How important is winning the faceoff?

Stéphane Caron[*]   Jean-Philippe Le Cavalier[†]   Samuel Perreault[‡]

Introduction

Length: 0.5 pages

Ici, on fait la mise en contexte et l'explication du problème.

Explications des prochaines sections.

Exploratory data analysis

Length: 1.5-2.5 pages

Ici, on présente le jeu de données sur lequel on travaille:

- informations de bases (nombre de matchs, nombre de faceoff, nombre de buts, etc)
- visualisations nous permettant de comprendre un peu mieux l'effet du faceoff sur d'autres aspects (tirs dans les x secondes suivant le fo, l'emplacement des buts marqués, etc)

Finalement, on présente quelques unes de nos hypothèses avec explications

- regarde seulement le data Erie
- faceoff en zone offensive

Pour conclure, on fait l'analyse (regression logistique) de l'effet du faceoff sur le fait de marquer un but a la fin d'une séquence.

On devrait être en mesure à ce point ci de confirmer si gagner le faceoff aide de manière significative à marquer un but. Mais, on veut en savoir plus . . .

Our approach

Length: 2-3 pages

So far, we have gathered plenty of evidence that winning offensive faceoffs significantly increases your chances of scoring a goal. Our intuition also tells us that this effect is most prominent in the very first seconds following the faceoff win and that, over time, the fact that one has won or lost the (most recent) faceoff looses its importance. To verify that, let us take a closer look at the exact times the goals in questions were scored.

(BARCHART OFF and DEF)

As shown in Figure~(BARCHART), and somewhat unsurprisingly, winning the faceoff tends to lead to more often to a quick goal. In particular, 4 goals were scored during the 4th second, that is, the timespan [3,4). **Comment y-axis on the right (giving probabilities). Interpret appropriately. Note presence of zero probabilities and explain why (due to the small number of data points).** This makes it hard to communicate the information about the content of barcharts like those of Figure~(BARCHART). To overcome such difficulty, let construct a smooth version of these latter using loess, a method for smoothing

---
[*]Intact(?) — Corresponding author: EMAIL
[†]Promu(?)
[‡]Department of Statistical Sciences, University of Toronto, Toronto, Canada.

scatterplots (*locally estimated scatterplot smoothing*). **Informal model description: we want curves, an estimate of the probability of a goal occurring during a given time interval (of one second). One curve per situation (like in the bar chart).** See Section~(SEC) for a more formal description of the underlying model. **Highlight that one needs to fine tune a smoothing parameter and explain how that was done.** The results are displayed in Figure~(LOESS).

(LOESS chart with various bandwidth, OFF and DEF)

1. discuss results in depth — GIVEN THAT THE SEQUENCE WAS LONGER THAN t.

2. SKIP PARAGRAPH. discuss weakness of loess: no confidence intervals for validating like we did previously.

3. discuss that not much data here, so possibly worthless anyways, (although CI from splines are conclusive).

4. discuss interesting questions that we cannot answer, need more data to use a more granular definition of situation.

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)

Technical details I.   Denote $Y_t \in \{0, 1\}$ the random variable indicating whether a goal occurred in the interval of time $[t, t+1)$, and let $\mathbb{E}(Y_t|x)$ be the expectation of Y_t (corresponding to the probability of a goal occurring in that timespan) given that the most recent faceoff was lost ($x = 0$) or won ($x = 1$). The loess curves in Figure~(LOESS) were obtained by fitting the *generalized additive model* (gam)

$$g\{\mathbb{E}(Y_t|x)\} = (1-x)f_0(t) + xf_1(t), \qquad g(z) = \ln\{z/(1-z)\}, \tag{1}$$

where g is the so-called logit function and, for both $k = 0, 1$, f_k is an unknown (i.e., to be estimated) nonlinear function of t that approximates $g(Y_t|x = k)$. The probabilities reported are obtained by solving this equation for $\mathbb{E}(Y_t|x)$ at each value of $t \in \{0, 1, 2, \dots\}$. The loess of sec.xx and the spline regression of sec.yy were fitted using the gam package, which fits generalized additive models.

Technical details II.   In Section~(SEC), we reproduce the analysis involving the model in (1), this time fitting the model using a splines-based technique (as opposed to a loess-based one). Loosely speaking, the two fitting procedures differ in the way the functions f_k are approximated. The splines-based method allows us to construct confidence intervals as well.

To perform the more granular analysis of Section~(SEC), we fit a model that includes a function $2 \times ZZZ = ZZZ$ nonlinear functions of t. We use the subscripts $k \in \{0, 1\}$ to indicate whether the faceoff was lost or won, and $\ell \in \{1, \dots, ZZZ\}$ to refer to each of the **ZZZ** situations specified in Section~(SEC). The resulting model is given by

$$g\{\mathbb{E}(Y_t|x, s)\} = \sum_{\ell=1}^{Z} \mathbb{1}(s = \ell) \times \left\{(1-x)f_{0\ell}(t) + xf_{1\ell}(t)\right\}, \tag{2}$$

where g is as in (1). Note that (2) could actually be expressed and fit as ZZZ distinct models. However, this formulation makes it clear how to include further covariates that are known to have a similar effect in multiple situations.