

Identifying Underrated Players in the Context of Major League Soccer

Barinder Thind, Alex Land, Brad Smallwood, Siddharth Arora

VANSASH Hackathon

VANCOUVER

September 22, 2018

Overview

- 1 Approach
- 2 Methodology
 - Getting Some Context
 - Shot Zones
 - Identifying Important Passes
 - Putting It Together
- 3 Results
 - Results
- 4 Next Steps
 - Next Steps
- 5 Conclusion
 - Conclusion
 - References

Approach

Approach

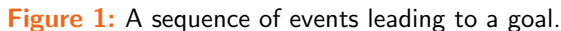
- On plays where a goal is scored, the sequence of events leading up to the goal may be helpful in identifying good players.
- We want to find the most important events in a sequence that leads to a goal, and the players that are involved.
- However, using a sequence of events has some issues:
 - ★ What constitutes a sequence? Possession? A certain number of events?
 - ★ Are all events in a sequence relevant for the goal that is scored? How do we decide that an event is relevant?

Approach

- We can split the pitch in to "zones" where goal attempts are commonly made, each of which has a similar number of attempts. Passes between zones are considered important passes in setting up a goal attempt.
 - ★ We don't have to worry about defining a sequence, since we only care about the direct assists to the goal.
 - ★ We don't have the issue of unrelated events in a sequence throwing us off.
- This allows us to focus our analysis on two important aspects of creating a scoring opportunity:
 - ★ The assisting player identifying and executing the pass.
 - ★ The scoring player getting open, receiving the pass, and scoring.

Methodology

- Initially, we looked at the sequence of events leading up to a goal to give us some context and help us brainstorm.



Shot Zones

- We used K-Means clustering to find a set of shot zones where a similar number of goal attempts are made.

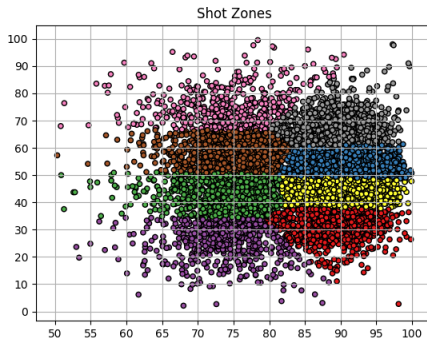


Figure 2: Shot zones.

Identifying Important Passes

- We filtered and manipulated the Whitecaps game event data to identify the important passes (ie. passes that travelled between zones) resulting in a goal.
 - ★ For this preliminary analysis we only look at the most recent inter-zone pass that resulted in a goal.

Overall Method

- The overall approach can be summarized in the following equation:

$$\sum_{j=1}^{NumberDangerZones} P_{player_i}(\alpha|\phi) + \lambda_j(Penalization)$$

Where α is that a goal is scored and ϕ is that they passed it into danger zone j

- The final goal is to look at the players who are successfully passing the ball into the danger zone and those passes are actually being converted to goals. We also are looking to find player who can actually get open

Results

Results

- For every 26 passes, a goal was scored.
- Of all passes in to the 2 most dangerous zones (in front of the net) that result in a goal, the average completion rate is 57

```
> summary(assist_players$Player)
```

```
Lee Nguyen      8
Romain Alessandrini 6
Alberto Josué Elis Martínez 4
Diego Valeri    4
Marlon Hairston 4
Michael Parkhurst 4
Ricardo Clark   4
Salvatore Zizzo  4
Andrew Farrell  3
Benny Feilhaber 3
David Villa     3
Graham Zusi     3
.
```

```
Héctor Daniel Villalba 6
Victor Vázquez          6
Bastian Schweinsteiger 4
Lloyd Ekow Sam          4
Matt Polster            4
Óscar Boniek García     4
Robert James R.J. Allen 4
Alex                    3
Anthony Jackson-Hamel   3
Danny Hoesen            3
Diego Fagúndez          3
Ibson                  3
.
```

```
Nicolás Lodeiro 6
Sacha Kljestan  5
Damien Perrinelle 4
Luciano Federico Acosta 4
Mauro Andrés Manotas Páez 4
Raheem Edwards   4
Rodney Wallace   4
Alexander Martin Callens Asin 3
Benjamin Angoua  3
David Accam      3
Giles Gordon Barnes 3
Ilsinho          3
.
```

Results




A player profile card for Lee Nguyen. On the left is a headshot of a man with dark hair and a goatee, wearing a black Los Angeles Galaxy jersey with the number 24. To the right of the photo is a large white circle containing the number 24. Further right, the text 'Lee Nguyen' is displayed in a large blue font, followed by 'Los Angeles Football Club' and 'Midfielder' in a smaller blue font. On the far right, a list of personal details is provided: 'Real Name: Lee Nguyen', 'HT: 5' 8" WT: 150', 'Age: 31 (10/07/1986)', and 'Birthplace: McKinney, TX'. At the bottom right is a Twitter icon followed by the handle '@LEENGUYEN24'.

Lee Nguyen
Los Angeles Football Club
Midfielder

24

Real Name: Lee Nguyen
HT: 5' 8" **WT:** 150
Age: 31 (10/07/1986)
Birthplace: McKinney, TX
[@LEENGUYEN24](#)

- Lee Nguyen attempted only 10 and completed on 8, meaning his completion rate is 80%.



Nicolás Lodeiro
Seattle Sounders FC
Midfielder

10

Real Name: Nicolás Lodeiro
HT: 5' 7" **WT:** 152
Designated Player
Age: 29 (03/21/1989)
Birthplace: Paysandú, Uruguay

- One of the other players we were able to identify.
- Was the MLS newcomer of the year for 2016.

Next Steps

- Try other zone configurations
- Incorporate other inter-zone passes made in the sequence leading up to the goal (ie. multiple assists)
 - ★ Passes made during the same possession that were significant in creating the play.
- Looking at other types of sequences that could have an impact on scoring opportunities

Conclusion

Conclusion

- We identified some players that may otherwise have gone unnoticed
- they seem to be good or underrated playmakers even though the "assist" statistic would not reflect that
- For the future, we can look at more zone split ideas and penalizations

References



Valuing Football Players' Passes by Leveraging Event Sequences

<http://www.scisports.com/news/2018/valuing-football-players-passes-by-leveraging-event-sequences>



Shot Matrix I: Shot Location and Expected Goals

<https://cartilagefreecaptain.sbnation.com/2013/11/13/5098186/shot-matrix-i-shot-location-and-expected-goals>

Thanks for Listening!

And thank you to the Vancouver Whitecaps, SFU (and affiliates), and the VanSASH Organization team for creating this event!