Authors:

Oliver Schulte, Zeyu Zhao; School of Computing Science, Simon Fraser University, Vancouver, Canada

Mehrsan Javan, Philippe Desaulniers, Christopher Boucher; SPORTLOGiQ, Montreal, Canada

Title: Model-Based Clustering for Identifying Exceptional Players in the NHL Draft

## Introduction

* Two Goals
  + Build predictive model
  + identify cohorts (cite Weissbock, Sony)
  + Unification: identify cohorts s.t. we can build good predictive models.
* More accurate predictions, non-linearity
* Predictively valid cohorts.
* Problem: Cohorts should be interpretable. Interpretability even more important than accuracy.
* Solution: Equation tree (LMT). What’s the terminology for these tree models.
* Get good results in terms of tree, Spearman.

## Methods

Explain LMT. Show model tree.

[Figure 1. Model tree.] [Figure 2: visualize of 2 clusters in terms of TOI, GP. Maybe pick top 3 players in cluster]

(have to write this)

|  |  |  |
| --- | --- | --- |
| Draft Year | LMT classification accuracy | LMT correlation |
|  |  |  |
|  |  |  |

Mention Shuckers, draft order

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Cluster Definition | Cluster equation | Cluster Size | Top Player | Strongest Point | Weakest Point |
| Show path | Show equation (maybe top three players) |  |  | (e.g. age vs. average age) |  |
|  |  |  |  |  |  |

# Cutting Floor

## How to trade a player: Clustering and Ranking NHL Players

# Sloan Instructions

Abstracts must contain fewer than 500 words, including title and body.

* Abstracts may include up to two tables or figures combined (e.g. 1 figure and 1 table, or 2 tables)
* Each abstract should contain the following sections:
  + Introduction – What question is this research trying to answer? Why is it an important question for the industry?
  + Methods – Description of relevant statistical methods used, including data sources or data collection procedures
  + Results – Description of actual (not promised) results along with relevant statistics
  + Conclusion – The overall takeaway from the study, including how the results will impact the sports industry
* All abstracts must be submitted in one PDF through the [2017 Abstract Submission online submission page](http://www.sloansportsconference.com/activities/research-papers/research-paper-abstract-submission/)

Notes: Sep 18 form OS. Great comments. We’re working on the tables. Should think of a name for the impact. Candidates:

1. Goal Impact

2. Expected Goals Added

3. Goal Probability Added

Each of these is similar to what has been used before.

Average impact per game:

|  |  |
| --- | --- |
| ClusterId | Avg(Impact\_per\_Game) |
| 1 | 2.431 |
| 2 | 2.521 |
| 3 | 0.002 |
| 4 | 0.284 |
| 5 | 2.204 |
| 6 | 2.851 |
| 7 | 2.626 |
| 8 | 2.678 |
| 9 | 0.290 |
| 10 | 2.729 |
| 11 | 3.181 |