Delving into Corsi: Building predictors for shot attempts in hockey

Patrick Rooney

November 6, 2017

A player's Corsi-for percentage in hockey is a measure of how many shots his team attempts while he is on the ice:

$$CF\%(X) = \frac{\text{shot attempts for X's team while X on-ice}}{\text{shot attempts for both team while X on-ice}}$$

Note: Shot-attempt = shot on goal OR shot missed / blocked

Questions:

- Can we predict shot-attempts using the $CF\%$ statistic?
- Can we build a *better* predictor, using all shot attempt data?
- Most importantly: is the $CF\%$ a good measure of how a player influences shot attempts?

Our Data Set

Distribution of Shot-attempts

We want to build a shot-attempt predictor, based on data from the 2015-2016 NHL season. There were 136,500 attempted shots in this season, and a total of 900 players (excluding goalies). We built three different algorithms, using three different

1. A logistic regressor using only six features: the cumulative $CF\%$ of home and away players, the average salaries, and the average playing time.

2. A logistic regressor using individual 1,800 binary features. Each player has two features: whether he is on-ice at home, or away.

3. A random forest algorithm using the 1above,800 features.

Random Forest Predictor