How bad should I win next faceoff?

Big Data Cup 2021

Philippe Blouin-Leclerc, Stéphane Caron, Jean-Philippe Le Cavalier et Samuel Perreault

05 March, 2021

During the course of a hockey game, there are a lot of faceoffs. The outcome of these events usually dictates which team will control the puck at the beginning of a sequence. Thereby, we can say that every faceoff are important. However, many would agree to say that some of these faceoffs are key moments, and thus have even more importance. As an example, a faceoff in the offensive zone, trailing by one goal with less than a minute to play would appears to be quite important. In this analysis, we will try to quantify the importance related to winning faceoffs. In other words, we want to confirm if winning a faceoff in the offensive zone is increasing the chances to score a goal during that sequence. On top of that, we would like to specifically answer this core question

> To what extend is winning offensive zone faceoffs increase the chances to score a goal? What are the effects on the seconds following the faceoff? Is winning an offensive zone faceoff even more important in man advantage situations?

We targeted offensive zone faceoffs, because their impact on scoring a goal seemed more obvious as the team is already close to the opponent net. To answer that question, we decided to use the Erie Otters data made available in the scounting dataset. Our rationale for using this dataset was to use data of one single team, which from our perspective, simplifies the interpretation and the conclusions. Moreover, the `scounting` dataset is the one having the most observations for a single team.

In the next sections, we are going to try confirm the impact of winning offensive zone faceoffs on a sequence-by-sequence basis. Afterward, we are going to dive deeper in the effect of winning such faceoffs, by looking at seconds-per-seconds or man advantage situations. In conclusion, we will apply our approach to more data using NHL API (add link).

Sequence-by-sequence

As in all data science projects, we first need to understand the data we are working with. As mentionned in the introduction, we focused on the `scouting` dataset. For simplicity, we will use some abbreviations for few repeated terms in this analysis:

- FO: Faceoff
- OZ: Offensive Zone
- DF: Defensive Zone
- NZ: Neutral Zone
- GF: Goal For Erie Otters
- GA: Goal Against Erie Otters
- Erie: Erie Otters

In the table 1, we can see basic informations about the `scouting` dataset that appear relevant to our analysis.

In order to tackle our initial problem, we had to structure the data in a way that we can easily analyze each sequences. After doing so, we can now extract additionnal informations about our dataset (see table 2 and add context such as the position of the faceoff and the)

Table 1: High-level features of the scouting dataset

Min. game date	Max. game date	Games	FO	Erie won FO	GF	GA
2019-09-20	2020-03-08	40	2441	1103	148	145

Table 2: Contextual data for faceoff situations

FO won Erie	PK			Even			PP		
	FO	GF | GA	% success	FO	GF | GA	% success	FO	GF | GA	% success
Defense									
FALSE	129	12	9.3 %	283	19	6.7 %	7	0	0 %
TRUE	71	8	11.3 %	292	19	6.5 %	10	1	10 %
Offense									
FALSE	10	0	0 %	390	24	6.2 %	93	10	10.8 %
TRUE	4	0	0 %	245	13	5.3 %	110	16	14.5 %

From the table 2, we can see that we don't have a lot of goals scored by Erie Otters for sequences that started in the offensive zone (63 goals) or goals against Erie for sequences that started in the defensive zone (59 goals). Still, that might be sufficient for making conclusions on per-sequence basis. However, we strongly doubt that it will be sufficient for drawing conclusions on a more granular basis.

Another thing we can notice from 2 is the success rate differences between won and lost faceoffs. From the data we have, we can already see the effect of winning the faceoff on man advantage situations. For even strength situations, the impacy is not that clear ...

To see if the effect of winning the faceoff in the offensive zone is significant, we fitted a logistic regression. From the parametric anova table, we can conclude that (at least for Erie Otters) the effect is significant:

```
## Anova for Parametric Effects
##              Df Sum Sq Mean Sq F value    Pr(>F)
## faceoff_win   2 372.08 186.038   185.6 < 2.2e-16 ***
## Residuals   850 852.00   1.002
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Our approach

Length: 2-3 pages

Ici, on explique comment on va tenter de pousser plus loin. On voudrait savoir plus que le simple fait que "ca aide" ... à quel point ça aide?

Brève explication de notre appproche (GAM).

Résultats du modèle illustrer dans le graph, avec interprétation.

More results

Length: 1 pages

On refait notre analyse mais sur le data de la NHL.

Potentiellement expliquer quelques hypothèses supplémentaires (ex: prend le data de toutes les équipes car environ le même weigth, expliquer différence dans le data si le cas)