

Points Prediction Models for NHL teams

Writer: Eric Foote

September 16, 2018

Supervisor: Dr. Stewart

1 Motivation

"Hockey statistics is in its infancy" [1] from what started out in the earliest day as a Yahoo group [2] has grown to a few teams having an analytics department that influences everything from personnel choices to player scouting. (e.g. Toronto) People want to forecast various aspects of the game for a varying number of reasons from who will score the most goals, which teams will have the most wins in the 82 game regular season to even who will win in the playoffs. The work that I am proposing is using various regression models to determine the points of all 30 NHL teams. For reference the points that I am referring to are in the case of a win a team is given 2 points and a single point for an overtime or shootout loss. The data is going to consist of 9 years from 2008-2009 to 2016-2017 for the usable data and the previous season 2017-2018 for testing. The different techniques that I want to apply to this data set are sum of squares, logistic regression, gradient boosted decision trees and if time permits a look at some different regression trees.

2 Work done

Currently a majority of time has been dedicated to the data collection process this is due to unfamiliarity with the JSON (Javascript object notation) data format, its structure when imported into R Studio and the lack of documentation on the NHL's part. However, the 10 years of player data and team records has been successfully collected as of right now. Also, research has been done into the various techniques that people are using to create current models using this data.

3 Work to be done

I am going to be starting out doing a sum of squares regression model on the dataset and determining how successful that particular model, then move on to do the logistic regression model and then finally move on to the gradient boosted decision tree. A rough time line is stated below to give a look in to roughly how long work is going to be spent on each individual topic, overall analysis and report writing.

Item	Rough Week Length
Sum of Squares	1
Logistic Regression	2-3
Gradient Boosted Tree	2-3
Other Regression Trees	1-2?
Analysis	1.5
Report Writing	1.5

References

- [1] Joshua Weissbock Forecasting Success in the National Hockey League using In-Game Statistics and Textual Data School of Electrical Engineering and Computer Science Faculty of Engineering University of Ottawa, Ottawa, Canada, 2014
- [2] Robert Vollman Aug 24, 2016
<http://www.hockeyabstract.com/thoughts/abriefhistoryofhockeyanalyticsconferences>