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 In the most basic case treated in this report, the underlying model takes the form

$$g(Y_t|x) \approx (1-x)f_0(t) + xf_1(t),$$

where $Y_t \in \{0,1\}$ indicates whether a goal occurred in the interval of time $[t, t+1)$, $x \in \{0,1\}$ indicates whether the faceoff was won or not, g is the so-called logit function and f_k is an unknown nonlinear function of t that estimates $g(Y_t|x=k)$. The different methods presented in this report are distinct in the way the functions f_k are estimated. The probabilities reported are obtained by solving this equation for Y_t. The loess of sec.xx and the spline regression of sec.yy were fitted using the gam package, which fits generalized additive models.