Measurement of Passing in the Ontario Hockey League

Matthew Smith

March 2021

Introduction

Junior and professional hockey are faced with unique challenges brought on by the COVID-19 pandemic. These challenges provide decision makers with the opportunity to innovate and experiment with new tools and methodologies to answer hockey questions.

Inspired by Toumi et al. (2021), **completion probability over expectation (CPOE)** uses Bayesian Additive Regression Trees (Chipman et al. (2010)) to estimate the probability that a *direct* (or *tape-to-tape*) pass will be completed in the context of a number of inputs.

We apply CPOE to explore the following questions:

1. Who are the best tape-to-tape passers in the OHL?
2. Who are the best 2021 draft eligible tape-to-tape passers?

Data

Scouting data

Scouting data is read and cleaned to account for differences in player names between the scouting dataset and OHL player demographic dataset. We create a few new variables that help us identify player sequences, whether a team has a man advantage, whether a team has a lead, and if the pass was successfully completed.

OHL Demographics

Understanding performance for those eligible for the 2021 NHL draft will require additional information. We created a loop to get OHL demographics data, which will support model development.

First Name	Last Name	BirthYear	Position	Age	Ht	Wt	Shot	Rookie
Martin	Has	2001	D	19	6.04	187	R	
Nick	Malik	2002	G	18	6.02	174	R	
Christian	Stevens	2002	C	18	5.11	181	R	1
Jan	Mysak	2002	C	18	6.00	180	L	1

Cleaning

We reasoned that direct passes are a better representation of passing skill set as it clearly shows intent to move the puck to a specific player. Goalies are excluded from the data set that is fed into the model.

A manual review of a subset of player and team names within the data are reconciled to ensure the joins occur correctly.

Variables removed from the scouting data include game_date, Home.Team, Away.Team, Period, Clock, Home.Team.Skaters, Away.Team.Skaters, Home.Team.Goals,

Away.Team.Goals, Team, Detail.1, Detail.2, Detail.3, Detail.4, Event, Player, and Player.2. Variables removed from the OHL data include Shot, Date of Birth, Height and Weight. Only completed observations are included in the final data (i.e., player exists in both the scouting and OHL data, along with date of birth and shot). Age is calculated as the difference between the the date of birth listed in the OHL data and the current date of the data. Distance between the player and the opposing team's net is calculated as the position between Player.2 and the center of the opposing team's net.

The final data provides us with 21,834 direct passes.

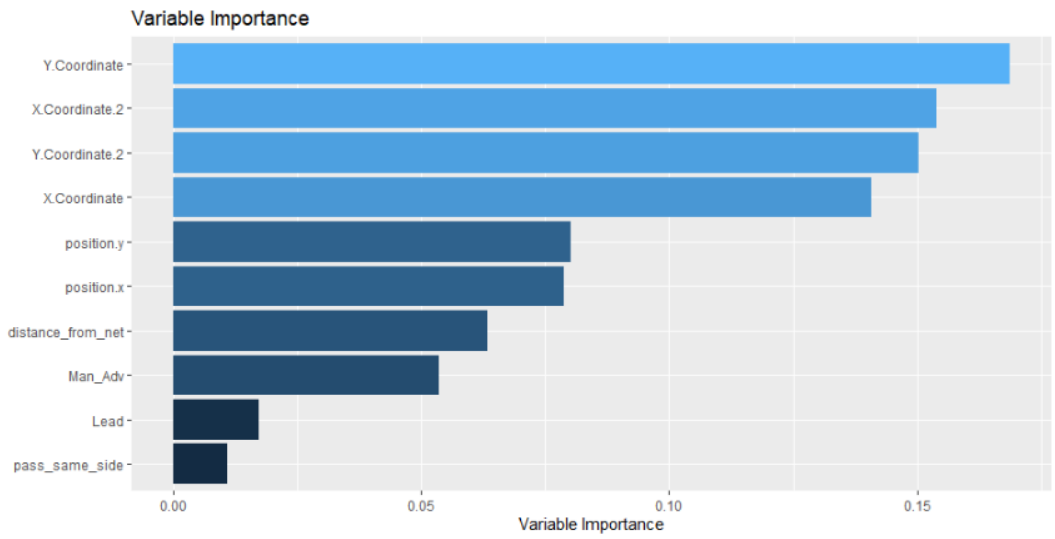
Methodology

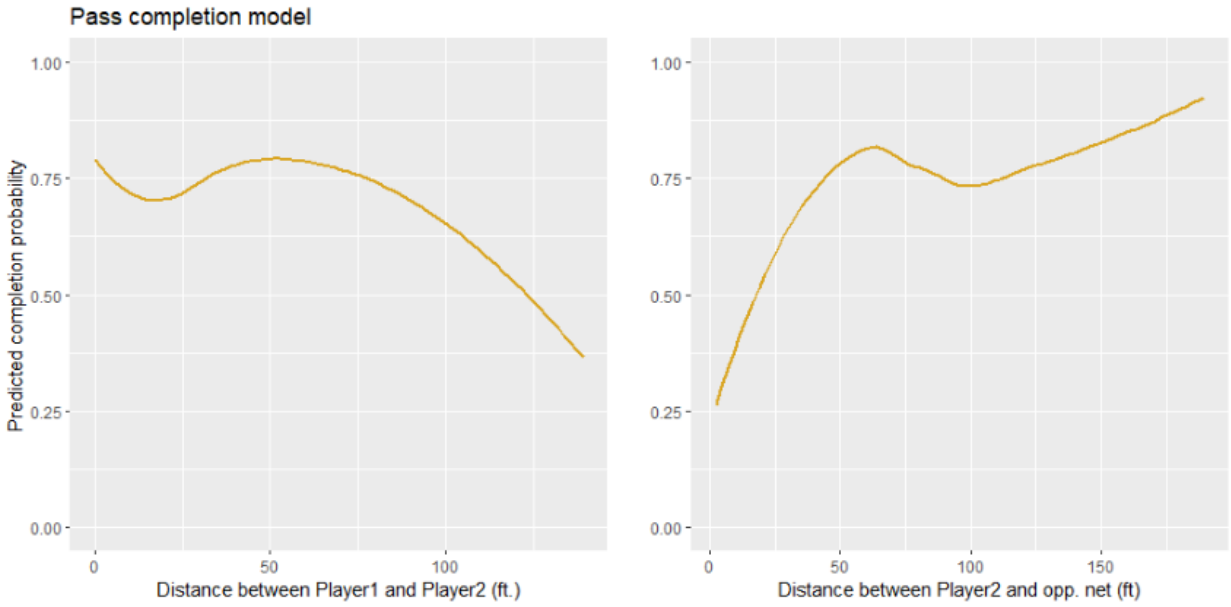
Bayesian Additive Regression Trees (BART) is a sum-of-trees model that we can use to predict the distribution of passing success in the context of a number of inputs. BART has been noted to perform better than similar ensemble and boosting models even in the presence of irrelevant inputs.

The inputs of the model are:

- Passer coordinates
- Receiver coordinates
- Distance between passer and receiver in feet
- Distance between receiver and the opposing net in feet
- Passing team score differential
- Passing team man advantage
- Passer position
- Receiver position
- Passer and receiver have the same shot (L/R)

Below are visuals showing the importance of each input in the model, the expected completion probability by distance between players, and the expected completion probability by the distance between the player taking the pass and the opposing team's net.





Results

Completion probability over expectation equal to 0% suggests that a player is completing passes we expect them to complete. If a player completes 50% of passes that have a predicted probability of 50%, the player will have a 0% CPOE.

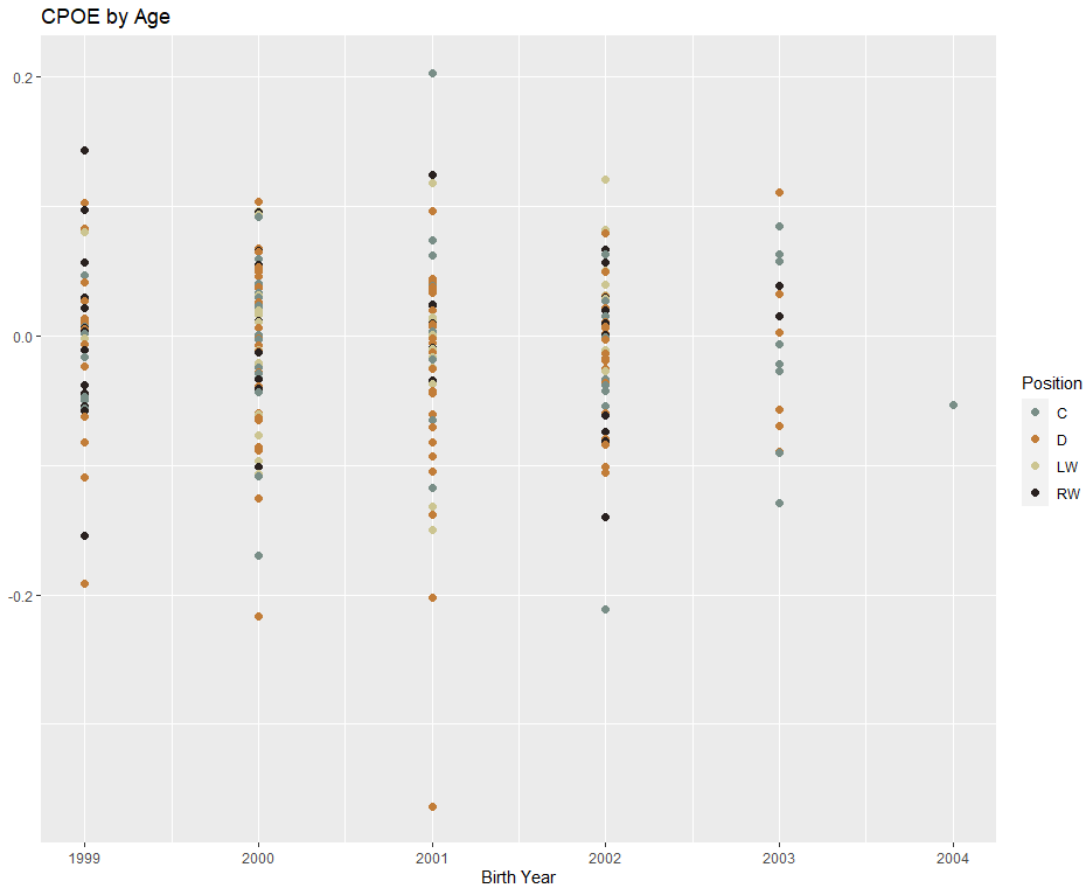
1. Who are the best tape-to-tape passers in the OHL?

In 2019, Marco Rossi was the best tape-to-tape passer in the OHL. An actual completion percentage of 93.33% and expected completion percentage of 72.99% suggests that he completes passes regardless of difficulty.

In other words, Marco Rossi completed 20.34% of passes over expectation.

Team	Position	Player	BirthYear	Actual	Expected	CPOE
Ottawa 67's	C	Marco Rossi	2001	93%	73%	20%
Ottawa 67's	RW	Joseph Garreffa	1999	74%	60%	14%
Kitchener Rangers	RW	Ryan Stepien	2001	86%	74%	12%
Saginaw Spirit	LW	Cole Perfetti	2002	85%	73%	12%
Ottawa 67's	LW	Jack Quinn	2001	82%	70%	12%
Barrie Colts	D	Brandt Clarke	2003	92%	81%	11%
London Knights	D	Alec Regula	2000	91%	81%	10%
Owen Sound Attack	D	Brady Lyle	1999	90%	80%	10%
Saginaw Spirit	RW	Cole Coskey	1999	82%	72%	10%
Saginaw Spirit	D	Mason Millman	2001	86%	77%	10%

The visual below shows CPOE across position and birth year for those with at least 20 direct passes.



2. Who are the best 2021 draft eligible tape-to-tape passers?

Brandt Clarke is the best 2021 draft eligible passer in the OHL. The NHL has Clarke listed as the 3rd best prospect in the 2021 draft.

Team	Position	Player	BirthYear	Actual	Expected	CPOE
Barrie Colts	D	Brandt Clarke	2003	92%	81%	11%
Owen Sound Attack	C	Deni Goure	2003	76%	67%	8%
Sudbury Wolves	C	Landon McCallum	2003	75%	69%	6%
Guelph Storm	C	Danny Zhilkin	2003	80%	74%	6%
Sudbury Wolves	RW	Chase Stillman	2003	77%	73%	4%
Sarnia Sting	D	Ryan Mast	2003	76%	73%	3%
Guelph Storm	D	Daniil Chayka	2002	90%	88%	2%
Sarnia Sting	RW	Ty Voit	2003	72%	70%	2%

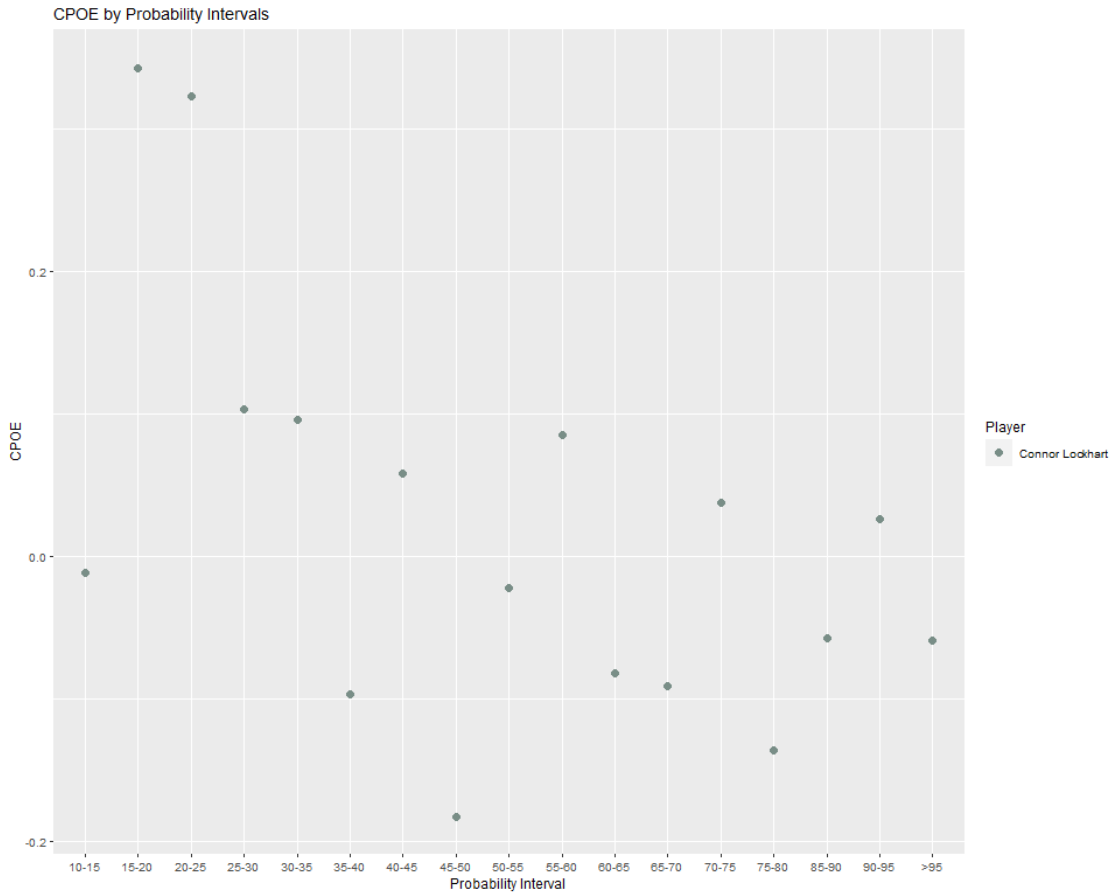
Windsor Spitfires	RW	Matthew Maggio	2002	59%	58%	1%
Owen Sound Attack	D	Sam Sedley	2003	81%	81%	0%

We are defining draft eligible as players born in 2003 that have at least 20 direct passes in the scouting data.

One players not listed in the top CPOE scores is center Connor Lockhart, who Lines.com ranks 63rd best prospects in the upcoming draft.

Team	Position	Player	BirthYear	Actual	Expected	CPOE
Erie Otters	C	Connor Lockhart	2003	68%	70%	-2%

Partioning the data by probability intervals (e.g., 0-5%, 5-10%,…,90-95%, 95-100% completion probability) might help us better understand performance.



Probability interval CPOE requires at least 5 passes to occur within the respective interval.

Connor performs exceptionally well in the challenging situations. His 34.30% CPOE in the 15-20% interval and 32.34% CPOE in the 20-25% CPOE interval are the highest CPOE scores among draft eligible players. In fact, his CPOE in these probability intervals are the highest CPOEs of any interval in the entire league. However, he struggles to complete

passes that we'd expect him to complete >95% of the time. In practice, a video coach can apply CPOE to measure improvement over time.

Application

CPOE provides decision makers with the ability to quantify passing ability. Coaches will be able to quickly highlight opportunities for player improvement, and how they may want to utilize specific players in set plays.

Modeling several years of data can help us further our understanding of player development as it relates to passing. More specifically, historical modeling can help NHL decision makers identify how CPOE in major junior translates into NHL. If historical data suggests CPOE in major junior is an indicator of CPOE in the NHL, GMs will be able to optimize draft capital.

Lastly, BART can be applied to NHL player tracking data, allowing us to better understanding passing in the NHL.

Contact

Github
Twitter