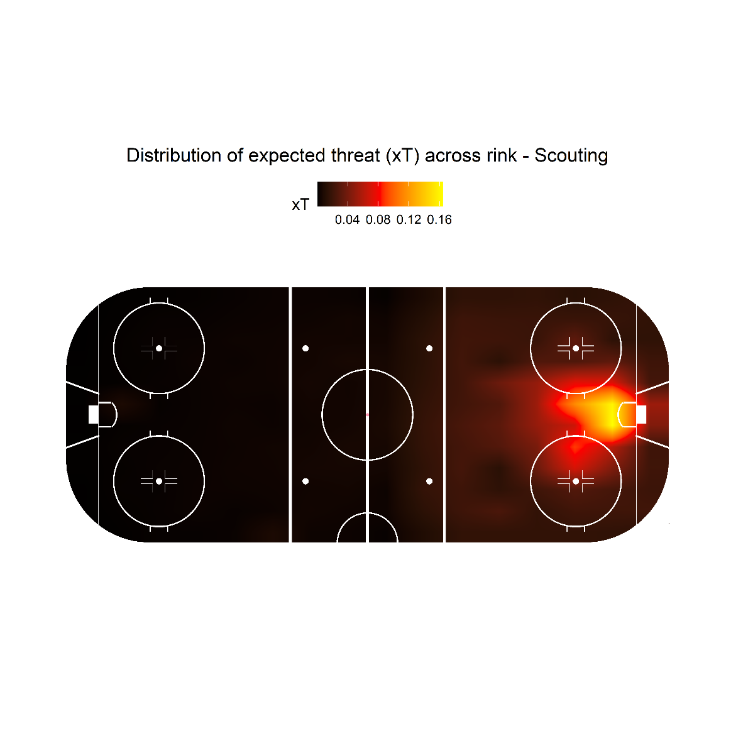
**Introduction**

How can we measure offensive production beyond assists, goals, and expected goals in hockey? Imagine a five-play possession that ended in a goal; the credited players would be the scoring player and up to two assisting players. However, how are we crediting 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? These are the questions that our analysis answers.

We borrowed a soccer analytics concept called Expected Threat (xT), developed by Karun Singh to grade offensive production based on the field location at the start vs. the end of the play. In addition to adapting xT to hockey, we expanded it by adding time-remaining-in-period and turnover probability. We name the new metric: Net Expected Threat (nxT).

**In soccer, the original xT model recognizes that at any point, a player possessing the ball has two options: to shoot (attempt to score) or to move the ball (dribble 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. xT generated (xTg) considers all possibilities and assesses by how much a player helped the team place the ball in a position to score, regardless of the actual outcome of the possession. To understand xT from a hockey perspective, interchange "ball" with "puck" and "dribble" with "zone-entry." One of the many benefits of this metric is the ability to divide credit. Take our five-play scoring-possession example. In that scenario, we can divide the xTg of each move action over the possession's total xTg. 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.

**Problem statement**

Hockey analysts utilize standard metrics such as assists and goals to grade offensive hockey players. The analytics revolution brought metrics like expected goals (xG). However, there is still a need to quantify each move action's impact to create a scoring threat. Expected Threat (xT) is a partial solution, but there are some limitations. xT only considers field position when assessing probabilities of shooting, scoring, attempting to move, and successfully moving the puck. For example, our xG model suggests that time-remaining-in-period significantly alters score probability regardless of field position. 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 effect 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 and incorporates time-remaining-in-period. By successfully addressing these limitations, 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 regardless of position.

**Solution**

We decided to add time-remaining-in-period to an xG model, along with field location, to address one of xT's limitations. Then, we used the xG model to generate score-probability matrices and implemented them in our xT model. At this stage, our xT model is very similar to Singh's, except for adding an extra variable to generate the score-probability matrix.

Accounting for the probability of a turnover required a dynamic programming approach, similar to Singh's process when creating the move transition matrix T\_x\_y for his xT model. 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. By multiplying each cell's probability of turnover by their corresponding hypothetical xT and summing the results, we obtain Opposing Expected Threat (oxT).

We can account for the probability of turnover by obtaining the difference between the possession team's Expected Threat (xT) and the opposing team's Expected Threat (oxT). The resulting metric is called Net Expected Threat (nxT), which is our proposed solution.

nxT = oxT - xT

nxT tells us the level of threat that each team represents at every point based on field location (cell) and time-remaining-in-half. To quantify each move action's impact, we calculate the difference in nxT before and after each play.

nxTg = nxT after the play - nxT before the play

By using oxT, we can measure the impact of a turnover by doing:

nxTg in turnover = (oxT at the end of the play) – nxT

nxT is the Expected Threat that team *i* had before the turnover (as the possession team). oxT is the Expected Threat that team *i* has after the turnover (now as the defensive team). This formula is different from equation b), which is the formula for nxTg when the play is successful.

As equation c) shows, oxT is needed to estimate the effect of turnovers. The previous methodology of xT was limited to equation b) and ignored unsuccessful actions: when the possession team loses the puck.

**Methodology**

Thanks to the SoccerAction team for their open-source module, we leveraged their xT model as a blueprint and codebase to build our own hockey xT module. Special thanks to Karun Singh for developing the methodology behind xT.

**Preparing Data**

The first step is to generate the xT. The methodology in this section will be very similar to Singh's. To prepare the data, we need to estimate each play's start and end location and identify move actions. Move actions in our model are pass-attempt and zone-entry. Finally, we need a success binary variable to determine whether the move action or the goal attempt was successful.

**Splitting field**

The next step is to split the field into zones or cells; this facilitates the analysis and avoids overfitting. We used a 16 x 8 grid. We created four different time-remaining bins (𝑡): more than 15 minutes, between 10 and 15 minutes, between 5 and 10 minutes, and less than 5 minutes.

Every cell has a different:

* Move probability 𝑚𝑥𝑦𝑡: the probability of opting to move in a cell (𝑥,𝑦) during time 𝑡. Estimated by dividing the number of move actions over the number of total actions (move actions + goal attempts)
* Shoot probability 𝑠𝑥𝑦𝑡: probability for attempting to shooting in a cell (𝑥,𝑦) time 𝑡. Following a similar approach to move probability
* Move transition matrix 𝑇𝑥𝑦: the probability of a player moving from the current zone to any other location. Each zone or cell contains a matrix of length 16 x 8, totaling 128 different probabilities. For sample size reasons, this computation does not take time-remaining-in-period into consideration
* Goal probability 𝑔𝑥𝑦𝑡: Score probability in a cell (𝑥,𝑦)(x,y) during time 𝑡 when a player shoots. Here is where our initial 𝑥𝑇 model differs from Singh's. Instead of calculating the score-rate per area, we developed an Expected Goal 𝑥𝐺 model including 𝑥 location, 𝑦 location, Euclidean distance, and time-remaining-in-period (𝑡). Then, we created score probability matrices for each bin using our 𝑥𝐺 model to predict score probability in different scenarios.

Chart

Description automatically generatedChart, bar chart

Description automatically generated

Chart

Description automatically generatedChart, bar chart

Description automatically generated

Example of Move transition matrix T\_x\_y in cells (11,4) and (4,6). The figure includes shoot and goal probability

Chart

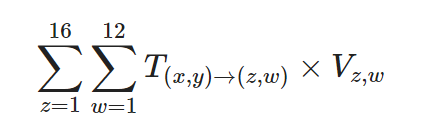
Description automatically generatedChart

Description automatically generated

**Deriving xT**

Following Singh's methodology, we assign V\_x\_y as the value of zone (x,y). As we mentioned, at any point, a player decides to either move or shoot. If a player shoots, we simply use score probability g\_x\_y as expected payoff.

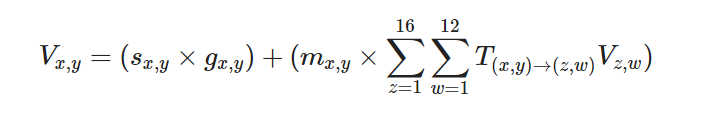
If the player moves, there are 128 possible destinations. Suppose the player moves to a destination (z,w), which is the end location of a play that started in (x,y). In that case, the expected payoff is the value of zone (z,w): V\_z\_w. We estimate the probability of moving to each area (z,w) and the respective expected payoff V\_z\_w. Finally, we proportionally weigh the payoffs. We use the transition matrix T\_x\_y to estimate the expected payoff of moving the ball with the following equation:



8

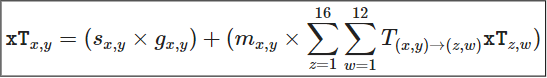
Once we know the expected payoff of shooting and moving the ball using the following equation

8



,t

We assign a value to each location based on shooting threat and the potential to generate threat later in the possession. Therefore, the value of V\_x\_y is the Expected Threat (xT).



**Accounting for the time remaining in the period**

During our model, the average number of iterations was 10. Since we added an extra feature to the model: *t,* we solve this equation four times to utilize the corresponding score-probability matrix. For that, we split the training set into four subsamples based on time-remaining-in-half. Finally, we train to model correspondingly, obtaining thus m\_x\_y\_t, s\_x\_y\_t, and g\_x\_y\_t.

**Dynamic programming comes in handy**

We need to know xT of all possible end locations (xT\_z\_w) to solve the xT\_x\_y equation. Singh proposes assigning xT = 0 to all cells (x,y) and 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 values from the previous 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*\_*t*.

**The wonders of Net Expected Threat**

As detailed during the solution segment, we developed an oxT and nxT. To generate oxT, we followed the same approach to create the 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 and weighed the hypothetical xT that the other team would have in that scenario. By doing that, we obtained oxT.

To obtain hypothetical xT, we used the xT matrix from our original xT model. By calculating the Hadamard product of L\_x\_y, xT\_t and summing over all the entries, we obtain the value oxT\_x\_y\_t.

oxT\_x\_y\_t = sum(L\_x\_y ⊙ xT \_t)

Then, to obtain the nxT for a given cell, we calculate the difference between the two values xT\_x\_y\_t and oxT\_x\_y\_t

nxT\_x\_y\_t = xT\_x\_y\_t - oxT\_x\_y\_t

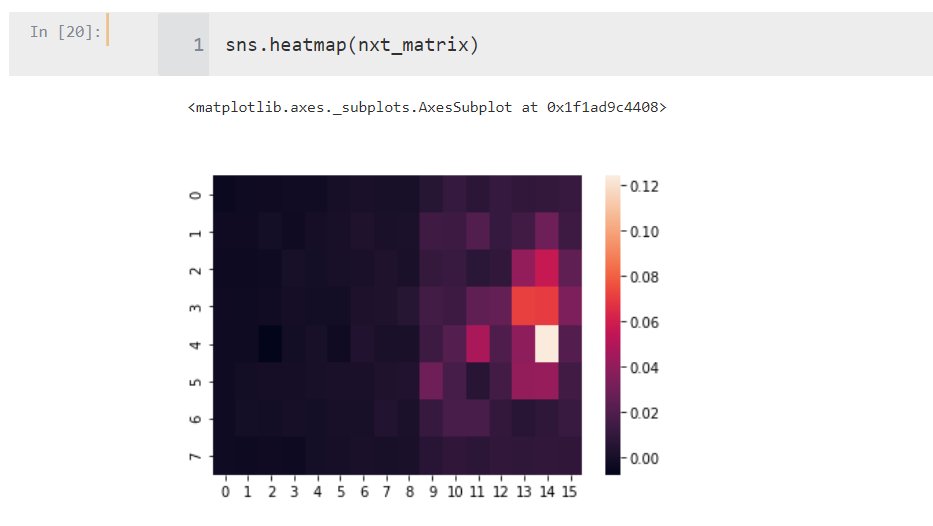
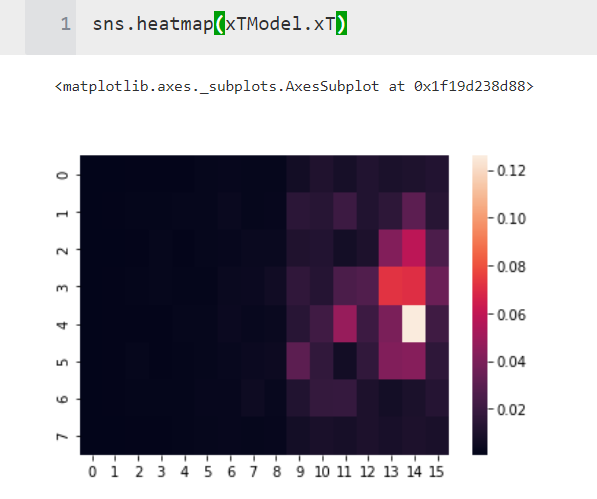
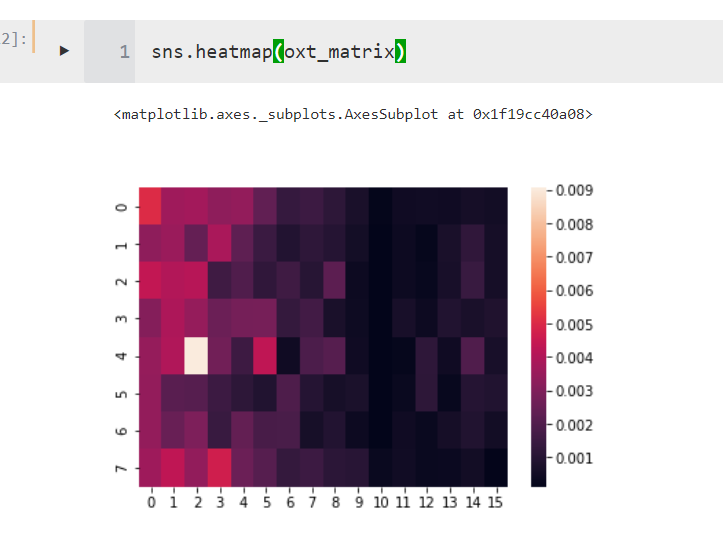
We have four different matrices xT, depending on the bin of *t*; therefore, it is necessary to run the process four times.

Figure x presents L\_x\_y. Figure z presents xT, oxT, and nxT grid respectively.

**A picture containing qr code

Description automatically generatedA picture containing chart

Description automatically generated**

****

**Results**

To utilize nxT to grade players, we obtain the top and bottom 10 players and teams in nxTg in our sample. Also, we visualize an example of how to use xT for the proper division of credit during a given play.

**Action points**

Using this metric, we were able to grade thousands of players in the scouting league in 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, players making mistakes are penalized accordingly.

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.

We are presenting examples of visualizations to analyze players and teams using nxTg.

*\* I am sure Evan can add something cool here\**