Quantifying offensive passing ability with Expected Primary Assists

Dan Morse

3/5/2021

Introduction

Points are the long-standing measure of talent in hockey. After all scoring goals is better than. . . not scoring goals. But goals and assists are far from perfect measures.

Recently, the hockey analytics community has instead focused on metrics like **expected** goals, derived primarily from shot location data. Expected goals (xG) have proven to be more predictive of future success than goals alone.

But what about assists?

At times, assists can be the focal point of a goal. This pass[1] from Jamie Drysdale is a good example of an important assist. Then there are other instances where an assist is counted despite not really being a huge factor in the goal, like this Connor McDavid goal[2] that credited Tyson Barrie with an assist. These two plays are counted the same for both Drysdale and Barrie, despite the fact that Barrie's "contribution" doesn't even make it into the highlight clip.

Thanks to the Stathletes' Big Data Cup, we've got a litany of passing details that I have attempted to compile into a new model to help distinguish between these two plays: Expected Primary Assists (xPA). Using the provided 40-game sample of Erie Otters data, we'll attempt to properly quantify just how impactful passes like the one from former Otter Jamie Drysdale linked above.

Method

Calculating the likelihood of a pass leading to a goal is essentially a two-part problem:

1. What is the probability that the pass is completed?
2. What is the probability that the resulting shot goes in the net?

Pass Completion Probability

For the first part, we'll construct an expected completion model similar to what you'll find in the NFL analytics community[3]. I utilized extreme gradient boosting from the R package xgboost for this model. The features selected were some basic game information like time remaining, strength state, and score differential, along with various classifications of some of the more common types of passes. The hyperparameters were identified using 5-fold cross validation before being input into the final model.

```
## [1] "CV test AUC: 0.6964"
```

---

[1] https://twitter.com/StevenEllisTHN/status/1158374638403149824?s=20

[2] https://www.youtube.com/watch?v=gcfIqClMstY

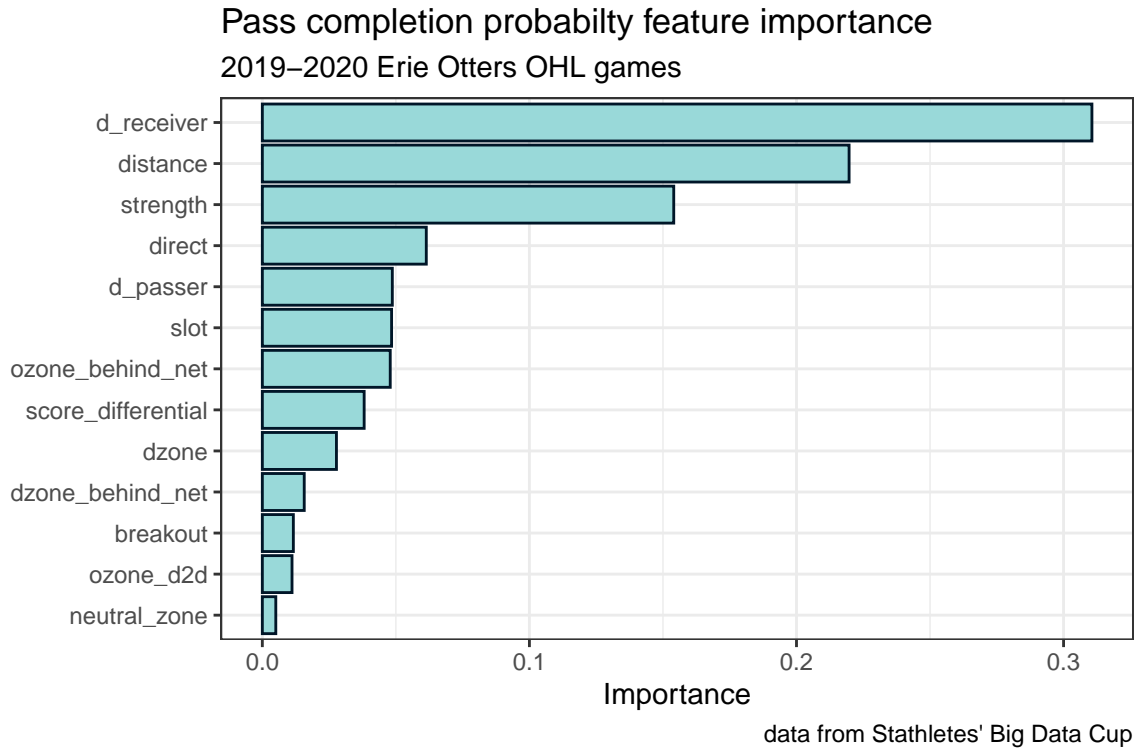[3] https://www.opensourcefootball.com/posts/2020-09-28-nflfastr-ep-wp-and-cp-models/

Figure 1: Passing model

Whether or not a pass targets a defenseman is the most important factor in predicting a pass completion. Generally, passes to defensemen are done at the blue line or while regrouping in your defensive zone, both high-completion probability situations. The pass distance and strength state are intuitively and empirically important. The most significant pass classification came with our slot pass, which makes sense as it's generally more difficult to find open space right in front of the net to get a pass through[4].

Expected Goals

Now that we've got a baseline to determine if a pass will be completed or not, we need to quantify how likely it is that a completed pass will lead directly to a goal. Enter: a new expected goals model.

Expected goals models in the NHL are numerous, but one thing they all omit is passing data. Josh and Luke of evolving-hockey.com opted to include information from events preceding a shot[5], which captures some of the pre-shot movement, in their model. But the raw NHL data doesn't include passing data. Fortunately for us, the Big Data Cup does.

Alex Novet has shown in the past[6] that including passing data can indeed improve an expected goals model's accuracy, so we're going to lean into that quite a bit for this one.

```
## [1] "CV test AUC: 0.8243"
```

The shooter's proximity to the net remains the most important feature, as is true with all other expected goals models in hockey that I can find. The angle to the net of the shot-taker is about as important as a

---

[4]According to the final model, passes across the slot averaged a 54.7% completion probability vs 73.6% for all other passes

[5]https://rpubs.com/evolvingwild/395136/

[6]https://hockey-graphs.com/2019/08/12/expected-goals-model-with-pre-shot-movement-part-1-the-model/

Expected goal feature importance

2019–20 Erie Otters OHL games
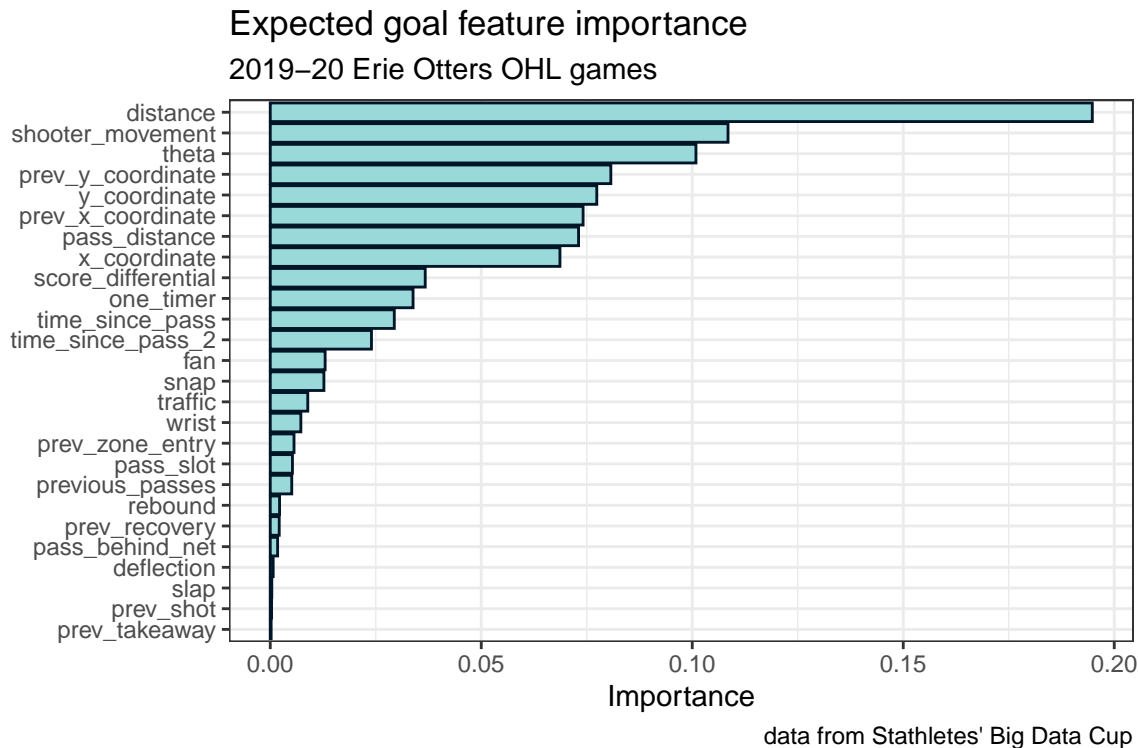
data from Stathletes' Big Data Cup

Figure 2: xG model

new feature called "shooter movement." Shooter movement only has a value when a pass was immediately preceding the shot. It is definied as the distance between the shot location and the intended pass destination from the previous event.

Pass distance (for shots with a pass immediately prior) and whether or not the shot was a one-timer also prove to be important factors. As they aren't publicly available in full right now, these aren't normally included in public xGoal models.

The AUC (area under the curve) evaluation of the cross-validation results is better than that of the Evolving Wild xG model (0.824 vs EW's 0.782)[7], which is encouraging in that it appears our new features are aiding the model's accuracy. But it's important to remember that this is only a 40-game sample and could be prone to over-fitting. Testing on future seasons would give us a better idea if this is the case.

We now have the two pieces we originally set out for when searching for xPA: pass completion probability and expected goals. We can combine to estimate the likelihood of a pass turning into a primary assist with the formula

$$xPA = P(CP) * xG_2$$

where $P(CP)$ is the probability of a completed pass and xG_2 is the expected goals of a shot from the receiver's location.

But what about passes to dangerous areas that are, for whatever reason, incomplete and don't have a shot recorded afterwards? Or passes that are completed but the recipient decides not to shoot?

We can account for these cases by calculating the expected goals value of *theoretical* shots taken after a pass. Our data just needs some small adjustments, and some educated guesses. We can adjust most of the variables fairly easily to account for the new phantom shot, but five others require some extra work:

---

[7]https://rpubs.com/evolvingwild/395136/

- shot type
- traffic
- one-timer
- shooter movement
- time since pass

Shot type, traffic, one-timer, and time since pass were all selected at random, weighed by how often they occured in our full dataset of actual shots taken. Shooter movement was sampled from all the shots within one standard deviation of the mean shooter movement in our shots data, and adjusted to ensure all shot locations remained within the bounds of the ice sheet.

The variables were randomly selected and the expected goals calculated 500 times, and with the average xG value being taken for each phantom shot.

Results

With that all taken care of, we can now calculate the expected goals and assists of every pass & shot in this data, and compare it to the goals, primary assists, and secondary assists scored on the OHL's official website[8]. I've narrowed it down to 5-on-5 stats as that's when the majority of the game is played, and other situations change the play style dramatically.

Erie Otters Advanced Stats
select 2019-20 games, min. 50 passes, 5-on-5 situations

	Position	Goals	xG	PA	xPA	PAOx	xPA/100	Passes	CPOx
Defense									
Jamie Drysdale	D	5	3.2	11	5.9	5.1	0.70	847	3.7
Drew Hunter	D	2	1.7	1	3.9	-2.9	0.64	605	2.7
Jacob Golden	D	1	1.7	3	3.4	-0.4	0.51	668	1.8
Kurtis Henry	D	2	1.8	3	3.9	-0.9	0.41	955	2.6
Jack Duff	D	2	1.9	5	3.7	1.3	0.38	977	-0.7
Luke Beamish	D	0	0.3	0	0.8	-0.8	0.35	228	5.2
Cameron Morton	D	2	1.1	4	1.9	2.1	0.34	555	-0.5
Brendan Kischnick	D	0	0.9	1	0.5	0.5	0.20	247	-0.9
Forwards									
Maxim Golod	LW	10	9.6	15	12.2	2.8	1.42	860	1.0
Hayden Fowler	RW	6	4.7	4	8.3	-4.3	1.27	651	-6.7
Kyen Sopa	RW	5	4.2	2	2.9	-0.9	1.24	234	3.8
Austen Swankler	C	8	8.0	6	8.5	-2.5	1.20	708	-3.8
Danial Singer	RW	6	6.1	5	2.0	3.0	1.19	168	2.4
Brendan Sellan	LW	6	8.5	6	7.3	-1.3	1.15	634	1.2
Daniel D'Amato	RW	10	9.9	6	7.4	-1.4	1.14	650	-0.7
Emmett Sproule	LW	6	7.7	4	7.6	-3.6	1.13	670	-1.4
Brendan Hoffmann	RW	8	9.1	4	5.4	-1.4	1.10	491	0.6
Elias Cohen	C	5	4.6	3	4.5	-1.5	0.97	462	4.5
Noah Sedore	LW	1	1.2	1	2.7	-1.7	0.93	291	0.0
Chad Yetman	C	18	14.3	8	5.8	2.2	0.86	674	6.1
Connor Lockhart	C	3	4.0	6	4.6	1.4	0.86	532	-0.9

It looks like we're measuring something close to primary assists! A linear model shows that xPA explains around 57% of the variance in an individual player's primary assist total at 5-on-5 and around 77% in all

[8]https://ontariohockeyleague.com/stats

situations.

One of the more noticeable differences is in how stark a contrast there is between forwards and defensemen. Taking each player's expected primary assists per 100 passes made (xPA/100), the best defenseman on the team is still a less dangerous passer than the worst forward on the team, on a per pass basis. That makes sense considering defensemen are making a significant portion of their passes on the perimeter of the offensive zone, where the expected goals are low, while forwards send more passes to the middle or down low.

On a contextual level, xPA/100 is measuring how "dangerous" a player's passes were on average.

Jamie Drysdale, a top-10 NHL draft selection, deservedly garnered most of the attention on the blue line last year, and that's backed up by his expected goals (xG) & expected primary assists (xPA) numbers. What the xPA can show us here that might otherwise go unnoticed is that Drew Hunter was nearly as dangerous of a passer as Drysdale despite only notching one primary assist at even strength. His high completion percentage over expectation (CPOx) indicates he was completing his passes at a high rate, implying that the lack of primary assists was driven by bad shooting luck. CPOx does appear to measure a skill rather than variance, as the first half-season CPOx for players correlated well with their second half-season CPOx.[9]

We can also use this as a stylistic measure. Plotting the xPA rate and the xG rate gives us a good idea of which players prefer to shoot from dangerous areas and which prefer to pass.
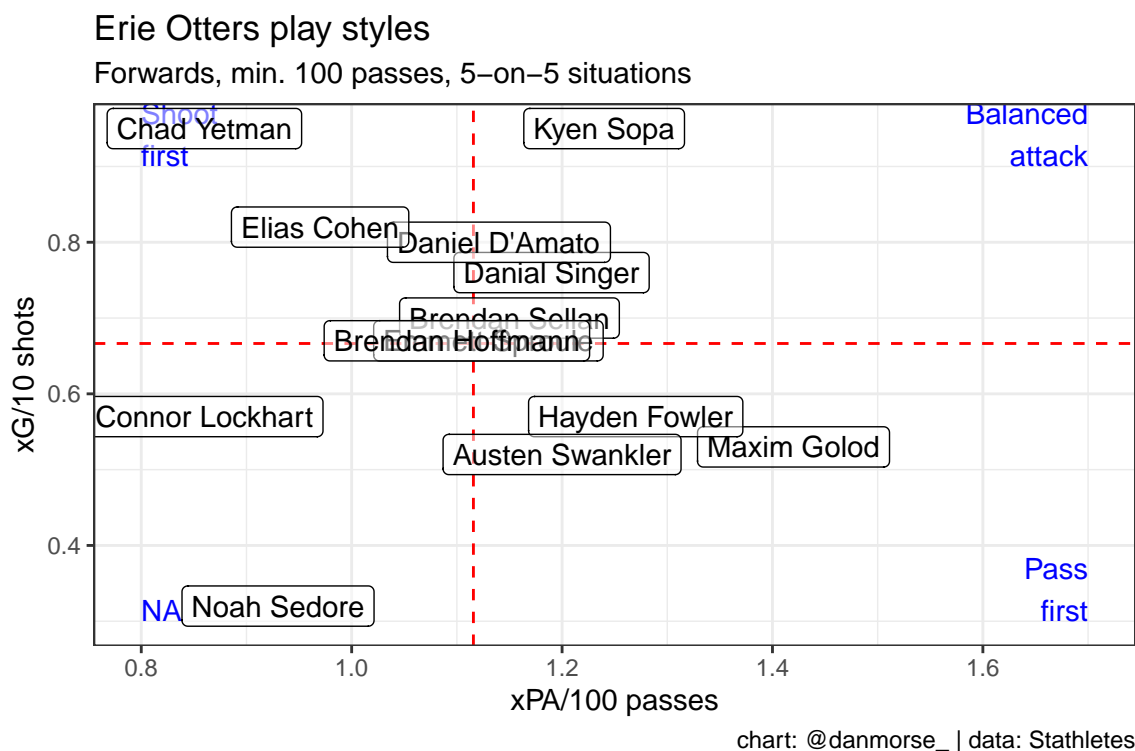

Figure 3: Player styles

The Otters' leading goal scorer (both in expected and observed goals) was Chad Yetman. Yetman was also the 2nd least likely forward to pass into dangerous areas on the team. Meanwhile, his frequent linemate Maxim Golod was on the opposite end of the spectrum, opting to pass more often than all but one other forward on the team. The pair playing on the same line produced the most goals on the team last year, with Golod netting the primary assist on nearly half of Yetman's goals at even strength.

```
## [1] "Yetman goals with Golod as A1: 44.4%"
```

---

[9]for players with at least 190 passes in both half-seasons: correlation coefficient = 0.5, r-squared = 0.476

Maxim Golod fed Chad Yetman last year
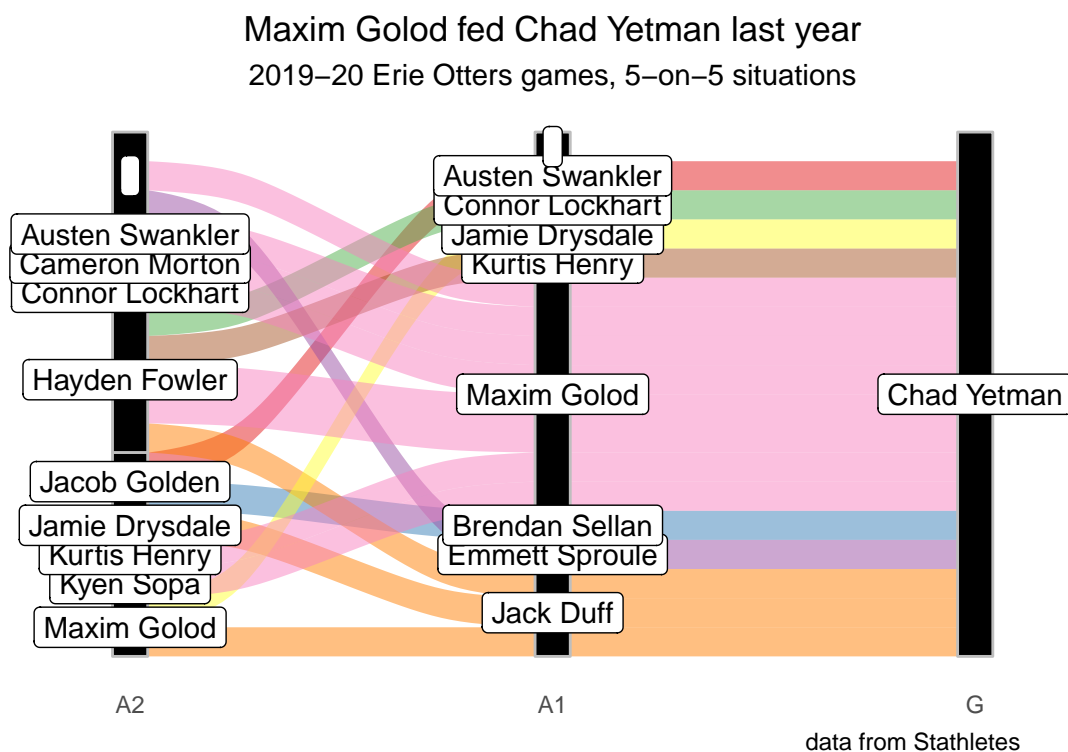2019–20 Erie Otters games, 5-on-5 situations

Figure 4: Chad Yetman's goals

Coaches frequently attempt to make forward lines with good chemistry. A breakdown like one in Fig. 3 could aid in that search, as we've already seen that pairing a player whose game is favored by xPA with a player who looks great by xG can lead to a winning combination. Based on this data, getting Hayden Fowler on a line with Connor Lockhart could be an idea worth exploring. Lockhart's xG rate isn't all that impressive, but his xPA rate makes it clear that if he's anywhere near the net he's opting for a shot over a pass. Letting him center a line with Fowler, the second-most aggressive passer on the team, should set up Lockhart in more favorable shooting situations.

Conclusions

Expected primary assists did help us find at least one underappreciated player in Drew Hunter. Testing this across an entire junior league has the potential to highlight more of these potential late-round prospects in the NHL draft. They can also help classify play styles for any player in the league fairly easily, which could be of use in identifying trade-deadline acquisition targets and in finding new forward line combinations.

Appendix

All the code for the models and figures can be found in a GitHub repository at https://github.com/danmorse314/Expected-Primary-Assists

6