Moving Beyond Assists: A Bayesian Analysis of Passing Ability

Marc Richards

University of Pittsburgh, Department of Statistics

1 Introduction

Current publicly available statistics on passing in junior hockey leagues are limited to assists; and some leagues don't even breakout primary and secondary assists. Further, simple counts of assists don't tell the whole story of a player's passing ability; for example, one could make a good pass that creates a great scoring opportunity but the receiver doesn't score or misses the pass entirely. Suffice it to say, evaluation of player's passing ability at the junior hockey level is lacking. This makes it difficult for NHL teams to scout these players objectively and for junior hockey coaches to evaluate their own players. Thus, utilizing the event-based dataset provided by Stathletes, we explore Erie Otters players' passing abilities based on the risk and reward.

2 Pass Risk

First, we define the **pass risk** as the likelihood a pass is completed.

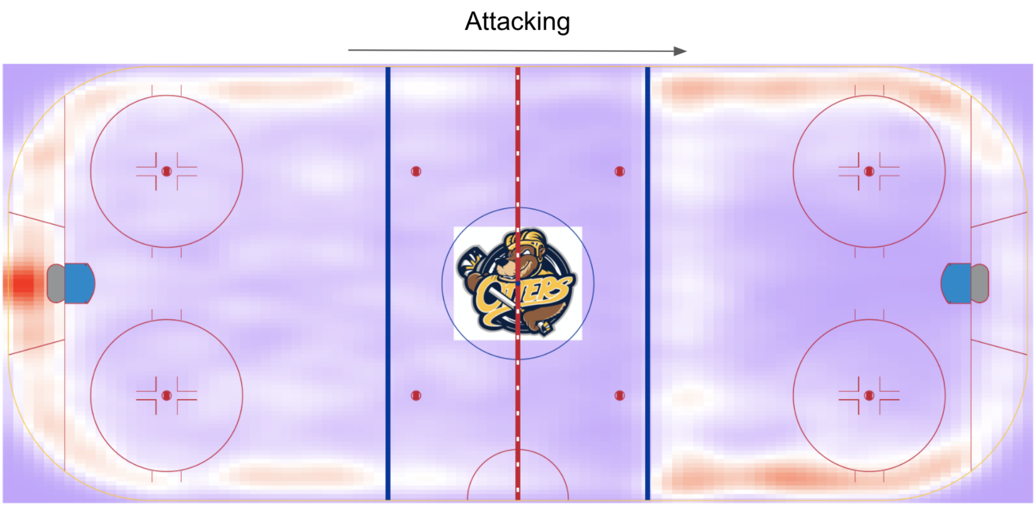
2.1 Expected Completed Pass Model

The purpose of this model is to estimate the probability of a successful pass. Expected Completed Pass models have been widely used in Soccer [American Soccer Analysis] and even in the NHL [Richards RITSAC 2019]. In order to model the probability of a successful pass, we generate a number of features, including but not limited to: the distance of the pass, the angle of the pass, the location of the passer and passing target, the indicator of a direct or indirect pass, and the strength state. The provided dataset consisted of 32,668 passes of which 23,778 were completed. Given the interactive nature of the independent variables, we consider a Bayesian Additive Regressive Trees (BART) model.

We compare BART to the popular Gradient Boosted Trees via the popular XGBOOST, Logistic Regression and a naive model based on predictive accuracy. For comparison, we tested with a 75% training, 25% testing split. To assess the predictive accuracy, we use the log-loss and area under the curve. The Table below shows the loss metrics for each averaged over 10 train/test splits with standard deviations in parentheses. BART out performs all three on both loss perspectives.

Model	AUC	Log-Loss
Naive	0.500 (0.000)	0.584 (0.005)
Logistic	0.651 (0.003)	0.570 (0.008)
XGBOOST	0.752 (0.006)	0.501 (0.006)
BART	0.760 (0.004)	0.497 (0.006)

To help understand how the location of the pass affects the completion probability, the Figure below shows a heat map of the location of the pass by the estimated completion probability. Red generally represents a higher completion probability while blue represents a lower one. The Figure shows that behind a team's own net is where the most passes are completed.

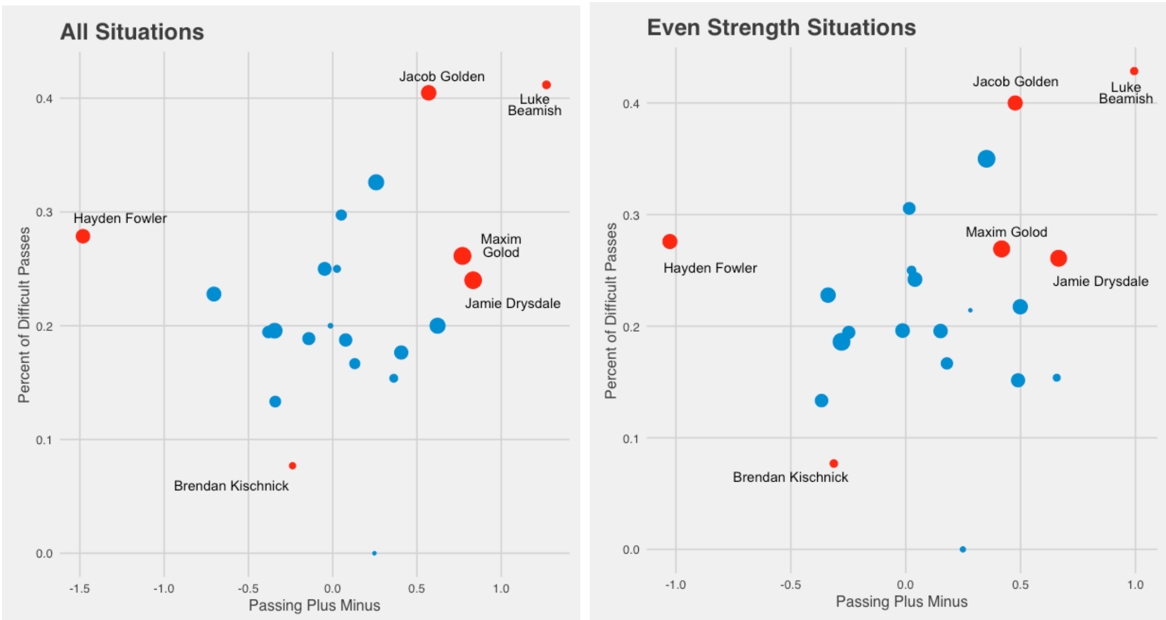


2.2 Pass Risk Metrics

In order to assess the difficulty of a pass, we introduce **Difficult Pass Completion Percentage** (DPP) which measures how many difficult high risk passes a player makes and completes. A difficult pass is arbitrarily defined as a pass that is in the 70th plus percentile of the most high risk passes derived from our completion probability model. Thus, DPP is simply the percent of completed difficult passes divided by the number of difficult passes attempted.

Next, because we modeled the risk associated with completing a pass on average, we expect that players with more passing talent would be more likely to execute a risky pass compared to a less skilled passer. Therefore, we define **Passing Plus Minus** (PPM) as the sum of the difference in the completion probability and the observed outcome of whether or not the pass was completed. We standardize these values over the number of games played.

To illustrate these pass risk metrics, the figures below show the PPM vs DPP for passes made in all situations and Even Strength only situations. A few notable players have been identified with red dots. Among those are team assist leaders and NHL prospects Jamie Drysdale and Maxim Golod. They have among the highest passing plus minus and completed the highest percentage of difficult passes. Defensemen Jacob Golden and Luke Beamish stand out as particularly strong passers, especially at completing difficult passes in a moderate sample size. While Brendan Kischnick shows up at the opposite end of the spectrum as a player that performed well below average in both. Hayden Fowler is interesting for the fact he has attempted and completed a high number of difficult passes but his PPM is among the worst on the Erie Otters.



3 Pass Reward

Previous research in both soccer and hockey has defined pass reward as the likelihood that the pass results in a goal in the next 10, 20 or 30 seconds. While this is a defensible approach, we look to take a different approach to overcome a lack of data as well as to model the uncertainty. Therefore, we define **pass reward** as the likelihood that given the receiver shoots the puck he scores.

3.1 Expected Hypothetical Goal Model

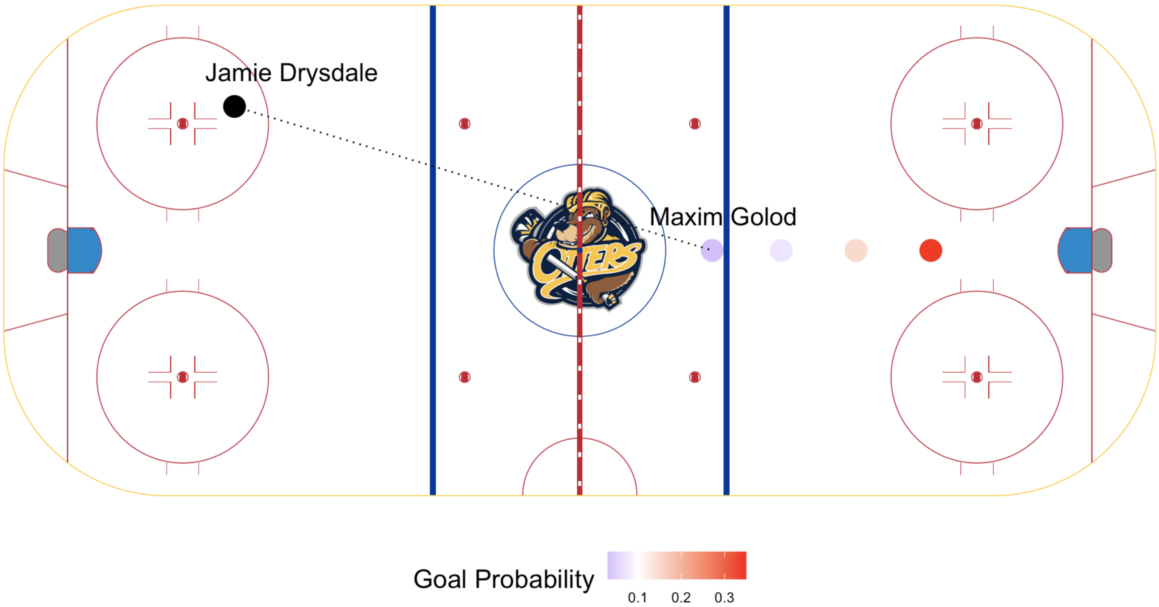
In order to account for the passes that do not result in a shot, we introduce an expected hypothetical goal model. The first step in this model is to estimate the probability of a goal. To do so, we utilize a number of features that previous Expected Goals Models in other leagues have found to be significant [EvolvingWild 2018]. These features include, but are not limited to, the strength-state, the X and Y location of the shooter, the shot type, game state, and the previous event. The dataset included $\sim$ 4,900 shots of which $\sim$ 230 resulted in goals. Similar to the Expected Completed Pass Model, we utilized a tree-based approach in Chipman et al. (2010)'s Bayesian Additive Regressive Tress (BART).

Again, to justify the use and generalizability of the model, we compare BART to a Gradient Boosted Trees model via XGBOOST, Logistic Regression and a naive model. Following the same procedure as we did in section 2.1, we average the accuracy across 10 train/test splits. The Table below shows that BART again outperforms all three models in both loss metrics.

Model	AUC	Log-Loss
Naive	0.500 (0.000)	0.215 (0.025)
Logistic	0.783 (0.004)	0.213 (0.042)
XGBOOST	0.791 (0.004)	0.207 (0.002)
BART	0.811 (0.023)	0.178 (0.002)

3.2 Unobserved Covariates

Next, we consider the hypothetical scenarios when a player receives a pass but does not shoot it and when a player shoots the puck at a different location. To understand this, the Figure below shows a Jamie Drysdale pass to Maxim Golod for a breakaway. In this instance, Golod actually loses the puck and does not get a shot off. However, that does not change the potential opportunity created. Now, if we just fixed the unobserved covariates; for example, we just assumed that Golod shot the puck at the blue-line where he received the pass, then we may underestimate the true pass reward. If he was on a breakaway, he may hypothetically shoot it closer to the goal, increasing the goal probability.



In order to obtain an Expected Goals estimate for these hypothetical scenarios, we must first estimate the unobserved covariates. For each counterfactual shot off the pass, we first divide the covariates into two groups: those that we observe and those that we do not observe. The Table below shows a list of observed and unobserved covariates at the time of shot.

Observed Covariates		Unobserved Covariates	
Period	Strength State	Time Elapsed from Pass	Shot Type
Goal Differential	Previous Event (Pass)	One Timer indicator	Traffic indicator
Previous X Location	Previous Y Location	X Location	Y Location

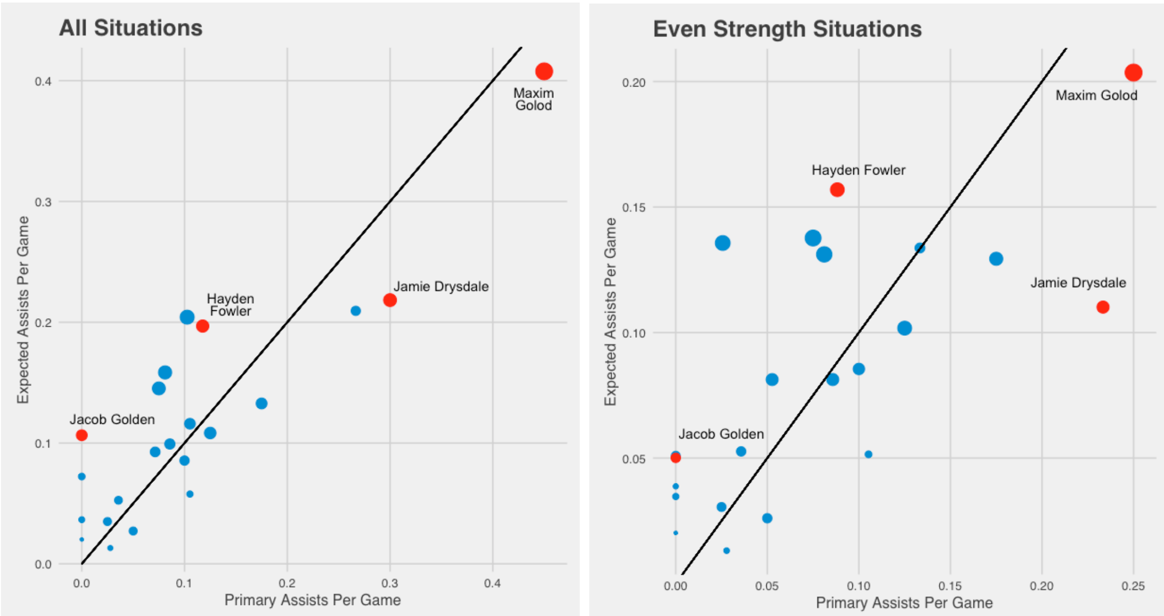
More formally, we let $X = (X_{\text{observed}}, X_{\text{unobserved}})$ and sample $X_{\text{unobserved}}$ from the empirical distribution of the actual observed shots. Thus, just as in Deshpande and Evans EHCP model, if we know the true value of f, then the Expected Hypothetical Goal Probability could be approximated by:

$$\text{EHGP}(X) = \mathbb{E}[F(X_{\text{observed}}, X_{\text{unobserved}})] \approx \frac{1}{M}\sum_{m=1}^{M} F(X_{\text{observed}}, X^{(m)}_{\text{unobserved}})$$

where $X^{(1)}_{\text{unobserved}}, \ldots, X^{(m)}_{\text{unobserved}}$ are the draws of unobserved covariates from the empirical distribution, $F(.) = [1 + e^{-f(.)}]^{-1}$ is the forecasted probability, and the expectation is taken over the empirical distribution of unobserved covariates. Rather than setting to a fixed quantity, we average over the uncertainty in the unobserved values allowing us to account for the uncertainty.

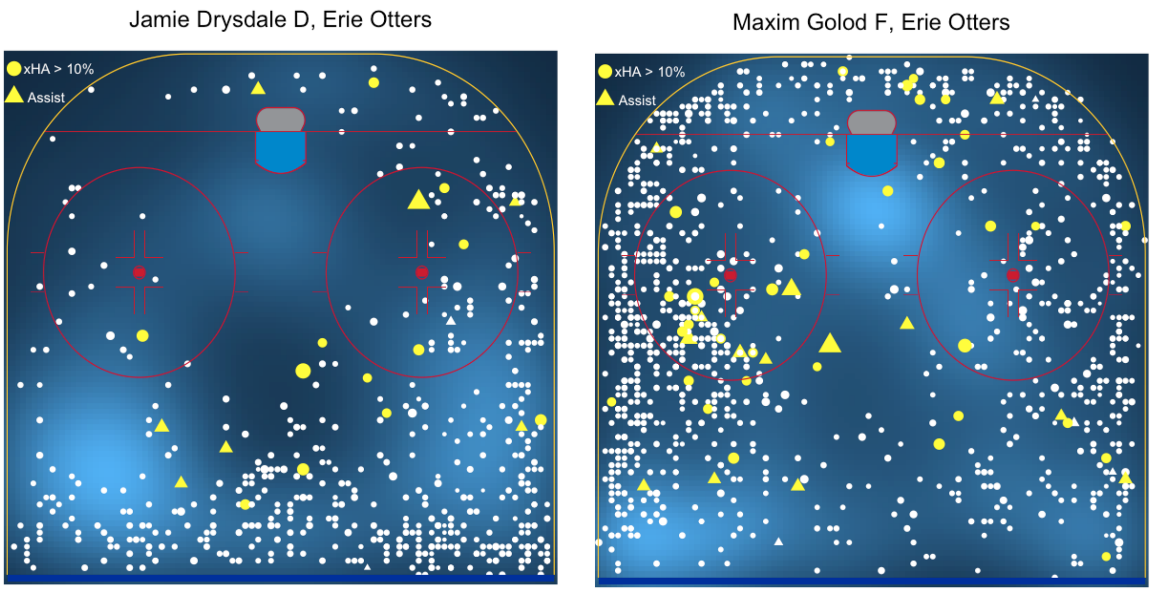
3.3 Pass Reward Metrics

Focusing on passes that actually resulted in a shot, we first introduce **Expected Assist** that is defined only for shots off of passes as the probability that the receiver scores. This improves upon Primary Assists, because it helps reduce the "luck" associated with getting assists from low probability shots. The figure below illustrates which players produced more or less assists than expected.



Its unsurprising to find that Jamies Drysdale and Maxim Golod are among the top in Primary Assists Per Game and Expected Assists Per Game. Interestingly, Jacob Golden has significantly more Expected Assists Per Game than Primary Assists, in-line with his strong pass risk numbers.

Next, we define **Expected Hypothetical Assist** as the hypothetical probability that the targeted player scores after receiving the pass. This gives us a measure of pass reward for every pass made. To illustrate, the Figures below show passes for Drysdale and Golod, respectively, in the offensive zone and includes indicators for high reward passes and assists with an underlying heat map of the pass target locations. The lighter spots are areas these players pass more often to while the darker are less often.



4 Pass Risk vs Reward

Lastly, we provide an aggregate of the Pass Risk and Reward metrics defined in sections 2 and 3 for the 2019-20 Erie Otters with the addition of **CP%** and **CPOE**. CP% is Completion Percentage and CPOE is Completed Passes Over Expected. Drysdale, the number 6 pick in the 2018 draft, was among the most accurate passers while also driving the most value by Expected Hypothetical Assist. Golden and Bemish are defensivemen that drove more value than their 0 primary assists off passes suggests.

			PASS RISK				PASS REWARD		
PLAYER	PASSES	GP	CP%	CPOE	PPM/60	DPP	A1/60	xA/60	xHA/60
Jamie Drysdale	1397	30	0.81	1.79	0.83	0.24	0.30	0.22	1.11
Maxim Golod	1437	40	0.76	2.14	0.77	0.26	0.45	0.41	1.10
Chad Yetman	1132	39	0.78	2.14	0.62	0.20	0.10	0.20	0.78
Hayden Fowler	891	34	0.64	-5.65	-1.48	0.28	0.12	0.20	0.71
Drew Hunter	862	28	0.77	1.32	0.41	0.18	0.07	0.09	0.71
Jacob Golden	1031	34	0.78	1.88	0.57	0.40	0.00	0.11	0.69
Luke Beamish	328	10	0.80	3.87	1.27	0.41	0.00	0.07	0.69
Austen Swankler	949	37	0.67	-2.75	-0.70	0.23	0.08	0.16	0.64
Kyen Sopa	336	15	0.76	1.62	0.36	0.15	0.27	0.21	0.62
Kurtis Henry	1135	40	0.74	0.91	0.26	0.33	0.03	0.04	0.57

DATA SOURCE: STATHLETES BIG DATA CUP 2019-2020 SEASON

5 Discussion

In this analysis, we showed a novel way to model the uncertainty around pass reward through an Expected Hypothetical Goal Model. Furthermore, we introduced metrics for evaluating passing ability as well as ways to visual these metrics and patterns. Lastly, we provided an aggregate for these metrics to identity the most skilled passers as well as the players that drove the most value from their passing in the 2019-2020 season.

There are a number of areas to improve upon in this analysis. Particularly, the notable exclusion of the location of the defending players on the ice. In addition to the tracking data, more data on the other teams in the OHL should provide more stable predictions, reliable models, and a better relative understanding of the values of these metrics for each player. For NHL evaluators, previous years data would prove valuable in order to quantify how well some of these passing metrics translate to the NHL. Furthermore, an analysis of passing development over time would be interesting in understanding player development in passing. Passing is also a two-way street. Catching a pass is a skill as well; thus, an exploration of the receivers of passes could provide valuable. Lastly, model improvements can be made, particularly in regards to simulating the unobserved covariates. We simply used the empirical distribution, but other imputations methods could prove to be more accurate.

6 Appendix

All code can be found at our github: https://github.com/mrichards284/BigDataCup

References

[1] "Not All Passes Are Created Equal:" Objectively Measuring the Risk and Reward of Passes in Soccer from Tracking Data http://mirkwood.cs.edinboro.edu/ bennett/class/csci480/fall2017/notes/p1605.pdf

[2] Expected Hypothetical Completion Probability. Deshpande and Evans. https://arxiv.org/abs/1910.12337

[3] A New Expected Goal Model for Predicting Goals in the NHL. Evolving Wild. https://rpubs.com/evolvingwild/

[4] An Updated Expected Passing Model. Matthias Kullowatz. https://www.americansocceranalysis.com/home/2 updated-expected-passing-model

[5] Expected Completed Pass. Marc Richards. https://hockey-graphs.com/2019/09/21/2019-ritsac-slides-and-video/

[6] Measuring Goals Added Through Passes and Skating in the NHL. Nathan de Lara. https://hockey-graphs.com/2019/09/21/2019-ritsac-slides-and-video/

[7] BART: Bayesian Additive Regression Trees. Chipman et al. https://arxiv.org/abs/0806.3286

[8] XGBOOST: A Scalable Tree Boosting System. Chen and Guestrin. https://arxiv.org/abs/1603.02754

[9] Beyond the Goal: Visualizing Expected Assists in Soccer. Benoit Pimpaud. https://medium.com/nightingale/b the-goal-visualizing-expected-assists-in-soccer-28df81dafe09