BSAVE: Bayesian Skater Action Value Expectation

Alex Novet

1. Introduction

The Stathletes Big Data Cup datasets provide previously unavailable junior and women's hockey event data. With this data, we can ask what was the value of each action that a player took with the puck.

This action valuation is a topic that has been studied across multiple sports and with multiple methodologies. In particular, VAEP [1], xT [2], and count-based approaches have been prominent [3]. For this dataset, I suspect that the most promising route is an approach similar to VAEP. It uses a richer feature set than the puck's location and relies on a window of events rather than a chain of possession. This lets us appropriately credit situations where possession is strategically conceded, such as dump-ins before a planned forecheck.

Separately, there have been several recent works in sports analytics that have taken a Bayesian approach to modeling. In particular, Bayesian Additive Regression Trees (BART) [4] have been successfully used in both football [5] and hockey [6]. BART has demonstrated strong predictive ability and is generally adept at handling nonparametric relationships and unbalanced datasets, both of which we face predicting goals here. BART creates posterior distributions rather than point estimates in predictions, which offers a more detailed understanding of possible outcomes. And it can be used to investigate causal inference questions of interest [7].

Thus, in this paper I implement a BART model in a VAEP-style framework and introduce a metric to estimate the impact of any action taken by the skater with possession of the puck. In the spirit of sports analytics' increasingly esoteric acronyms, I call this metric Bayesian Skater Action Value Expectations, or BSAVE. Specifically, BSAVE relies on models that estimate the likelihood that each team will score a goal in the next 15 events. We then isolate the impact of each event by calculating the net likelihood of a goal for or against and comparing it to the same net value before the event occurred. That impact is the BSAVE for that event.

This paper proceeds as follows: In section 2, I develop my methodology. Of particular importance is an effort to derive "skate" events from the existing data. In section 3, I review the performance of the BART models. In section 4, I illustrate what we can learn about player skill and style from the model outputs. Section 5 uses BSAVE for causal inference by studying the impact of having traffic in front of the net during a shot. Finally, I conclude in section 6 with some closing observations and discussions of future work.

2. Methodology

<u>Observations</u>

I used almost all of the 5v5 actions in the original dataset. I removed drawn penalties as these do not reflect an action taken with the puck, though in the future we could predict them as value-creating events like we currently do with goals. For convenience, I removed the first 3 events in each period, as they do not have preceding information to use in their features.

The most significant change made to the original dataset is the inclusion of “skate” events to represent when a player has skated the puck to a new location. These events were derived from the existing dataset by detecting when the player with possession at the end of one event had possession at the start of the next at a different time and location. This addition was critical to ensure that all of the key options available to a skater – skate, pass, shoot, or dump – were included. With skates included, we can safely remove “zone entries”, as this event type is now entirely duplicative of dump-ins, passes, and skates that are already represented.

<u>Independent Variable</u>

Generally, the goal of BSAVE is to understand the likelihood that a goal will be scored “soon”. In soccer, VAEP defines “soon” as within the next 10 events. In this analysis, I chose a binary label based on whether a goal is scored in the next 15 events. This length was chosen because it represents 95% of all possession lengths (Figure 1). This choice seems reasonable because it ensures that we almost always get the full chain of possession, and for shorter possessions, we potentially incorporate chains of events where a team deliberately gives up possession in a more tactical area, such as a dump-in, then regains the puck and scores.

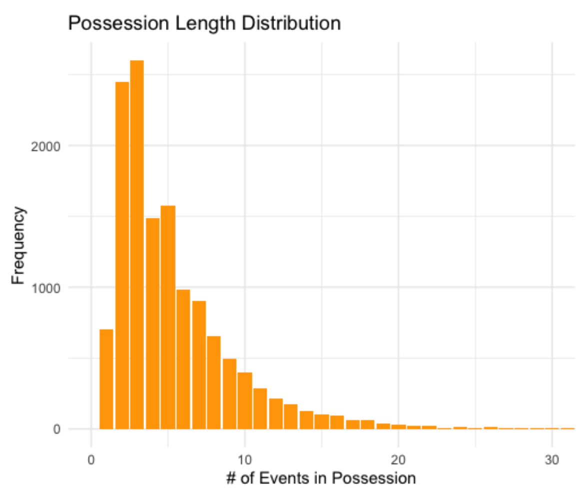

Figure 1

<u>Features</u>

To predict whether a goal will be scored in the next 15 events, I use a total of 54 variables from 4 general categories:

- Game context: The period, score differential, and home or away status
- Event details: The event’s type, time, x and y coordinate, distance, angle, and side from the net, and whether the event was successful. If the event is a shot, we also include the type of shot and whether it had traffic and/or was a one-timer. If the event is a skate or completed pass, we include the end location’s x and y coordinates.
- Prior event details: For the 3 preceding events, we look at the event details listed above except, for simplicity, we omit the shot, pass, and skate specific features

- Complex feature: For the 3 gaps between the 4 events included above, we calculate the time and distance between events and whether they changed team or the side of the goalie on which they occur. We also determine the event's "speed", or distance travelled per time elapsed.

Modeling

With this data, I built Bayesian Additive Regression Models for logistic dichotomous outcomes using the BART package in R [8]. I specified a sparse Dirichlet prior so that BART will conduct variable selection from the options supplied. Little hypermeter tuning was conducted, so this model represents an "out of the box" implementation that could be improved upon. I fit two models: one predicting the likelihood of a goal for the possession team and one predicting the likelihood of a goal against.

Finally, I created three comparisons for each of the two BART models. The first is a simple naïve baseline where every prediction is the average likelihood across the entire dataset. To get stronger comparisons, I also trained a logistic regression and a random forest.

3. Model Evaluation

To determine how well the BART models perform out of sample, I looked at their performance when trained on the first 29 games and then used to predict events in the final 10. The model predicting goals for the possession team appeared to be acceptably calibrated (Figure 2) and outperformed both the naïve baseline and the two comparison models: while the random forest had a stronger AUC, the BART model had a superior log loss and RMSE (Table 1).

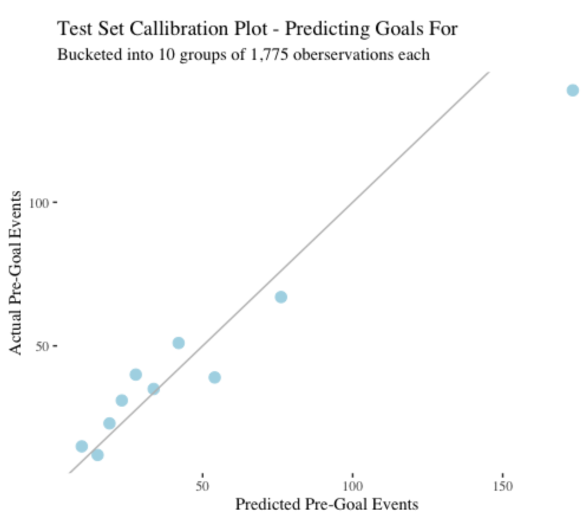

Figure 2

Test Performance – Goals For			
Model	**Log Loss**	**AUC**	**RMSE**
BART	0.112	0.133	0.1560
Log. Regression	0.114	0.103	0.1562
Naïve Baseline	0.119	0	0.1575
Random Forest	0.126	0.248	0.1572

Table 1

Unfortunately, the model predicting goals against the possessing team has admittedly weaker performance. While it does outperform the alternatives, it does not beat the baseline by a large margin (Table 2) and its calibration plot is disappointing (Figure 3). This likely occurred because the possession team is naturally trying to avoid goals against, so we have less insight into what the opponents without the puck are doing; this inherently makes predicting goals against more challenging. In addition, the data is especially unbalanced in this model. This could potentially be improved in a larger dataset or with sampling techniques to address the imbalance. In the meantime, the performance is acceptable enough for directional use here, but we are better served by indexing more heavily on the first model.

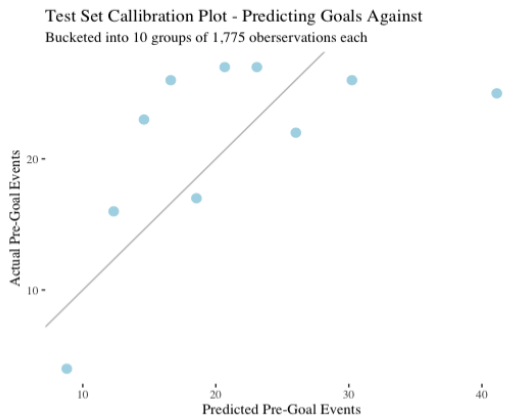

Figure 3

Test Performance – Goals Against			
Model	**Log Loss**	**AUC**	**RMSE**
BART	0.064	0.00036	0.1089
Naïve Baseline	0.065	0	0.1089
Log. Regression	0.066	0.00033	0.1090
Random Forest	0.095	0.00600	0.1102

Table 2

4. BSAVE Analysis

The BART models provide the probability that a goal will be scored by either team shortly after each event. We can now calculate BSAVE, i.e., the isolated impact of each event, by taking the net probability of an upcoming goal and subtracting the same probability from the prior event. This can be expressed as such:

$$BSAVE = (F_t - A_t) - (F_{t-1} - A_{t-1})$$

Where F_t is the probability for the possessing team scoring a goal in the 15 events including and following event t, A_t is the probability that the opposing team scores a goal in the same 15 events, and F_{t-1} and A_{t-1} are the analogous values for the event prior to event t. These BSAVE values can be aggregated to find the impact of each player. For example, we can review the top Erie Otters by average BSAVE contributed per game (Table 3)

Player (5 gp min)	Events / GP	Avg. BSAVE / Event	Total BSAVE/GP
Drew Hunter	60	0.0074	0.44
Jack Duff	72	0.0048	0.35
Luke Beamish	65	0.0052	0.34
Kurtis Henry	70	0.0048	0.34
Jamie Drysdale	81	0.0041	0.33
Jacob Golden	59	0.0054	0.32
Chad Yetman	55	0.0053	0.29
Austen Swankler	60	0.0048	0.29
Brendan Kischnick	44	0.0062	0.27
Brendan Hoffmann	44	0.0053	0.24

Table 3

Furthermore, we can break down the contributions by event type to understand the role of each player. For example, Figures 4 illustrates the frequency and average value of each player's passes, and Figure 5 gives a summary of Jamie Drysdale's contributions per game.

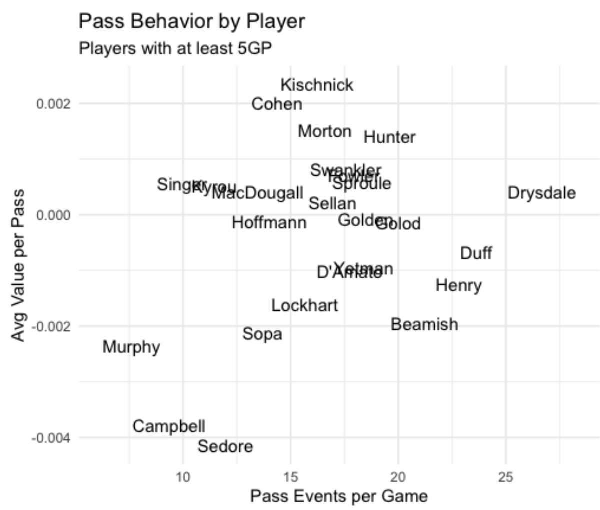

Figure 4

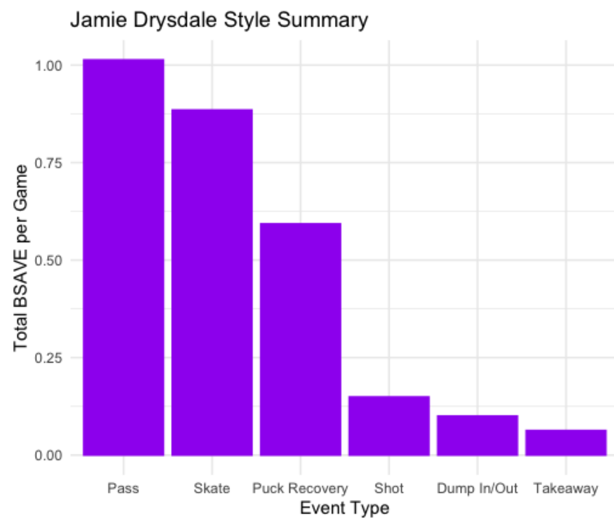

Figure 5

We can see the crucial role of Drysdale's playmaking, as he made the most passes for the Otters while maintaining an overall positive value as he distributed the puck. His skating also adds a significant amount of value. General managers can use this data to better understand the skillset and impact of players available for drafting.

We can also evaluate tactical decisions like which players are choosing good situations to dump the puck (Figure 6). While Kurtis Henry and Luke Beamish dump the puck a similar number of times each game, Henry is generating meaningful value when he dumps the puck whereas Beamish's dumps have negative value. Their coaching staff could review the situations in which Beamish is dumping the puck to see if they should advise him to try alternate tactics.

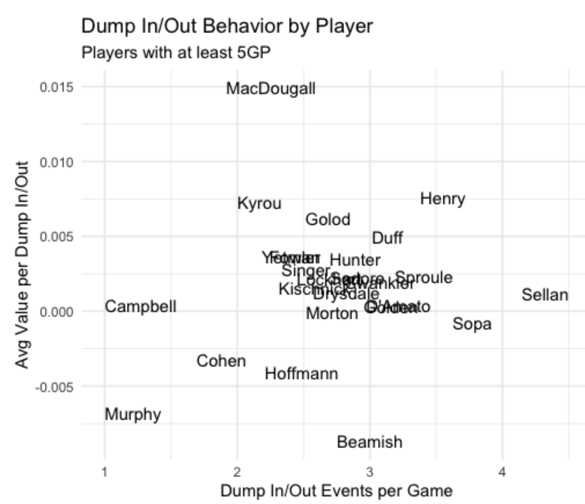

Figure 6

5. Causal Inference: The Value of Traffic in Front

In addition to measuring player contributions, BART models can be used to measure the causal impact of particular variables. To demonstrate, I studied the value added to a shot when it has traffic in front of the net. I duplicated all of the shots in which this occurred, created a counterfactual in which there was not traffic but the rest of the data stayed the same, and generated a prediction of this shot's BSAVE. By comparing this to the original BSAVE value for that shot, we can see that traffic increased the BSAVE of these shots by a point estimate of 0.016 and a 95% confidence interval of 0.002 to 0.057 (Figure 7).

Furthermore, we can dig deeper into these shots to see when traffic has the biggest impact. Initial estimates suggest traffic may be especially effective conditional on shots that are already relatively close to the net, but the wide confidence intervals indicates that we should not make any such conclusions from the available data (Figure 8).

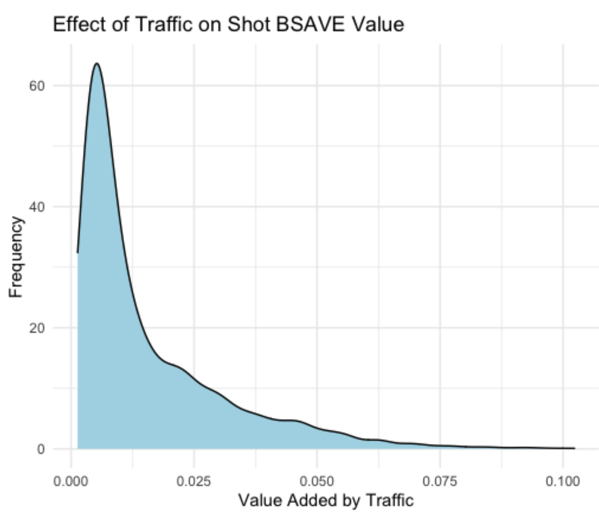

Figure 7

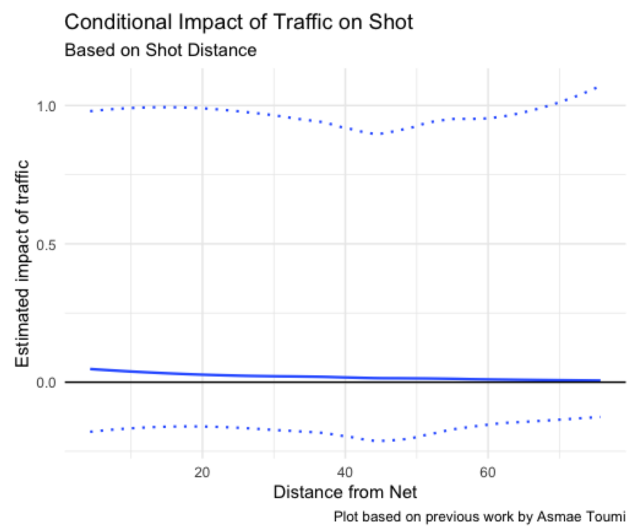

Figure 8

6. Conclusion

BSAVE offers a Bayesian approach to measuring the impact of every play made with the puck. This can be used to understand a player's overall contributions, assess their playing style, and even measure the impact of certain tactical choices.

This framework can be improved. Data could incorporate stoppages: as it stands, some faceoffs and puck recoveries are poorly scored because the prior event does not include the result, e.g., the puck recovery after a save is compared to the situation at the shot, not the save. We also credit players with takeaways but do not penalize players with giveaways. Finally, more could be done with the Bayesian structure of the model, both to confirm that the assumption of normal errors is met and to incorporate the posterior distributions into player analysis.

All code used in this project will be made available at https://github.com/anovet/BDC21. I would like to thank Alyssa Longmuir and Daniel Weinberger for visualization advice and Stathletes for providing this opportunity.

References

[1] Decroos et al. *Actions Speak Louder than Goals: Valuing Player Actions in Soccer* *https://arxiv.org/pdf/1802.07127.pdf*

[2] Singh, Karun *Introducing Expected Threat (xT)* https://karun.in/blog/expected-threat.html

[3] Van Roy et al. *Valuing On-the-Ball Actions in Soccer: A Critical Comparison of xT and VAEP* https://tomdecroos.github.io/reports/xt_vs_vaep.pdf

[4] Chipman et al. *BART: Bayesian Additive Regression Trees* https://projecteuclid.org/journals/annals-of-applied-statistics/volume-4/issue-1/BART--Bayesian-additive-regression/10.1214/09-AOAS285.full

[5] Deshpande, Sameer and Evans, Katherine *Expected Hypothetical Completion Probability* https://arxiv.org/abs/1910.12337

[6] Toumi, Asmae and Lopez, Michael *From Grapes and Prunes to Apples and Apples: Using Matched Methods to Estimate Optimal Zone Entry Decision-Making in the National Hockey League* https://rpubs.com/atoumi/zone-entries-nhl

[7] Carnegie, Nicole Bohme *Introduction to Bayesian Additive Regression Trees for Causal Inference*
https://www.youtube.com/watch?v=9d5-3_7u5a4

[8] Sparapani et al. *Nonparametric Machine Learning and Efficient Computation with Bayesian Additive Regression Trees: The BART R Package* https://www.jstatsoft.org/article/view/v097i01