On the Attack: Using Analytics to Unlock the Secrets of Successful Zone Entries in Hockey*

Anton Olivestam[1], Axel Rosendahl[1], Hampus Svens[1], and Lukas Hellberg[1]

Linköping University liu.se

Abstract. In this paper, we use machine learning methods to see what types of offensive zone entries and their upcoming play sequence lead to shots with high xG values in hockey. Our results showed a clear difference in sequence characteristics regarding the length of the sequence, the type of zone entries, the number of events, and in what zones those events took place for good goal-scoring opportunities compared to worse ones.

Keywords: Zone entries · Neural Networks · xG · Play sequence

1 Introduction

In this paper, we represent an analysis of different offensive zone entries and the upcoming play sequence to see which type of sequences that leads to shots with high expected goal value (xG). This we will do by creating a neural network model to analyze the sequences and then gather statistics to see which zone entries lead sequences resulting in a high xG value.

An offensive zone entry is defined as the act of a team carrying, passing, or dumping the puck across the opposing team's blue line, i.e. entering the offensive zone. Wilderoth *et al.* found that no goals during SHL seasons 2018/19, 2019/20, and 2020/21 were scored without being in the offensive zone given that both teams were at full strength and that the goalkeeper has not been pulled [1]. This makes entry into the offensive zone crucial in the game of hockey.

Similar work has been done by Chatel [2]. Chatel divided the game into sequences where each sequence started from every possession change. The post looked at what particular outcome a type of zone entry would lead to frequently. It did this by looking at how the expected goal (xG) value changed. Chatel captured this into a decision tree that shows the risk and reward for each action in different zones.

2 Background

Offensive zone entries can be divided into three types. (1) a dump-in, (2) a carry-in, and (3) a pass. However, a further division of these zone entries was made in this study. A carry-in was split into two versions: wideCarry and middleCarry.

* Supported by Linköping University.

These different versions describe where along the blue line the player skated into the offensive zone with the puck. A pass entry was split into two different versions: widePass which meant that the puck moved more than half of the rink size in the sideways direction and shortPass which means that the puck moved less than half the rink size in the sideways direction.

Table 1. Statistics for the different offensive zone entries.

Entry	Percentage of total entries	Average xG on shots	Shots	Goals
wideCarry	37%	0.05	44.5%	3.3%
middleCarry	16%	0.07	47.4%	5.2%
widePass	4%	0.05	64.7%	4.7%
shortPass	6%	0.07	48.7%	9%
dumpIn	37%	0.06	17.4%	1%

In Table 1 we have displayed some statistics gathered from our dataset, which is based on 20 SHL games, to give a hint of what to expect from our analysis. Here we could see that a WidePass most often leads to a shot on net. However, a shortPass is the entry that most often leads to a goal and also has the highest average xG. Further, we can also see that a middleCarry more often leads to a goal and has a higher average xG than a wideCarry.

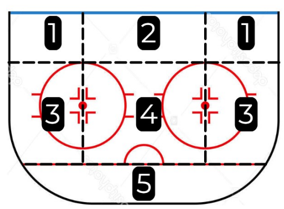

Fig. 1. Zone split.

Further, the offensive zone was divided into five different sub-zones in order to make it easier to analyze the data. In Figure 1 a visualization of the zones is made. The two zones 1 is called upperWide, zone 2 is called upperMiddle, zones 3 is called middleWide, zone 4 is called slot, and zone 5 is called behindGoalCrease.

3 Algorithms

In this paper, we present one model created in order to analyze the entry types and the sequences after. This model is an LSTM model created to predict xG based on a certain entry type and the sequence after.

3.1 Recurrent Neural Network

Recurrent Neural Networks (RNN) have been used in a huge variety of different areas due to their ability to predict the next event in a sequence [4]. However, a traditional RNN struggle to capture long-term dependencies due to the vanishing gradient problem [3]. To tackle this, LSTM (Long short-term memory) a more complex architecture that was specifically designed to overcome these issues, was introduced. Since our data consists of sequenced events and our goal is to predict the next event, or more specifically predict the xG after our sequence, LSMTs, and RNNs were a highly suitable tool [4].

4 Results

4.1 Predict xG

In Figure 2 we can see the performance of our LSTM model. The scatter plot suggests that our neural network has a decent level of predictive capability for xG.

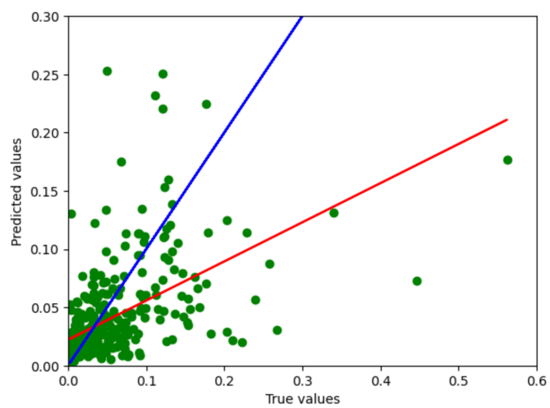

Fig. 2. The plot shows the predicted xG in relation to the true xG in the data. The red line is the regression line for all of the predicted xGs and the blue line is the optimal regression line that we strive to get the points to be close to.

Table 2. Average number of occurrences for five events for the 100 sequences with highest predicted xG and the 100 sequences with lowest predicted xG values.

Event name	High xG sequences	Low xG sequences
pass	28%	22%
reception	24%	15%
lpr, (loose puck recovery)	18%	24%
shot	5%	3%
puckprotection	4%	7%
average sequence length	23.52	9.66

In Table 2 we have extracted the sequences where our model predicted best xG values versus the sequences where our model predicted worst xG values and displayed how often five different events on average occurred in the two types of sequences. From Table 2 we can see that the biggest difference between the two types of sequences is that for those leading to a higher xG the number of receptions is higher whilst the number of passes is slightly higher. This means that the number of successful passes is a lot higher for high xG sequences. Another big difference is that the number of lpr's are more frequent for the low xG sequences which probably means that there are fewer direct passes and more lost pucks from the offensive team in those sequences.

Table 3. Average percentage of events in each zone of the offensive zone split.

Zone	High xG sequences	Low xG sequences
middleWide	26%	19%
slot	21%	13%
behindGoalCrease	20%	22%
upperWide	18%	32%
upperMiddle	15%	14%

In Table 3 we can see the average percentage of events happening in each zone from Figure 1 in the two different sets of sequences. What we can see in this table is that high xG sequences have more events in the slot and middleWide zones than low xG sequences. More events happening in the slot for high xG are expected since shots from this area are a good scoring opportunity. More events in the middleWide zone probably give a higher xG since a player controlling the puck in this zone has a lot of options, the player can shoot, pass to the slot, or pass over the central line to a player in the same zone but on the opposite side which all could create a great scoring opportunity. Also as expected low xG sequences have a lot of events in the upperWide zone which is furthest from the goal and a zone that is hard to create a scoring opportunity from.

4.2 Zone Entries

Table 4. Statistics for frequent patterns of zone entries.

Entry	Sequences	Slot shots	Receptions in middleWide	Ratio A	Ratio B
wideCarry	527	209	219	0.4	0.42
middleCarry	255	66	25	0.26	0.098
widePass	55	17	23	0.3	0.42
shortPass	87	27	152	0.31	1.7
dumpIn	37	0	123	0	0.59

In Table 4 we have calculated statistics for each zone entry based on the best zones and events to have in a sequence according to our results from the LSTM model. These zones were middleWide and slot and we added the reception event to the middleWide zone and the shot event to the slot zone. The ratio A and ratio B in the table are calculated as follows:

$$Ratio\ A = \frac{Shots\ in\ slot}{Total\ sequences}\ \ ,\ \ Ratio\ B = \frac{Receptions\ in\ middleWide}{Total\ sequences} \tag{1}$$

The sequences used in Table 4 are the same sequences that were used in our LSTM model which means that these are all the sequences that included a shot. From Table 4 we can see that the wideCarry most often leads to a slot shot amongst these sequences. The biggest difference between the zone entries are the number of receptions in the middleWide zone which on an average happens 1.7 times for a shortPass entry while the second most frequent is only 0.59 times. This makes shortPass a great entry since we could see from Table 2 and 3 that having more reception events and more play in the middleWide zone relates to a high xG value. This we can also relate back to Table 1 were we could see that a shortPass entry most often ended up in a goal and had the highest average xG out of all entries.

4.3 Summary

In this paper we have analyzed different offensive zone entries and with the help of a neural network gained further insights into sequences of play in ice hockey. We found clear differences in characteristics for sequences with a high predicted xG value compared to their counterpart sequences with low xG values. We found that the most significant difference between the sets of sequences where that more events of receptions meaning more successful passes, more events in the middleWide zone, and more events in the slot zone lead to a higher xG value. We then found that the entry that most often had receptions in the middleWide zone was shortPass entry which we also could clarify led to most goals our of the entries in our dataset. Further, we found that a wideCarry entry most often lead to a shot in the slot zone amongst the sequences that included a shot.

4.4 Future ideas and improvements

To further analyze zone entries it would be interesting to look at sequences before the zone entry as well and see how that affects the performance of a offensive zone entry. Another thing that could be improved in this paper is the size of the dataset used since our only included 2550 zone entries and 961 zone entry sequences including a shot which was used for the LSTM model. This could be reason to why we "under-predict" the high xG values since they are not that common in out small dataset.

5 Code appendix

The code for the LSMT model and scripts for calculating staistics used in this paper can be found in the following GitHub link:
https://github.com/Olivestam/linhac_zone_entries

References

1. Wilderoth, Erik, Ulf Johansson, and Arsalan Sattari. "Where not to lose the puck." Linköping Hockey Analytics Conference. 2022.
2. Chatel, T. (2020, March 26). Introducing offensive sequences and the Hockey Decision Tree. Hockey Graphs. https://hockey-graphs.com/2020/03/26/introducing-offensive-sequences-and-the-hockey-decision-tree/ . Last accessed 10 Apr 2023
3. Hochreiter, Sepp. "The vanishing gradient problem during learning recurrent neural nets and problem solutions." International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 6.02 (1998): 107-116.
4. Sherstinsky, Alex. "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network." Physica D: Nonlinear Phenomena 404 (2020): 132306.