$nxTg_{e,i} = oxT_{e,i+1} – nxT_{e,i} \tag{3}$$

nxT is the Expected Threat that team e had before the turnover (as the possession team). oxT is the Expected Threat that team e has after the turnover (now as the defensive team). As Equation 3 shows, oxT is needed to estimate the effect of turnovers. The previous methodology of xT was limited to Equation 2 and ignored unsuccessful actions: when the possession team loses the puck. In conclusion, we are accounting for both the risk of losing the puck and the threat that the opposing team would represent depending on turnover location.

4 Methodology

Thanks to Decroos et al. [2019] for their open-source module SoccerAction. We leveraged their xT model code as a blueprint to build our computations.

Terminology:

- Grid: 16-length x 8-width grid with 128 different cells. We split the rink into zones or cells
- Move action: a pass attempt or a zone-entry attempt
- Time-remaining-in-period (t): We use four different time-remaining bins t_1 = More than 15 minutes, t_2 = between 10 and 15 minutes, t_3 = between 5 and 10 minutes, and t_4 = less than 5 minutes
- Move probability $m_{x,y,t}$: the rate of times a player opted to move in a cell (x, y) during time t
- Shoot probability $s_{x,y,t}$: the rate times of a player attempted to score in a cell (x, y) during time t

- Goal probability $g_{x,y,t}$: the scoring probability in a cell (x, y) during time t when a player shoots. Estimated using a score-probability model including variables: x, y, Euclidean distance, and time-remaining-in-period. The algorithm was random forest.
- Move transition matrix $T_{x,y}$: the probability of a player moving from the current cell (x, y) to another cell (z, w). Each cell contains 128 different probabilities
- Turnover transition matrix $L_{x,y}$: the probability of a turning over the puck from the current cell (x, y) to any other cell (z, w). Each cell contains 128 different probabilities.

Expected Threat

First, we generate the xT by following a similar methodology to Singh but adding the extra feature t. Let $xT_{x,y}$ be the expected payoff of a cell (x, y). For simplicity, we will assume all only one bin of t. The incorporation of t will be explained later.

1. If a player shoots, then $xT_{x,y} = g_{x,y}$. If we are uncertain of the player's decision, we multiply shoot probability $s_{x,y}$ times $g_{x,y}$
2. If the player moves, there are 128 possible destinations, each one with a different expected payoff. Suppose a player moves from (x, y) to (z, w). In that case, the expected payoff of (z, w) is $xT_{z,w}$ multiplied times the probability of moving to (z, w) from (x, y). Since there are 128 possible locations, we estimate $xT_{z,w}$ of all (z, w) locations and sum the results. The result is $xT_{x,y}$ when moving. We use the transition matrix $T_{x,y}$ to estimate this. If we are uncertain of the player's decision, we multiply the probability of moving $m_{x,y}$ times $xT_{x,y}$
3. If we add $xT_{x,y}$ when shooting and $xT_{x,y}$ when moving, both under uncertainty of decision, we get the total xT, which we simply call $xT_{x,y}$

Thus we can express Expected Threat in cell (x, y) as:

$$xT_{x,y} = (s_{x,y} \times g_{x,y}) + (m_{x,y} \times \sum_{z=1}^{16} \sum_{w=1}^{8} T_{(x,y)\to(z,w)} xT_{z,w} \tag{4}$$

Solving for xT

For simplicity, we will assume all only one bin of t. the incorporation of t will be explained later.

To solve Equation 4, Singh proposes assigning $xT = 0$ to all cells (x, y) and then evaluate the formula iteratively until convergence. At each iteration, we assess the new xT for each zone using xT values from the previous iteration. Each new iteration utilizes $xT_{x,y}$ values of the prior iteration as $xT_{z,w}$. Therefore, the more iterations, the higher the possible number of actions before the shoot-attempt we are considering. In other words, with 10 iterations, we are looking at up to 10 moves ahead of each scenario! When we end our simulation for every bin, we obtain an 8x16 matrix with 128 values, each one representing the xT of a given cell. This matrix is simply called xT. Figure 1 represents the dynamic between transition matrices for each cell and xT.

Accounting for the time remaining in the period

To incorporate t, we created four different score probability matrices utilizing a random-forest algorithm. Figure 2 explains the difference in score probability by every bin of t.

Since we added an extra feature to the model, we need to solve the equation four times to utilize the corresponding score-probability matrix. We split the training set into four subsamples based on time-remaining-in-period. Finally, we train to model correspondingly, obtaining thus $m_{x,y,t}$, $s_{x,y,t}$, and $g_{x,y,t}$. The two exceptions are move transition matrix $T_{x,y}$ and turnover transition matrix L_{xy}.

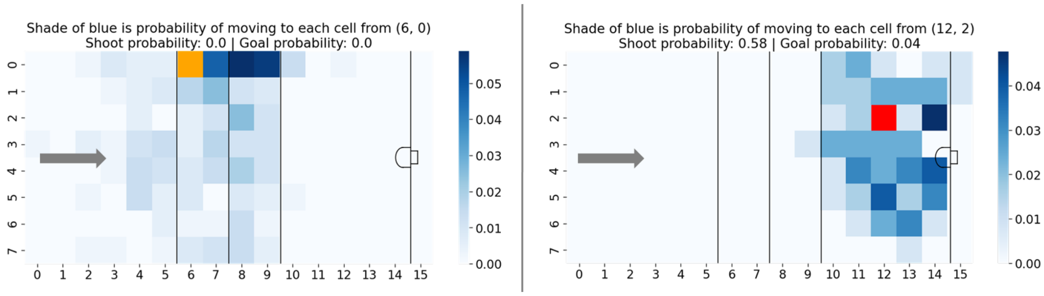

Figure 1: Move Transition Matrix of cells (6,0) and (12,2). Shade of red represents Expected Threat

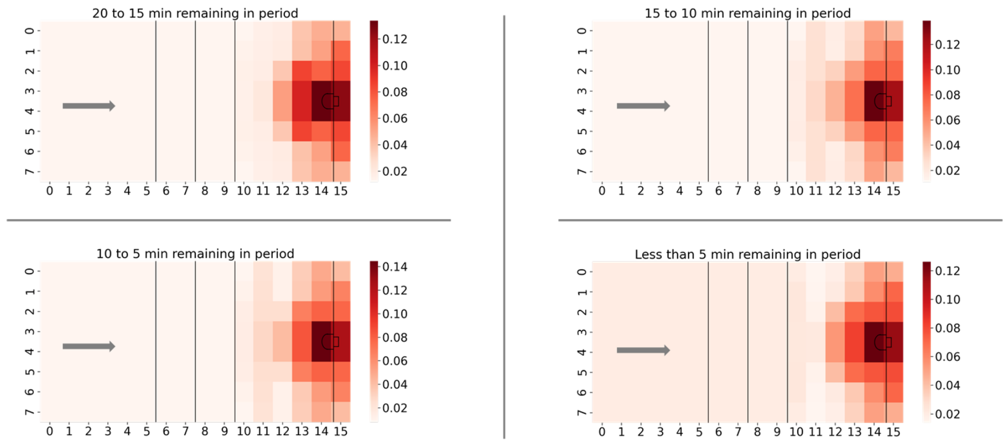

Figure 2: Score Probability Matrix depending on (x,y) and time remaining in half

$$xT_{x,y,t} = (s_{x,y,t} \times g_{x,y,t}) + (m_{x,y,t} \times \sum_{z=1}^{16} \sum_{w=1}^{8} T_{(x,y)\rightarrow(z,w)} xT_{z,w,t} \tag{5}$$

Opposing and Net Expected Threat

We developed Opposing Expected Threat (oxT) and a Net Expected threat (nxT) to solve our problem statement. To generate oxT, we followed a similar approach to the move transition matrix ($T_{x,y}$). Still, we utilized failed move-actions (turnovers) instead of successful ones. We call it the turnover transition matrix ($L_{x,y}$). We use $L_{x,y}$ to estimate the probability of turning over the puck at every possible cell in the rink. Figure 3 presents Turnover Transition Matrix for cells (14,3) and (5,0.)

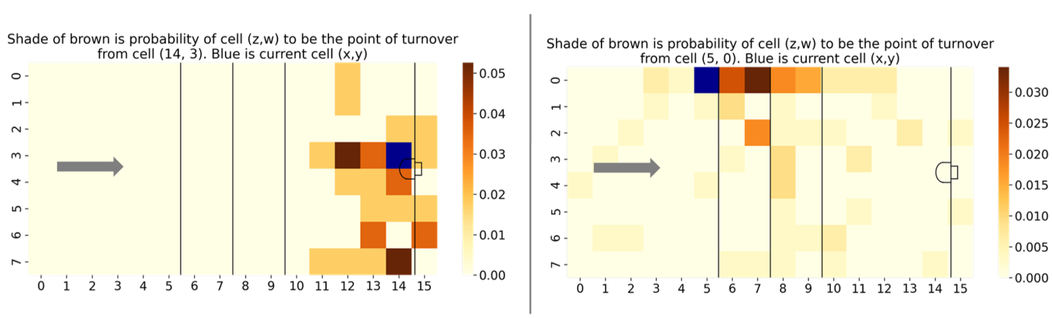

Figure 3: Turnover Transition Matrix of cells (14,3) and (5,0)

The next step is to use the xT grid we generated in our original xT model. By calculating the Element-wise product of $L_{x,y}$ and xT_t, and summing over all the entries, we obtain $oxT_{x,y,t}$, as Equation 6 shows. $oxT_{x,y,t}$ is the opposing expected threat in the cell (x, y) during time t. Think of $oxT_{x,y,t}$ as the defensive team's expected threat when the offensive team is in a cell (x, y) during time t. The sum of all entries of $L_{x,y}$ is the probability of turnover in point (x, y).

$$oxT_{x,y,t} = \sum_{i=1}^{16} \sum_{j=1}^{8} L_{x,y} \circ xT_t \tag{6}$$

To obtain the nxT of cell (x, y) during time t, we calculate the difference between $xT_{x,y,t}$ and $oxT_{x,y,t}$

$$nxT_{x,y,t} = xT_{x,y,t} - oxT_{x,y,t} \tag{7}$$

Since we have four different matrices of xT depending on the bin of t, we need to run the process four times. Figure 4 presents the xT, oxT, and nxT grid, respectively.

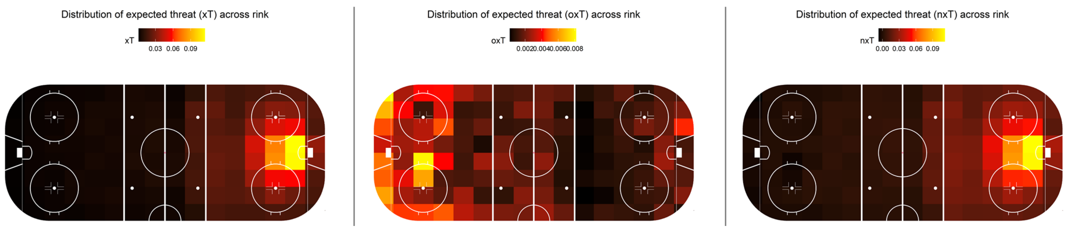

Figure 4: xT, oxT, and nxT observed in the sample between 20 and 15 minutes of time remaining in period

5 Results

Using Net Expected Threat, we can grade hundreds of player's offensive production regardless of their position. Every move action generates xT irrespective of how close the play happened to the opposing net and whether the possession ended in a goal or not. nxT can quantify the effect of both successful and failed move attempts; therefore, we penalize players making mistakes accordingly. One of our model's key features is that it correspondingly penalizes turnover based on location: a turnover near the possession team's net is more costly.

To utilize nxT to grade players, we obtain the top ten players in terms to total and average nxTg in our sample. Also, we visualize an example of how to use xT for the proper division of credit during a given play.

Top 10 Players in Net Expected Threat generated by play
Minimum 500 plays

Player	Team	nxT	plays	nxT per play
Maxim Golod	Erie Otters	8.16	1982	0.0041
Jamie Drysdale	Erie Otters	6.10	1661	0.0037
Drew Hunter	Erie Otters	3.44	1058	0.0033
Jacob Golden	Erie Otters	4.06	1280	0.0032
Jack Duff	Erie Otters	4.12	1374	0.0030
Kurtis Henry	Erie Otters	4.18	1430	0.0029
Chad Yetman	Erie Otters	4.14	1512	0.0027
Kyen Sopa	Erie Otters	1.31	527	0.0025
Cameron Morton	Erie Otters	1.70	701	0.0024
Daniel D'Amato	Erie Otters	2.64	1087	0.0024

Data: 40 game BDC scouting dataset.

Top 10 Players in Total Net Expected Threat generated
Minimum 50 plays

Player	Team	nxT	plays	nxT per play
Maxim Golod	Erie Otters	8.16	1982	0.0041
Jamie Drysdale	Erie Otters	6.10	1661	0.0037
Kurtis Henry	Erie Otters	4.18	1430	0.0029
Chad Yetman	Erie Otters	4.14	1512	0.0027
Jack Duff	Erie Otters	4.12	1374	0.0030
Jacob Golden	Erie Otters	4.06	1280	0.0032
Drew Hunter	Erie Otters	3.44	1058	0.0033
Emmett Sproule	Erie Otters	2.91	1225	0.0024
Brendan Sellan	Erie Otters	2.72	1199	0.0023
Daniel D'Amato	Erie Otters	2.64	1087	0.0024

Data: 40 game BDC scouting dataset.

Figure 5: Ranking players in terms of Total nxT and nxT per play. Data provided by Stathletes. Erie Otters 2019/2020 season

This data sample consisted overwhelmingly of Erie Otter games, as they provided their 2019-2020 season. Therefore, the data is very unbalanced, and the results should be analyzed accordingly. To make fair comparisons, we utilized a high number-of-plays threshold. Very few players outside of the Erie Otters fell into that category. Also, every non-Erie Otter team in the sample only faced the Erie Otters, so selection bias exists. In conclusion, our results are significant for analyzing Erie Otters players, but not so for players in other teams. A larger sample size with multiple games per team is needed to assess every player in the league correctly.

Other uses of nxT: division of credit

Suppose we take a single possession (a string of plays/actions). In that case, we can assign a credit percentage to each action. The possession we use as an example ended in a goal. The most impactful play was Chase Stillman's assist to Quinton Byfield.

nxT Credit Distribution in a Sunbury Wolves goal

Player	Event	nxT added	net nxT change %
Emmett Serensits	Pass	0.009331764	11.68
Chase Stillman	Zone Entry	0.015886822	19.88
Chase Stillman	Pass	0.054676825	68.44
Quinton Byfield	Goal	-	-

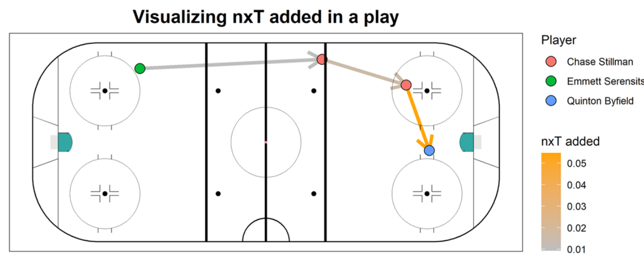

Figure 6: Division of credit

We invite teams and scouting departments to implement a version of this metric to boost their scouting efforts and provide an extra layer of information to their evaluation and decision-making process. We create an open-source python module available **here** to facilitate our metric's implementation.

References

Tom Decroos, Lotte Bransen, Jan Van Haaren, and Jesse Davis. Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '19, pages 1851–1861, New York, NY, USA, 2019. ACM. ISBN 978-1-4503-6201-6. doi: 10.1145/3292500.3330758. URL http://doi.acm.org/10.1145/3292500.3330758.

Karun Singh. Introducing expected threat (xt). URL https://karun.in/blog/expected-threat.html.