NHL Project Progress Report

3/29/2021

Client: Sam Ventura

Minyue Fan, Steve Kim, Kexiong Shen, Linda Yang

Advisor: Brian MacDonald

Carnegie Mellon University

Agenda

Topic	Slides No.
Introduction	3
Data	4-6
Methods	7
Results	8
Next Steps & Roadblocks	9
Q&A	10

Introduction

- Recap of our project goal
    - Investigate and determine the effect of development path on player's development and performance in the NHL
    - Whether the different development path taken by players of similar traits /qualities would make a difference

Data Overview

- Web scraping in R:
  - Source: https://www.eliteprospects.com/
  - All players who played in NHL, NCAA, USHL, or AHL from 2001 to 2020
    - *contains some data earlier than 2001*
- Two data sets:
  - Players' biographical information
  - Players' performance data each season
    - *box score statistics*
- More data if needed:
  - Further back in time?
  - More player performance measurements (e.g. playoff appearance)

Data Description

- Biographical information:
  - 15786 rows (# of players); 7 columns

	Player	Position	DateofBirth	Height	Weight	Nation	Shoots
1	Scott May	C	Jan 08, 1982	5'10" / 178 cm	187 lbs / 85 kg	Canada	R
2	Kent Gillings	F	Jun 14, 1979	5'10" / 177 cm	194 lbs / 88 kg	Canada / Ireland	R
3	Tyler Kindle	D	Feb 20, 1978	5'8" / 173 cm	165 lbs / 75 kg	USA	L
4	D'Arcy McConvey	C	Oct 23, 1981	5'10" / 177 cm	185 lbs / 84 kg	Canada	L
5	Lloyd Marks	C	Oct 21, 1977	5'8" / 173 cm	174 lbs / 79 kg	Canada	L
6	Jason Deskins	C	May 06, 1979	5'10" / 178 cm	185 lbs / 84 kg	USA	-
7	Jim Abbott	LW/C	May 03, 1980	6'1" / 186 cm	185 lbs / 84 kg	USA	L

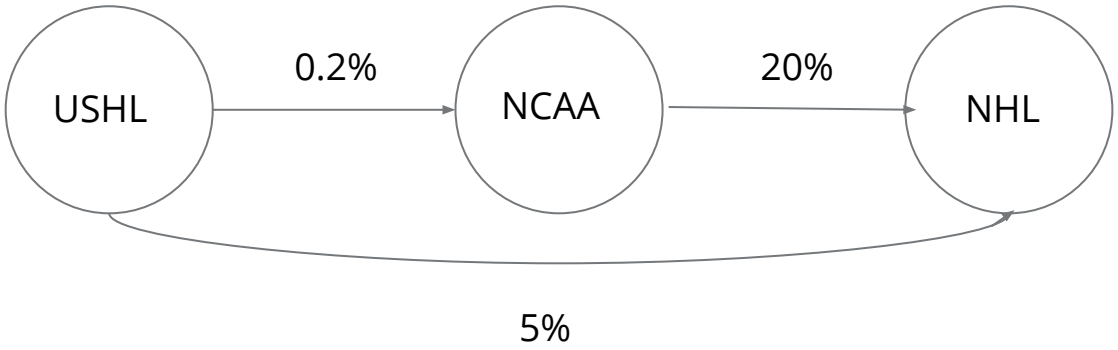
Data Description

- Player performance:
  - 266326 rows (# of players * # of seasons they played); 10 columns

	Player	Season	Team	League	Games	Goals	Assists	TotalPoints	PenaltyMinutes	PlusMinus
1	Scott May	1998-99	South Surrey Eagles	BCHL	45	10	28	38	23	
2	Scott May	1999-00	South Surrey Eagles	BCHL	54	42	42	84	86	
3	Scott May	2000-01	Ohio State Univ.	NCAA	37	9	9	18	26	-3
4	Scott May	2001-02	Ohio State Univ.	NCAA	40	12	17	29	42	4
5	Scott May	2002-03	Ohio State Univ.	NCAA	43	10	25	35	56	5
6	Scott May	2003-04	Ohio State Univ.	NCAA	41	15	19	34	42	4
7	Scott May		St. John's Maple Leafs	AHL	5	1	1	2	2	3
8	Scott May	2004-05	St. John's Maple Leafs	AHL	16	0	1	1	21	-3

Methods

1. Logistic Regression
2. Bayesian Additive Regression Trees (BART)
3. Markov Chain Model



Results

Logistic Model

- Outcome variable: If a player played in 10 or more games after being drafted into the NHL
- Predictor variables:
  - Player success metric standardized by game
    - *Plus minus per game during the year before being drafted into the NHL*
    - *Goals per game during the year before being drafted into the NHL*
    - *Penalty minus per game during the year before being drafted into the NHL*
  - Position: forward vs backward
  - League before being drafted into the NHL
- Results: Unfortunately, we found no significant predictors
- Next steps: There were a few players in the data set that transitioned in and out between NHL and other Pre-developmental leagues. We need to further clean the data and address these discrepancies to Improve the accuracy of our findings

Next Steps & Roadblocks

- Setting up a remote joint CMU account (contacting Carl Skipper)
  - Web-scraping takes much longer than anticipated
  - Need not only to scrape by league, but also by individual player
- Modelling, using BART and Markov Chains
  - Attempt to model how the different development paths for players with similar backgrounds / stats
  - Attempt to isolate the development path specifically
- Expand to increase more leagues
- Dealing with missing data
  - Missing information on some crucial stats, such as plus-minus

Q&A

Thank You!