Wide Attack Battles

Lori He

skillcorner [DATE]

1 Analysis Scope

Season: WSL 2023-2024

Object: Actions after a player receiving a pass in the Wide Attack Area.

- Wide Attack Area: outside of the box and in the attacking third.

Methods:

- Define the success (attacking-wise) of the sequence following the pass.
- Quantify the difficulty of context as to gaining success
- Modeling to reveal the variance of result that can be explained by the context
- Extract cases where an attacker gained success under difficult to attack context, and where a defender prevent success under easy to attack context.
- Rank players accordingly

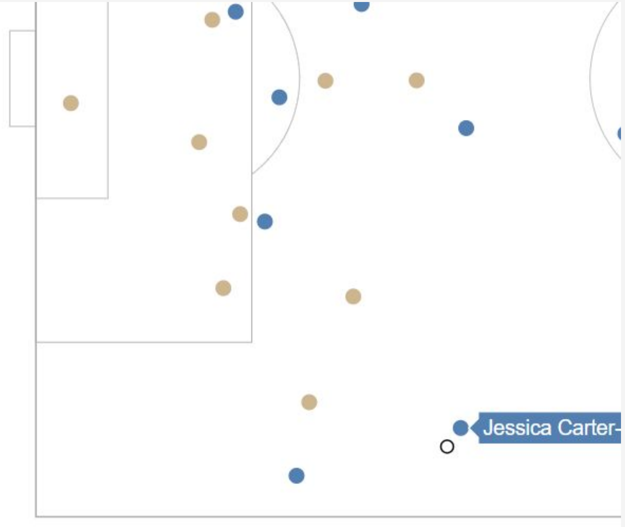
1.1 Define Success of Sequence

Start of sequence: the frame when the distance between the ball and the receiving player is less than 2 meters.
End of sequence: the second pass or shot event after the start of sequence.

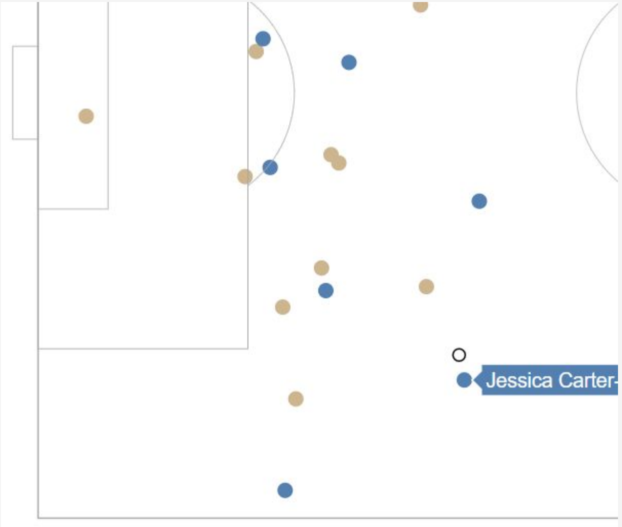
A sequence is defined as success if it fits the criterias below:

- The ball's new location is 2 meter closer to goal than where it was at the start of sequence.
- The attacking team continuously holds possession until the ball moves to the location described above.

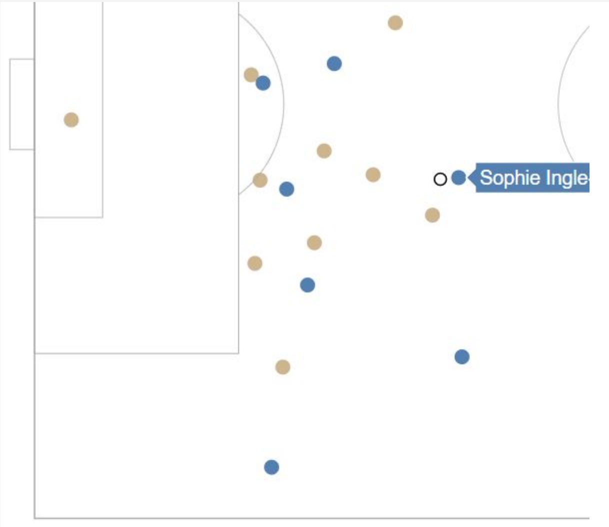
Start of sequence
Ball to goal distance: 42.04 m



Start of sequence
Ball to goal distance: 40.38 m



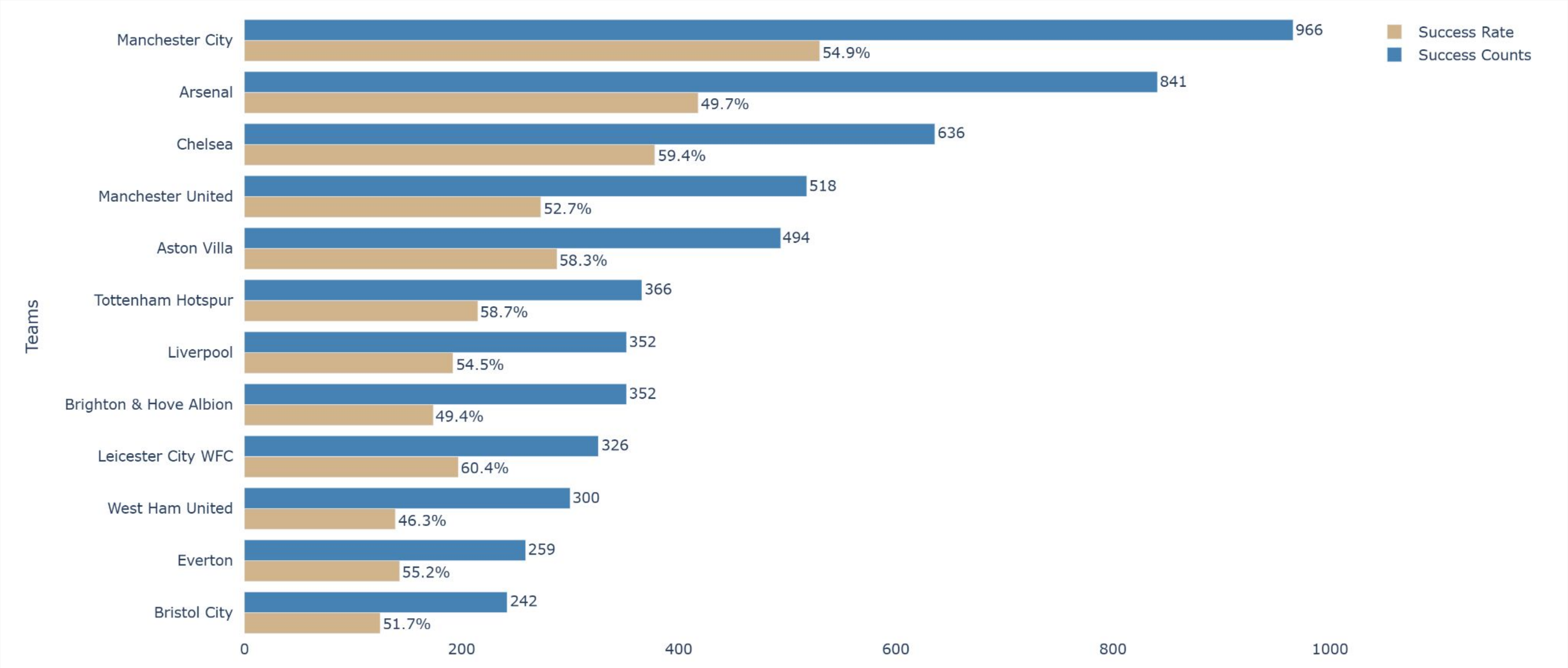
Start of sequence
Ball to goal distance: 34.68 m - **Success**



During this project, the location of possessing player was used to indicate the ball location, as event frames were marked during the passes at which the ball was away from where the player last possessed it.

1.2 Team Performance

Manchester CIty and Arsenal seat on the top as to the count of successful sequences.
However, some teams with a much lower counts ends with an nice success rate - Leicester City succeeded 60.4% of all the wide attack sequences.



2 Context Variables

As the team performance differs, each sequence faces difference attacking difficulties. Here are a few variables considered to have influences on the sequence result.

- Score difference
  The scoring situation at the moment can affect players' intention of being aggressive or not. It also potentially reveals the skill level difference between the teams.

- Count of defender in the box
  With a sufficient number of defenders in the box, the main defender of the sequence has more freedom to leave a space behind and press hard, thereby increasing the attacking difficulty.

- Count of attacker (above the attacking point)
  From the 2 meters behind ball location to the goal line, all the attackers within this zone are considered as potential helpers to assist the sequence.

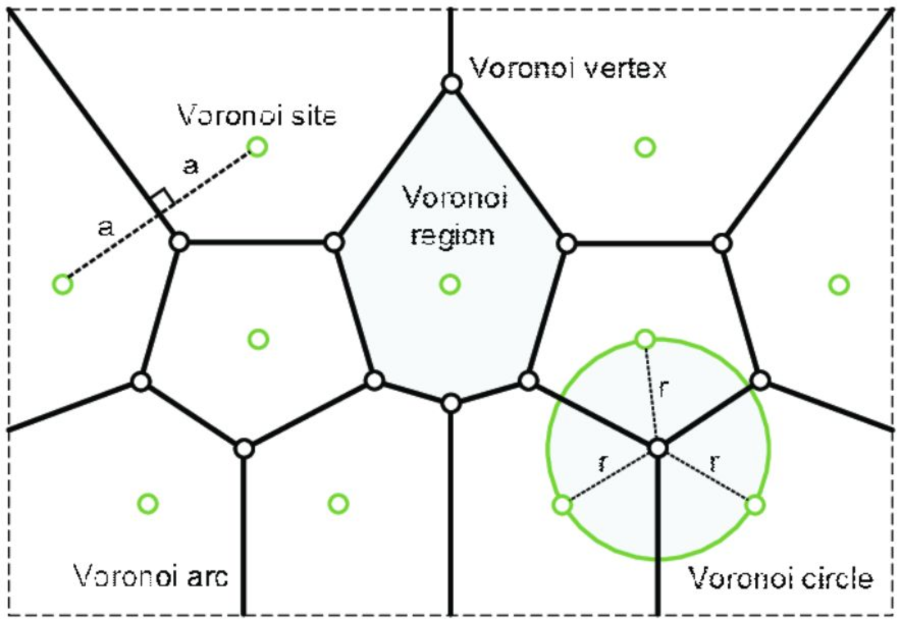
- Team
  Whether playing at a home stadium is commonly considered to have influence on the team's performance

- Sum of Freedom
  The space in which the attacker can move to with relatively lower threat. This is calculated using Voronoi diagram and explained in the next page.

2.1 Sum of Freedom

[A Voronoi diagram is a type of tessellation pattern in which a number of points scattered on a plane subdivides in exactly n cells enclosing a portion of the plane that is closest to each point.]

Simply said, any point on the edge between two green circles in the pic has equal distance to each circle. Any point within a green circle's Voronoi site is closer to this circle than any other circles.



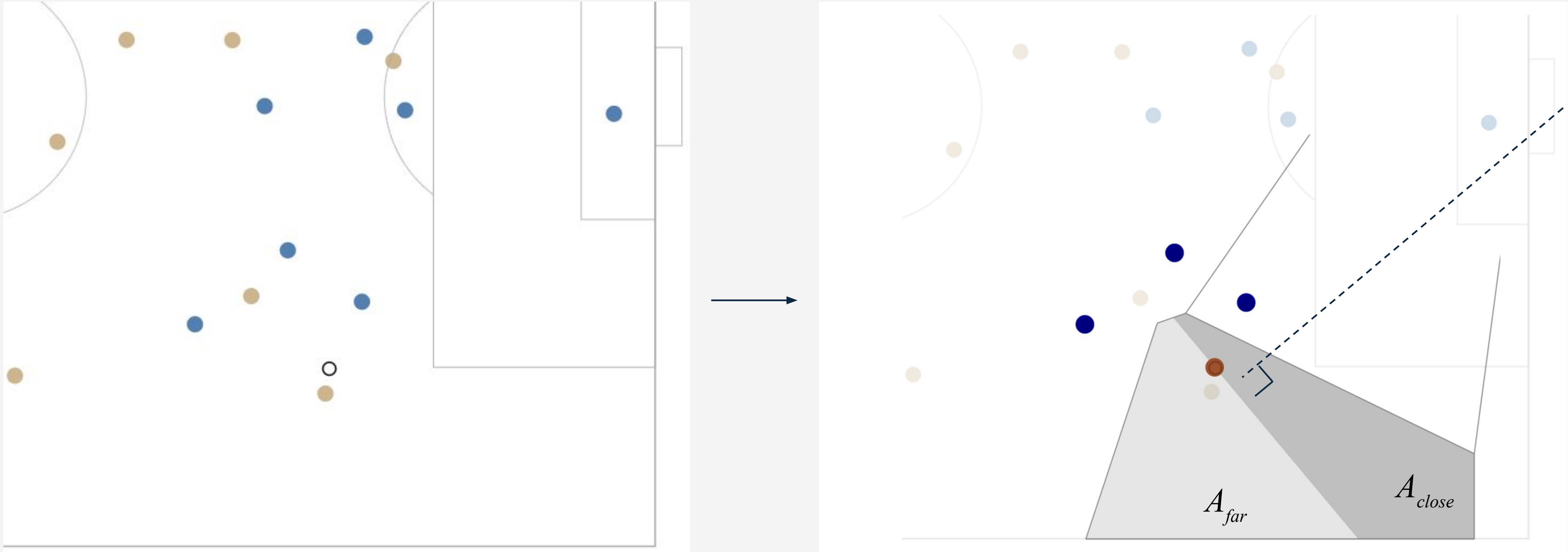
2.1 Sum of Freedom

For each sequence, a Voronoi site was calculated on the ball and nearby defenders. The site is divided by a line perpendicular to the line between ball and center of goal. The Sum of Freedom is defined as a sum of the site area closer to the goal and the are further from the goal, a heavier weight was applied to the closer area.

The boundary of the Voronoi diagram was the sidelines, goal lines, and 20 m away from the ball to four directions.

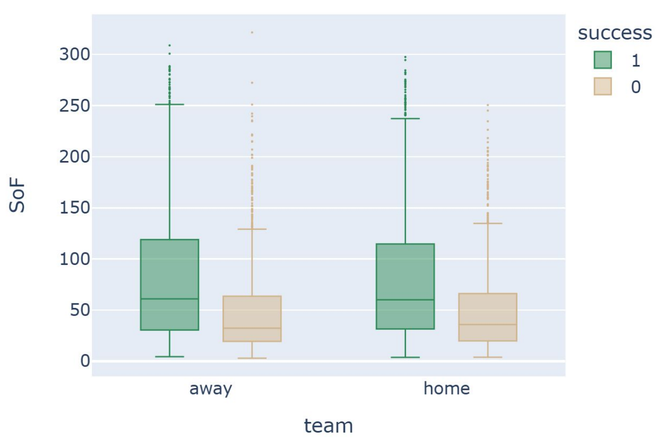
$$SoF = log(A_{far} + 1) + A_{close}$$

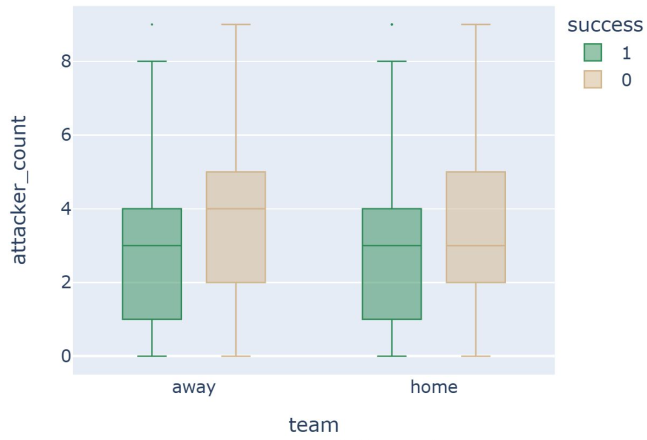
Eg1. SoF = 166.9 m^2 Result: Success

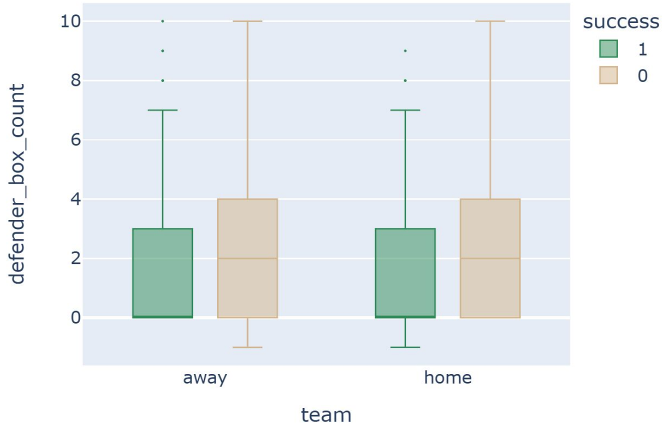


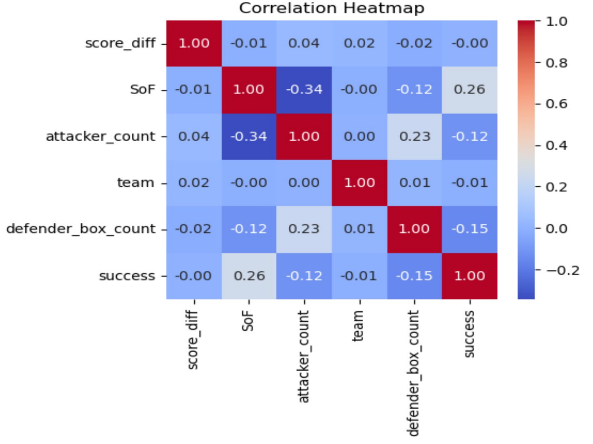
2.2 Feature exploration

A few feature exploration reveals the significance of SoF, defender_box_count, and attacker_count to the sequence result.









3 Models

Using the variables mentioned as features, a few classification models were trained to reveal how much variance of the sequence result can be explained by contextual influences.

In the report on the right, label 1 represents success, label 0 represent unsuccess.

```
Model: RandomForest
Accuracy: 0.6438
Classification Report:
              precision    recall  f1-score   support

           0       0.59      0.67      0.63       203
           1       0.70      0.63      0.66       249

    accuracy                           0.64       452
   macro avg       0.64      0.65      0.64       452
weighted avg       0.65      0.64      0.64       452

Confusion Matrix:
[[135  68]
 [ 93 156]]
```

```
Model: GradientBoosting
Accuracy: 0.6372
Classification Report:
              precision    recall  f1-score   support

           0       0.57      0.78      0.66       203
           1       0.75      0.52      0.61       249

    accuracy                           0.64       452
   macro avg       0.66      0.65      0.64       452
weighted avg       0.67      0.64      0.63       452

Confusion Matrix:
[[159  44]
 [120 129]]
```

```
Model: LogisticRegression
Accuracy: 0.6283
Classification Report:
              precision    recall  f1-score   support

           0       0.56      0.75      0.65       203
           1       0.72      0.53      0.61       249

    accuracy                           0.63       452
   macro avg       0.64      0.64      0.63       452
weighted avg       0.65      0.63      0.63       452

Confusion Matrix:
[[153  50]
 [118 131]]
```

```
Model: KNeighbors
Accuracy: 0.5774
Classification Report:
              precision    recall  f1-score   support

           0       0.53      0.52      0.52       203
           1       0.61      0.63      0.62       249

    accuracy                           0.58       452
   macro avg       0.57      0.57      0.57       452
weighted avg       0.58      0.58      0.58       452

Confusion Matrix:
[[105  98]
 [ 93 156]]
```

4 Player Evaluation

The variance that were not explained by contextual influences should be heavily influenced by the players' capabilities:

- Outstanding attackers would succeed on scenarios which the model considered to be difficult to succeed.
- Outstanding defenders would prevent an attacking success on scenarios considered to be easy to succeed.

3392 randomly selected sequences were left out before the model training and was used for player evaluation.

Among the four models, the GradientBoosting model was chosen to evaluate attacker, because of its better performance on label 0 (unsuccess), the RandomForest model chosen to evaluate defender, because of its better performance on label 1 (success).

4.1 Attacker Evaluation

Among attacker who have faced more 25 difficult scenarios from the evaluation set, Mary Fowler and Kirsty Hansen stood out with their over 0.6 success ratio.

pass_recipient_name	success_difficult_ratio	difficult_count
C. Kelly	0.569892	93
C. Foord	0.538462	65
L. Hemp	0.519231	52
S. Catley	0.404255	47
B. Mead	0.409091	44
Leila Ouahabi	0.465116	43
E. Toone	0.585366	41
N. Charles	0.512195	41
M. Fowler	0.628571	35
H. Blundell	0.375000	32
L. Galton	0.290323	31
K. Dali	0.483871	31
K. Hanson	0.607143	28
K. Casparij	0.370370	27
S. Mayling	0.518519	27
L. James	0.500000	26
K. Smith	0.384615	26
K. Robinson	0.400000	25
K. McCabe	0.240000	25

4.1 Defender Evaluation

Defender evaluation was focused on the 'paired players' who were the closest to the ball receiver at the start of sequence.

Among attacker who have faced more 20 difficult scenarios from the evaluation set, Ella Powell, Niamh Charles, Mayumi Pacheco, and Catherine Bott stood out with their over 0.4 prevent success ratio.

paired_player	unsuccess_easy_ratio	easy_count
Ella Mae Florence Powell	0.400000	55
Heather Payne	0.361702	47
Ffion Morgan	0.264706	34
Jamie-Lee Napier	0.303030	33
Hannah Blundell	0.193548	31
Emma Kullberg	0.266667	30
Sarah Emma Mayling	0.366667	30
Kirsty Smith	0.259259	27
Niamh Louise Charles	0.481481	27
Mayumi Pacheco	0.480000	25
Emma Koivisto	0.250000	24
Naomi Layzell	0.391304	23
Catherine Joan Bott	0.428571	21
Katie Robinson	0.285714	21
Clare Wheeler	0.250000	20