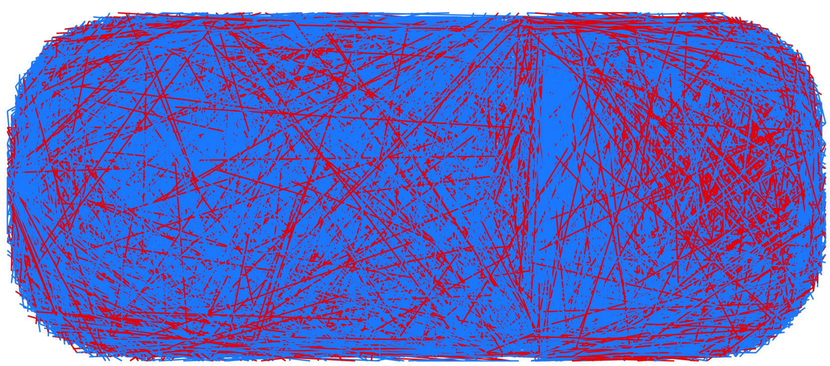
It's a Great Data for Hockey

Michael Egle, Iowa State University
2021 Big Data Cup

February 18, 2021



1 Introduction

Across all sports, most metrics are focused on scoring, and with good reason. More points leads to better chances of winning. Expected points added in football quantifies each play in terms of points by taking down, distance, and many other variables into consideration. The hockey equivalent is expected goals. When a player takes a shot, what's the probability that the average player would score in that scenario? These metrics are remarkably useful. But what about passing? Goal scoring is the most exciting element of hockey, but a majority of plays in a hockey game are the passes that lead up to a shot on goal rather than the shot itself. Some players in sports like basketball, soccer, and hockey just seem to have great intuition and can make smart decisions on the fly. In this paper, I ask the question: how can one attempt to quantify the decision-making of an individual hockey player based on the locations of the passing player and the intended pass recipient?

2 The Data

The data used for this project is a 40 game play-by-play sample of the 2019/2020 Erie Otters (Ontario Hockey League) season made public for the 2021 Big Data Cup. Alongside the basic time, period, and team name data, it includes the event at each point of the game, such as face-off wins, shots on goal, passes, etc. Most notably, this data includes coordinates for where each event occurs, including the start and end coordinates of each pass made during the game. These coordinates can be used to construct the metric.

3 Quantifying Decision-Making

Many metrics that employ expectation models focus on the outcome of the play. Unlike other metrics, mine only evaluates the player's decision-making process. Is the team better off because of a player's decision? At each point in time, a given player has x expected goals from their current location. At the same point, the player can choose to pass the puck to a teammate who has y expected goals at a different location. Each pass has a risk. There's a chance that the intended receiving player could mishandle the puck and the opposing team will gain possession. There's also the chance that an opposing player will intercept the puck before it arrives. I define a passing player's intuition score (I) for an individual pass as:

$$I = xG_2 * P(\text{completed pass}) - xG_1$$

Where xG_1 is the expected goals from the point of the passing player and xG_2 is the expected goals from the point of the receiving player.

4 Models

I use two logistic regression models. One will calculate the expected goals and the other will calculate the probability of a completed pass.

4.1 Expected Goals

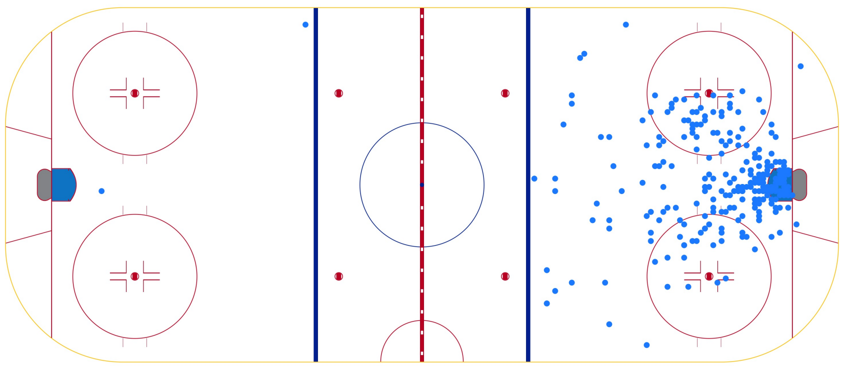


Figure 1: Location of each successful shot on goal in the data.

Intuitively, there's a higher frequency of goals as the shooting player gets closer to the goal. It also seems that there's a slight advantage to have an angle on the goal rather than a straight shot (perhaps because the shooting player has a better chance to get their shot behind the goaltender). As a result, the expected goals model uses the following information:

- Distance from goal (in feet)
- Whether the possession team has a power play
- How large the player advantage is if it is a power play
- Angle formed by the goal and the player in question

The number of goals an individual player or team scores above their baseline expectation can easily be calculated using this model, but that is outside the scope of this paper.

4.2 Pass Completion Probability

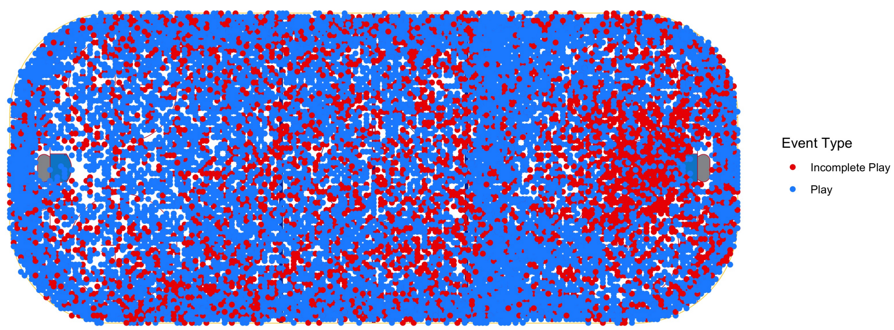


Figure 2: Location of each pass's endpoint in the data.

Passes with an endpoint closer to the goal are much less likely to be completed. For similar reasoning as shot distance, I decided to include pass distance to the model. The pass completion probability model uses the following information:

- Distance traveled by the pass (in feet)
- The puck's distance traveled in the x direction (in feet)
- Whether the possession team has a power play
- How large the player advantage is it it is a power play
- Distance from goal for passing player (in feet)
- Distance from goal for the intended receiving player (in feet)

5 Results

When calculating the results for each player, passes that move backwards in the x direction are filtered out since they are likely intended to be used to "reset" on offense which is irrelevant to what this metric is attempting to quantify. I define an individual player's intuition rate as the proportion of passes where the intuition score is greater than 0. The top ten players in the OHL in 2019/20 in intuition rate are listed below (Minimum of 40 passes in the dataset).

	Player	Team	Intuition Rate	Observations
1	Luke Beamish	Erie Otters	0.85	93
2	Cameron Morton	Erie Otters	0.79	86
3	Brendan Kischnick	Erie Otters	0.79	52
4	Francesco Pinelli	Kitchener Rangers	0.69	42
5	Fedor Gordeev	Guelph Storm	0.67	66
6	Jack Duff	Erie Otters	0.66	184
7	Jacob Golden	Erie Otters	0.64	356
8	Drew Hunter	Erie Otters	0.63	296
9	Cole Perfetti	Saginaw Spirit	0.62	42
10	Kurtis Henry	Erie Otters	0.62	239

Another way to evaluate players using this metric is by looking at each player's average pass intuition. However, I omitted the average because the numbers were very small and there wasn't much difference between all players.

It is important to note that these results are very likely biased in favor of the Erie Otters since they have the most observations in the dataset. As seen above, their players make up seven of the top ten in intuition rate. They appear in all 40 games while no team plays more than five games in this data. This metric can also be used to evaluate passing at the team level, as seen below (Minimum of 200 passes in the dataset).

	Team	Intuition Rate	Observations
1	Kitchener Rangers	0.51	569
2	Hamilton Bulldogs	0.51	232
3	Mississauga Steelheads	0.51	421
4	Windsor Spitfires	0.50	532
5	Saginaw Spirit	0.50	374
6	Sault Ste. Marie Greyhounds	0.50	233
7	Sudbury Wolves	0.48	359
8	Owen Sound Attack	0.48	345
9	Flint Firebirds	0.48	362
10	Erie Otters	0.47	7375

6 Conclusion and Future Work

In this paper I defined intuition as the difference of the current expected goals and the expected value of a pass. Then calculated the rate of decisions where the intuition rate is greater than 0 for each individual player and team. There were a number of limitations in creating this method of passing evaluation. As I mentioned earlier, this data only includes games in which the Erie Otters are playing, which might lead to some bias in the models. Of course, there are many more factors that can be useful in predicting the outcome of a pass that perhaps weren't included in the data. At the NHL level, tracking data could be very useful to create tools similar to this one. Such data would significantly improve the quality of this metric by taking into account how close the nearest defender is, whether the nearest defender has good positioning, and so many other factors. Despite the room for improvement, this metric provides a useful insight for the data.

7 Acknowledgements

I would like to thank everyone at Stathletes who worked so hard to put together a phenomenal competition. Open source data competitions such as this one are a great opportunity for the sports analytics community. I also thank the Erie Otters for graciously publishing their play-by-play data from the 2019-20 OHL season. To paraphrase the great Bob Johnson, it's always a great day for hockey.

8 Appendix

- All code can be found in this GitHub repository:

  https://github.com/michaelegle/2021BigDataCup

- Code for creating the hockey rink background can be found in this GitHub repository from Ross Drucker: https://github.com/rossdrucker/Playing-Surfaces