Big Data Cup 2022: Power Play Danger Index - Adding Context to the Power Play Structure Index

Tim Keller
University of Basel

22.5.2022

Abstract

Using the framework given by Matt Cane's Power Play Structure Index this project creates the Power Play Danger Index that adds the context of expected goals to the Power Play Structure Index.

1 Introduction

While scoring goals is still the main goal of a power play there are different factors that seem to affect the value of of a player advantage situation. One thing is getting into a structured setup and generating scoring opportunities from that setup.

In his Hockey-Graphs article "Measuring the Importance of Structure on the Power Play" (2017) [1] Matt Cane introduced the Power Play Structure Index to "measure a team's power play structure using shot location data" and "quantifies their ability to establish and shoot from a structured formation".
Despite only relying on shot location data the Power Play Structure Index is a better predictor to future success than previous goals but still has a lot of room to improve predictability and repeatability. Nevertheless the index is not perfect and there for sure exist multiple areas of improvement.

The first shortcoming of the index addressed should be the value of the shots considered. Setting up a lot of shots from the point would get a good Power Play Structure Index but the value of the shots are not particularly good. The exact opposite is the case is by setting up shots from the slot but the current version of the index does not know the difference. The concept really helpful for this problem is expected goals because it puts a values on every shot depending on different factors and therefore should add additional context to the index and presumably improves the performance.
Another important topic on the power play is creating space and using the additional space on the ice that is offered by the player advantage. The team on the penalty kill usually tries to keep things tight and desires to prevent players from having scoring chances from open space. But does that improve the performance of our index and is it worth to go away from the simplistic approach of the Power Play Structure Index?

The goal of this project was to take the framework that Matt Cane created with the Power Play Structure Index and iterate over different concepts of importance in a power play to be able to add further context and nuance to the index.

2 Methodology

2.1 Recreation Power Play Structure Index

With our available data we started to recreating the Player Structure value and from that the Team Structure Index directly with SQL.
To calculate the values we used the following formulas:

$$\text{Player Structure} = \frac{\sum \text{distance to avg shot location}}{\sum \text{Shots for player}}$$

$$\text{Team Structure Index} = \frac{\sum(\sum \text{Shots for player} * \text{Player Structure})}{\sum \text{shots of all players}}$$

2.2 Expected Goals Model

To put some further context into the structure of the power play I built an expected goals (xG) model to put some context to the Structure Index so that it not only is able to tell who is able to constantly create shots from the same positions but also constantly from good positions.

The feature selection for the xG model was inspired by the series of articles "Expected Goals Model with Pre-Shot Movement" (2019) [2] Alex Novet wrote on Hockey-Graphs. The importance of distance and angle to the net but also the importance of previous events are big part of my feature set:

- shot type
- x coordinate, y coordinate and distance to the goal
- angle to the goal
- shot type, x coordinate, y coordinate of the previous two events
- the time remaining in the game
- time difference, angle difference and difference for distance to the goal for the last and second to last event
- strength state

Despite them normally being excluded from xG models blocked shots were taken into account for this model to not further constrain the sample size and have xG values for all shots in the Structure Index.

A XGBoost classifier with grid search as parameter estimator was used for modelling with shots from the 2021 Big Data Cup datasets as training data but also a 10 KFold split of the current data to have a better ratio of training and testing data. Using AUC scoring the model received a value of 0.833.

2.3 Powerplay Danger Index

Now from the results of the xG model I decided to create an additional index for taking shots from more "dangerous" positions than average:

$$\text{Player xG Index} = \frac{\sum(xG - AVG(xG))}{\sum(shots)}$$

And a similar approach to a Team Index as in the Team Structure Index:

$$\text{Team xG Index} = \frac{\sum(\sum(Shots) * \text{Player xG Index})}{\sum(Shots)}$$

Continuing from there the goal was to put these two Index values into one. Now the question was how they should be weighted to each other and how they be combined.

With the xG Index aiming to be as high as possible and the structure index trying to be as close to zero as possible it just made sense to subtract the structure index values from the xG Index and therefore a value as high as possible would be the desired result.

The Structure Index values usually have a higher standard deviation therefore there needs to be a weighting lowering the structure value so either value has a factor in the final result. Using predictability as deciding factor I settled on weighting the Structure Index with the value 0.01 (which also has some shortcomings discussed in the findings) and came up with this final formulas for the Power Play Danger Index of players and teams:

$$\text{Player Danger Index} = \frac{\sum(xG - AVG(xG))}{\sum(shots)} - \frac{\sum(\text{distance to avg shot location})}{\sum(Shots) * 100}$$

$$\text{Team Danger Index} = \frac{\sum(\sum(Shots) * PlayerxGIndex)}{\sum(Shots)} - \frac{\sum(\sum(Shots) * PlayerStructureIndex)}{\sum(Shots) * 100}$$

3 Findings

3.1 Power Play Structure Index

Because of the size of the data set the interpret-ability of the results is very limited and tend to be very extreme with multiple players having only one or two shots. But with some caution we can say that Elisa Holopainen, Alina Müller and Megan Keller are examples for players with a really good power play structure. Lara Stalder, Hillary Knight and Michelle Karvinen take their shots on the power play from a lot of different position and therefore have a high player structure value (Figure 1).

Regarding Teams the differences of power play shots in the dataset range from 13 (ROC) to 39 (Finland) where ROC has the lowest Power Play Structure Index value and Switzerland the highest which seems surprising as Switzerland was praised as having a great power play (Figure 2).

3.2 Powerplay Danger Index

Using the same two measures Matt Cane used in the original article repeatability and predictive power we can look if the newly created power play Danger Index creates any additional value.

Using split-half Pearson correlation we can see that both the structure index and the xG Index outperform the danger index but not by a concerning margin:

- Team Danger Index: 0.498

player_name	team_name	powerplay_shots	sum_distance_to_avg_shot_location	player_structure_index
Elisa Holopainen	Olympic (Women) - Finland	8	44.291275	5.536409
Alina Muller	Olympic (Women) - Switzerland	3	23.949691	7.983230
Megan Keller	Olympic (Women) - United States	3	24.456078	8.152026
Nina Pirogova	Olympic (Women) - Olympic Athletes from Russia	3	25.347867	8.449289
Minnamari Tuominen	Olympic (Women) - Finland	11	125.630135	11.420921
Petra Nieminen	Olympic (Women) - Finland	4	47.481147	11.870287
Nelli Laitinen	Olympic (Women) - Finland	5	64.436039	12.887208
Anna Shokhina	Olympic (Women) - Olympic Athletes from Russia	6	78.080377	13.013396
Alex Carpenter	Olympic (Women) - United States	7	98.582263	14.083180
Angelina Goncharenko	Olympic (Women) - Olympic Athletes from Russia	4	61.166788	15.291697
Jenni Hiirikoski	Olympic (Women) - Finland	6	98.926297	16.487716
Phoebe Staenz	Olympic (Women) - Switzerland	4	74.141960	18.535490
Cayla Barnes	Olympic (Women) - United States	7	140.908468	20.129781
Lara Stalder	Olympic (Women) - Switzerland	8	166.416791	20.802099
Hilary Knight	Olympic (Women) - United States	3	67.567432	22.522477
Michelle Karvinen	Olympic (Women) - Finland	5	125.287689	25.057538

Figure 1: Player Structure calculated from average position of power play shots

team_name	team_shots	team_structure_index
Olympic (Women) - Olympic Athletes from Russia	13	12.661156
Olympic (Women) - Finland	39	12.975707
Olympic (Women) - United States	20	16.575712
Olympic (Women) - Switzerland	15	17.633896

Figure 2: Team Structure Index calculated from average position of power play shots

- Team Structure Index: 0.526
- Team xG Index: 0.583

Regarding the predictability of goals none of the index values outperforms expected goals in Pearson correlation but that is not necessarily the goal. With the way we set up the weighting we were able to outperform both the Structure Index and the xG Index in positive correlation to goal-scoring:

- Team Danger Index: 0.071
- Team Structure Index: -0.061
- Team xG Index: 0.04

A shortcoming of designing the Power Play Index like that, is that it results in a lot of negative values. Only one player in the dataset with at least 3 power play shots has a positive Player Danger Index (Figure 3) and no teams have a positive value for the Team Danger Index (Figure 4).

The biggest jump in the ranking from Structure Index to Danger Index made Michelle Karvinen from last in Structure Index and first in xG Index helps her to jump the board in the Danger Index. The biggest negative jump was by Alina Müller but, with a negative xG Index, she is still in the upper half for the Players Danger Index. On the

player_name	powerplay_shots	player_structure_index	player_xg_index	player_danger_index
Michelle Karvinen	5	25.057538	0.338378	0.087803
Petra Nieminen	4	11.870287	0.110047	-0.008655
Elisa Holopainen	8	5.536409	0.021792	-0.033572
Nina Pirogova	3	8.449289	0.016614	-0.067879
Phoebe Staenz	4	18.535490	0.106481	-0.078874
Megan Keller	3	8.152026	0.002266	-0.079254
Alina Muller	3	7.983230	-0.003548	-0.083380
Minnamari Tuominen	11	11.420921	-0.002256	-0.116465
Anna Shokhina	6	13.013396	0.003333	-0.126801
Jenni Hiirikoski	6	16.487716	0.029753	-0.135125
Angelina Goncharenko	4	15.291697	0.002437	-0.150480
Nelli Laitinen	5	12.887208	-0.030727	-0.159599
Alex Carpenter	7	14.083180	-0.042718	-0.183549
Lara Stalder	8	20.802099	-0.000805	-0.208826
Cayla Barnes	7	20.129781	-0.028250	-0.229548
Hilary Knight	3	22.522477	-0.030866	-0.256090

Figure 3: Player Danger Index calculated from average position of power play shots

teams side the differences are not that big ranking wise but the Danger Index helps to contextualize the differences between the teams further.

team_name	team_powerplay_shots	team_structure_index	team_xg_index	team_danger_index
Olympic (Women) - Canada	10	5.597423	0.012346	-0.043628
Olympic (Women) - Finland	47	11.908058	0.047045	-0.072035
Olympic (Women) - Olympic Athletes from Russia	19	10.004744	0.001936	-0.098111
Olympic (Women) - Switzerland	19	14.717957	0.019600	-0.127579
Olympic (Women) - United States	29	13.515063	-0.017030	-0.152180

Figure 4: Team Danger Index calculated from average position of power play shots

4 Further Thoughts

Initially the idea was to not stop at one iteration of improvement of the Power Play Structure index with all the potential of added context with the availability of the data. Another interesting and important dimension of the power play is finding and creating space while the team with the player disadvantage tries to keep everything tight. I tried to look at differences for the distance to the nearest defenders between events but was not able to bring the idea particularly far and also was not able to incorporate it into the idea of the index. But it is something I would love to further explore. Additionally the question needs to be asked if an index formula is still the right framework to express all the additional context in.

In the overall topic of structure interested to explore, using the tracking data, would

be the positions of all players of the ice when reaching the state of a formation on the power play by finding another spin on the Structure Index.
Finally, because of the limited number of shots in the data set, I tried to recreate the danger and structure indices by not only using shots but also using all Events in the offensive zone and a non-shot expected goals model, assigning expected goals values to every event in the data set. Unfortunately the new index performed worse than the index only using shots on the power play.

5 Conclusion

We were able to improve the initial Power Play Structure Index a little bit by providing some context adding expected goals to the equations and creating the Power Play Structure Index. With the availability of tracking and event data that does not have to be the last piece of context added to the equation but for sure is an important one because what value do consistent shots on the power play really have if it is not from a position with high chance of scoring a goal.

My code for the Big Data Cup 2022 is available on my GitHub repository: `https://github.com/TK5-Tim/Big-Data-Cup/tree/main/2022`

6 References

[1] Cane, Matt: `https://hockey-graphs.com/2017/02/14/measuring-the-importance-of-structure-on-the-power-play/`, 2017, last accessed 22.05.22

[2] Novet, Alex: `https://hockey-graphs.com/2019/08/12/expected-goals-model-with-pre-shot-movement-part-1-the-model/`, 2019, last accessed 22.05.22