Determining Pass Value and Efficiency in Hockey

Owen Ricketts and Devlin Sullivan

Over the course of recent years, the impact of analytics has had a visible effect on the tactics and decision-making employed by players and coaches. Whether it's the increased prominence of the 1-3-1 power play formation[1, 2], or the tendency for teams to be more selective with shots - increased access to data has exposed weaknesses in previously standard strategies at hockey's highest levels.

With the inclusion of both complete and incomplete pass tracking data made available for the Big Data Cup, we decided to tackle attempted and successful passes as a specific area of investigation. More specifically, we wanted to determine which offensive-zone passes contributed most effectively to goals, while factoring in the risk associated with failing to complete a pass.

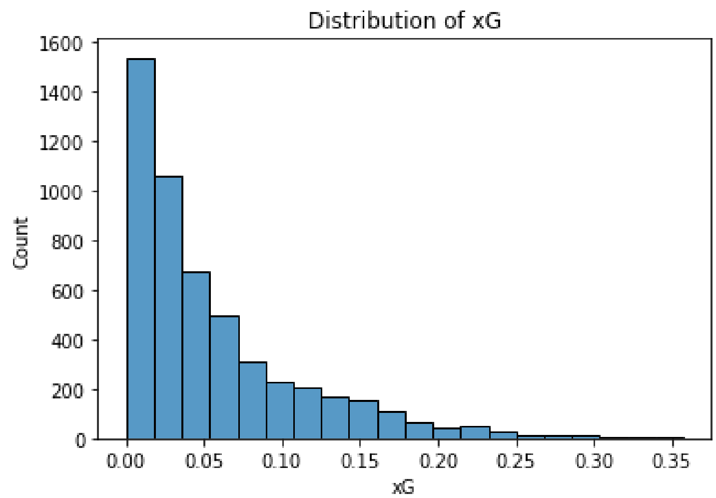
Existing methods of evaluating passes exist across other sports where goals are sparse, namely soccer. One example of how soccer attempts to quantify the value of a pass is Packing - which assigns value to the number of players that a forward pass bypasses or "takes out of the game"[3]. Though this approach has its drawbacks even within soccer analytics, its linear and simple approach is particularly ill-suited to analyzing passes in hockey, where passes can advance behind the net, along the boards, and many of the most dangerous passes are lateral.[4]

Due to these complexities, we aimed to develop a flexible and coordinate-based model for evaluating passes, offering insight on each pass' expected contribution to the team's goal scoring, as well as the expected likelihood of the pass being completed. We also wanted the development of this model to be applicable to lower-level and foreign leagues without robust tracking systems installed in order to be useful for assessing a team or individual's playmaking ability without relying on primary or secondary assists.

Approach

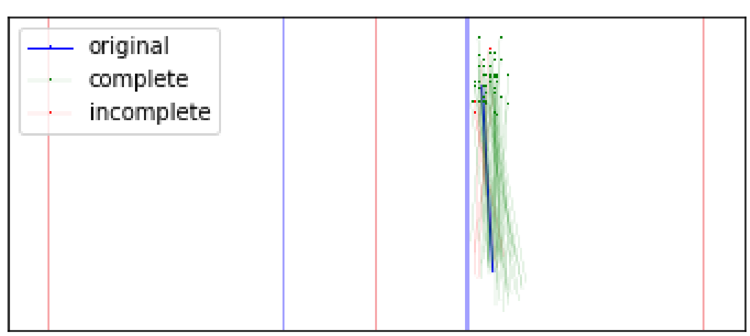
For this project, we used the Erie Otters' scouting dataset, though the same methodology could easily be applied to data from the Olympics or NWHL. The scouting dataset offers the ability to have a larger sample size for evaluating individual skaters, which is why we selected it. The data included was used to generate an expected goals (xG) model, as well as a model for estimating pass success and value. For the purpose of our analysis, we looked exclusively at even strength, offensive-zone events.

xG Model



The xG model we created was a cross-validated logistic regression (`LogisticRegressionCV` from Scikit-learn), which incorporated the following features: time since the last event, last event type, shot release stype, opposing team skaters, if the shot is taken on a powerplay, lateral movement since the last event, X/Y coordinate of the shot, and shot distance. In line with the model at Moneypuck, our most significant contributing factors to a shot's xG were the time since the last event, the distance of the shot, and the shot release type[5]. The model's Compute Area Under the Receiver Operating Characteristic Curve (ROC AUC) score was 0.793, which compares favorably with existing public research[6], but may be subject to underfitting based on the distribution of xG values.

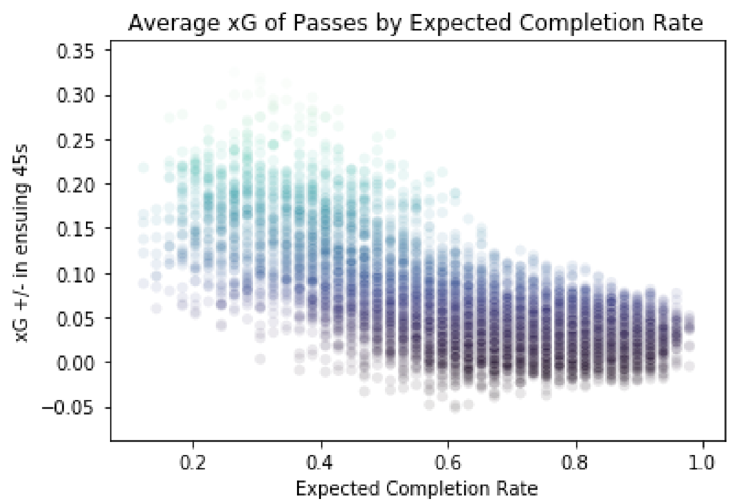
Pass Similarity Model



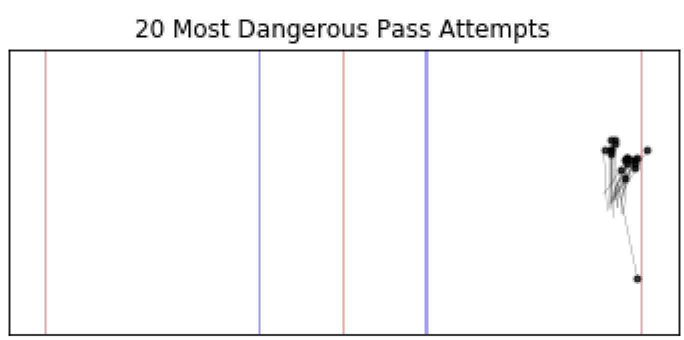
To evaluate pass similarity, we employed a K-Nearest-Neighbors model, the general framework of which had proven valuable in existing research by Daniel Weinberger[7]. To determine pass similarity, we wrote a k-nearest-neighbors model that selected each pass' closest matches by smallest combined distance between their respective origins and destinations. This graph highlights a single pass' nearest neighbors using this methodology. Each pass was evaluated based on it's 50 nearest neighbors (excluding indirect passes).

Applying and Evaluating the Models

In order to assign each pass a value based on its contribution to the team's attack, we utilized the xG model to calculate how much xG a team generated and conceded in the ensuing 45 seconds of game time after the attempted pass. As for the expected completion rate of the pass, we aggregated the expected completion rate of its neighbors. A larger dataset would have afforded the ability to weigh a pass' comparables based on how close they were to the original, but this approach was better suited for the given data.



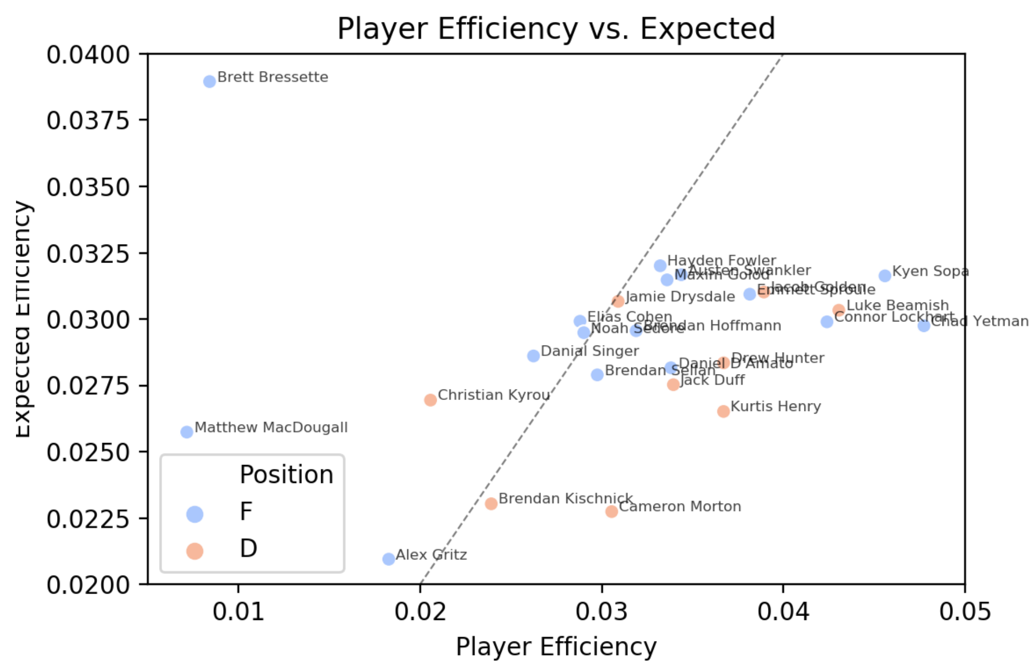
Our first main finding outlines the negative relationship between the expected likelihood of a pass being completed, and the quality of offense that this pass generates. Similarly, our initial findings highlight that the majority of dangerous passes originate from near or below the goal line, and their intended recipient is in the slot. Both of these findings validate the models used, and are consistent with prior public research.



Actionable Insight

Player Evaluation

By multiplying the xG +/- of a successful pass with the probability of a pass' success (Efficiency), we can calculate the expected value added by each pass attempted. This calculation can be used as a basis for determining which players consistently are able to optimize their passes such that they are both likely to be completed and contribute substantially to creating dangerous scoring chances. We can also determine whether or not players exceed their expected completion rate or outperform the projected offensive impact of their passes - likely a sign of strong passing accuracy and decision making.



According to the models we developed, these players generated the most value above expected for Erie. The data also tells us that Kyen Sopa was attempting high-value passes and still outperformed his expected completion rate, while Kurtis Henry was able to generate more offense

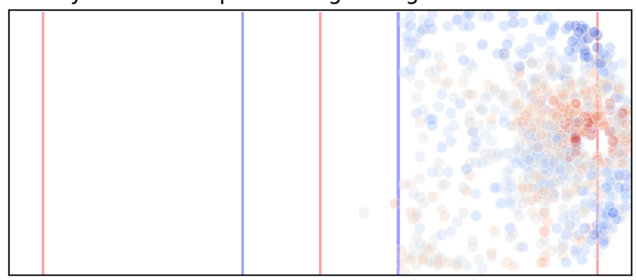
Player	Efficiency Delta	Position
Chad Yetman	0.017991	F
Kyen Sopa	0.013965	F
Luke Beamish	0.012720	D
Connor Lockhart	0.012504	F
Kurtis Henry	0.010190	D

than expected on less valuable passes, which were more often attempted by defensemen than forwards.

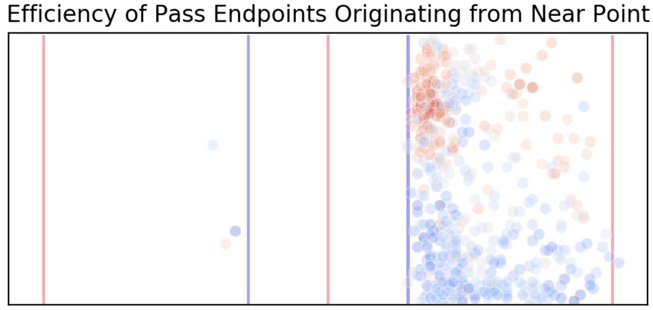
Strategy

Investigating the data also offered some valuable insights about which specific passes are more ideal for teams to play. As shown by the included figures, passes to the slot are ideal when the passer is below the goal line. However, when evaluating point passes, the optimal pass selection changes. While the few passes to the slot attempted still yield decent results, the most efficient passing choices are cross-ice passes (either D-to-D passes, which have a higher probability of completion, or passes down low that bypass the defense on the passer's side of the ice).

Efficiency of Pass Endpoints Originating from Below Goal Line



Efficiency of Pass Endpoints Originating from Near Point



<u>Conclusion</u>

In summary, this paper has outlined the following:

- Low probability, high reward 'home run' passes are worth attempting due to the non-linear relationship between pass difficulty and offense driven
- Efficiency above expected is useful to describe playmaking in the offensive zone that doesn't manifest itself through assists, though a larger sample of data is needed to determine the stability of this data
- Current pass selection from the point and below the goal line could be improved, though analysis of further context is needed to suggest that major problems exist with current offensive-zone passing patterns

Some future improvements on this project may include using more contextual data to refine model performance. Though KNN models are better suited for lower dimensionality, incorporating other features through a general additive model (or similar), such as Statcast's xwOBA model could improve model performance[8]. Furthermore, many of these features used in NFL Next Gen Stats Completion Probability[9], such as pass speed, time on the puck, on-ice traffic, and release type are avenues to improve model performance, and could be investigated with a larger sample size or more expansive data and potentially implemented to improve the model in the future.

References

1. Ward, Joe, and Joe Lemire. "How the 1-3-1 Power Play Works." The New York Times. January 27, 2019. Accessed March 06, 2021. https://www.nytimes.com/interactive/2019/01/27/sports/hockey/power-play-1-3-1-lightning.html.
2. Parnass, Arik. "How Can We Quantify Power Play Performance In Formation?" Hockey Graphs. April 24, 2016. Accessed March 06, 2021. https://hockey-graphs.com/2016/04/25/how-can-we-quantify-power-play-performance-in-formation/#more-8638.
3. Gardiner, Bobby. "Unpacking Packing." StatsBomb. August 03, 2016. Accessed March 06, 2021. https://statsbomb.com/2016/08/unpacking-packing/.
4. Boyle, Chris. "The Most Important Line on the Ice You've Never Heard Of." Ontario Minor Hockey Association. January 08, 2019. Accessed March 06, 2021. https://www.omha.net/news_article/show/486107-the-most-important-line-on-the-ice-you-ve-never-heard-of.
5. MoneyPuck.com -About and How It Works. Accessed March 06, 2021. http://moneypuck.com/about.htm.
6. Anderson, Cole. "Expected Goals (xG), Uncertainty, and Bayesian Goalies." Game Theory. January 03, 2018. Accessed March 06, 2021. http://www.crowdscoutsports.com/game-theory/expected-goal-xg-model/.
7. Weinberger, Daniel. "Passing Clusters: A Framework to Evaluate a Team's Breakout." Hockey Graphs. October 22, 2019. Accessed March 06, 2021. https://hockey-graphs.com/2019/10/22/passing-clusters-a-framework-to-evaluate-a-teams-breakout/.
8. Sharpe, Sam. "An Introduction to Expected Weighted On-Base Average (xwOBA)." Medium. September 24, 2019. Accessed March 06, 2021. https://technology.mlblogs.com/an-introduction-to-expected-weighted-on-base-average-xwoba-29d6070ba52b.
9. "Next Gen Stats: Intro to Completion Probability (Part II)." NFL.com. September 28, 2018. Accessed March 06, 2021. https://www.nfl.com/news/next-gen-stats-intro-to-completion-probability-part-ii-0ap3000000967238.