Sports Analytics Challenge

Evaluation/Identification the Patterns of Genepressing

Minghuan LIU

Apex Data & Knowledge Management Lab

Shanghai Jiao Tong University

minghuanliu.com

MINGHUAN LIU (ERIC)

- **Undergraduate**, Computer Science, Southwest Jiaotong University, China
- **Pre-admitted Ph.D. Student**, Apex Lab., Shanghai Jiaotong University, China

Research Interests: Reinforcement Learning, Multi-agent System, Data mining

Part-time **Assistant Coach** of Huya club for kids in Shanghai.

APEX LAB

- Data Mining (DM)
- Machine Learning (ML)
- Reinforcement Learning (RL)
- Natural Language Processing (NLP)
- Computer Vision (CV)
- Several papers in famous AI conferences (NIPS, ICML, ICLR, IJCAI, AAAI…)

Content

- 1. Definitions of Genepressing
- 2. Data and Challenges
- 3. Evaluate Genepressing
  - 3.1 Identify Ball Loss Only with Coordinates
  - 3.2 Evaluation Metrics
- 4. Identify the Pattern of Genepressing
  - 4.1 Visualization the Pattern
  - 4.2 Identify Patterns through Diagrams
- 5. Future Work
  - 5.1 How can we apply deep learning?
  - 5.1 How to use reinforcement learning?

1. Definitions of Genepressing

- Things take place within 8 seconds after the **ball loss**.
- Usually refer to that how the players of the team losing the ball **reacts** in terms of blocking the central areas/passing lines and forcing the opponent to play on the sides.

2. Data and Challenges

- Data: coordinate data (ball/players)
- Information:

Player

Player Id	Team Fixture	Match Fixture	Half	Time Value	Player X Position	Player Y Position
The unique id of each player	The team name of given item	The match name of given item	The period name of given item	The time counted with ms	The X coordinate of given item	The Y coordinate of given item

2. Data and Challenges

- Data: coordinate data (ball/players)
- Information:

Ball

Period Id	X Position	Y Position	Time Value	Home Name	Away Name	Season	Match day
The period of givn item	The X coordinate of given item	The Y coordinate of given item	The time counted with ms	The home team name.	The away team name	The season name	The match day name

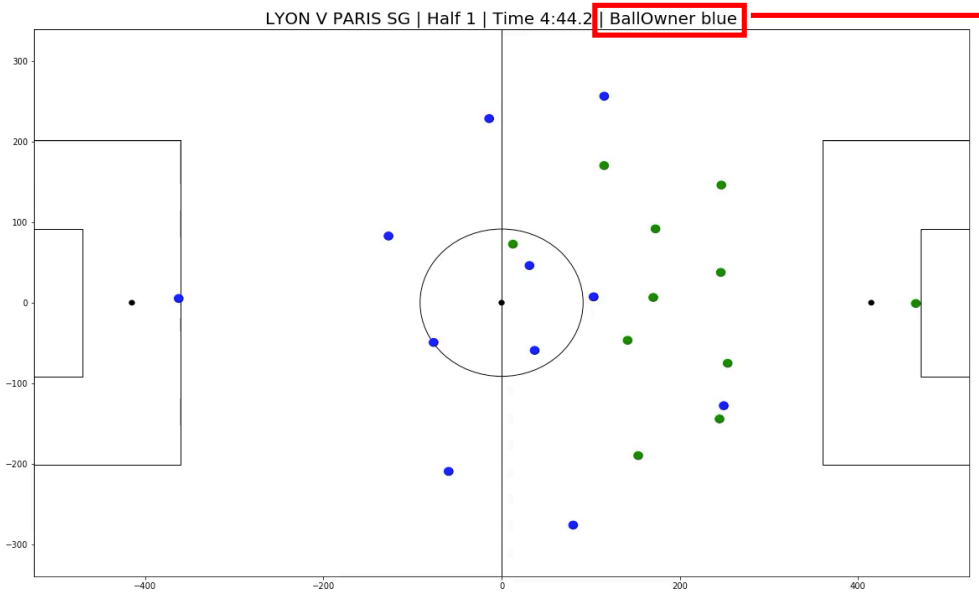
2. Data and Challenges

- Main Challenge
  - Without event data, need to identify the event by coordinates only
  - Ball and players coordinates are separate and need to be connected by hand
  - Data only contained 2D spatial information

3. Evaluate Genepressing

3.1 Identify Ball Loss Only with Coordinates

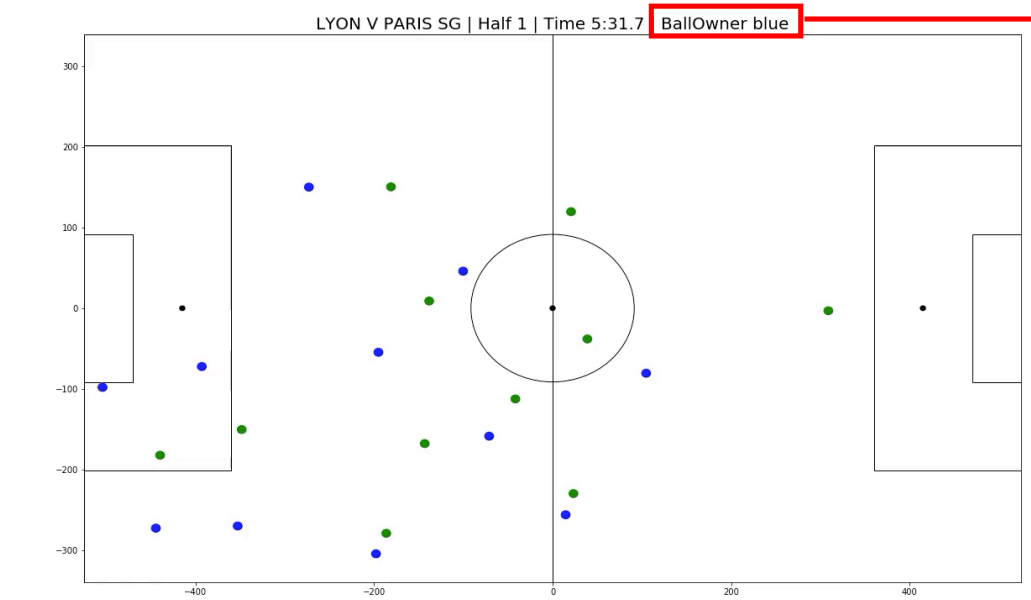
- 1. Identify the ball owner
  - Find the time that player and ball coincide and label the owner as the team of the players
  - Set the outside ball' s owner as None
  - Fill down the owner
  - Clean the noise
- 2. Clip the ball loss
  - 8s after the change of the ball owner
- 3. Extract information and result through the clipped data

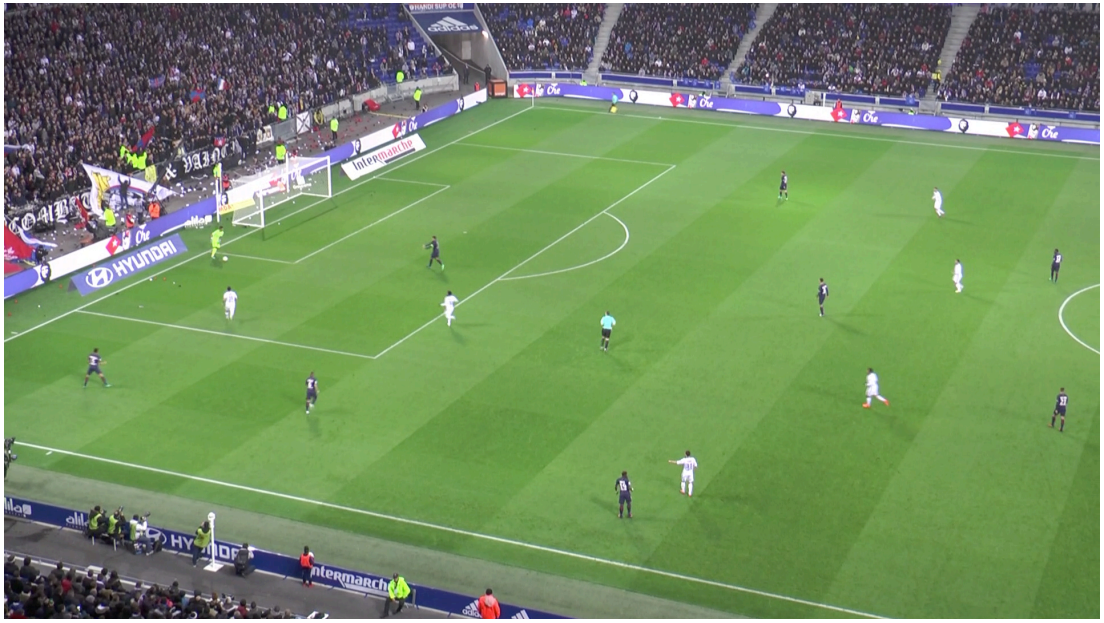


BallOwner blue

Ball loss happened!

Green is for the hometeam (Lyon)
Blue is for the awayteam (PSG)





BallOwner blue

Ball loss happened!

Green is for the hometeam (Lyon)
Blue is for the awayteam (PSG)

3. Evaluate Genepressing
3.2 Evaluation Metrics

- All metrics are **for the team who lost the ball** at the beginning of the 8s clip and it evaluates the **opponents**' behaviors

Average Moving Orientation

Average Pushing Distance

Average Pushing Velocity

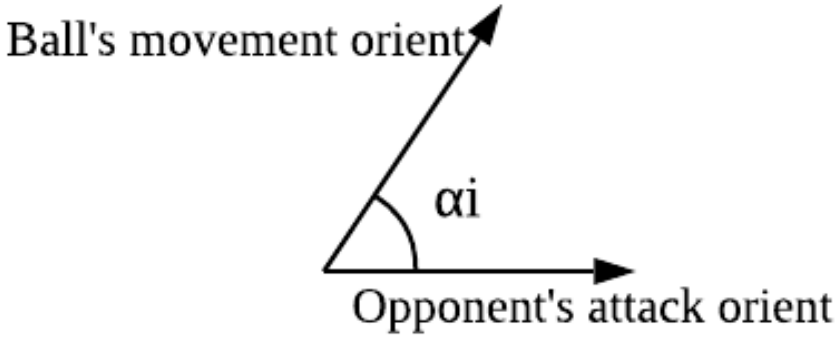
Average Distance from Threaten Zone

Average Minimum Distance from the Side

3.2 Evaluation Metrics – Average Moving Orientation

$$f_{amo} = \frac{1}{2}\left(\cos\left(\frac{1}{n}\sum_{i}^{n}\alpha_i\right) + 1\right)$$

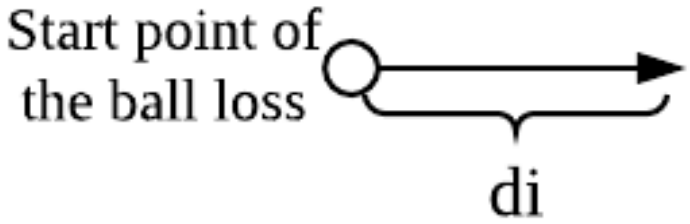
$\boldsymbol{n}$ *is the number of items*



3.2 Evaluation Metrics – Average Pushing Distance

$$f_{pd} = 1 - \max(0, \frac{\frac{1}{n}\sum_i^n d_i}{maxpd})$$

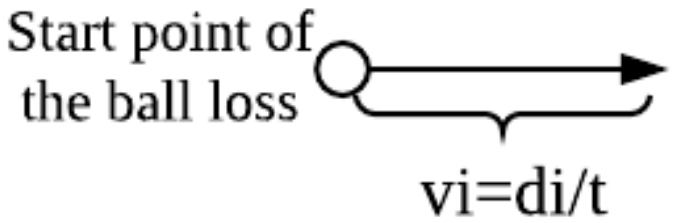
maxpd *is the max pushing distance seted by hand*



3.2 Evaluation Metrics – Average Pushing Velocity

$$f_{apv} = 1 - \max(0, \frac{\frac{1}{n}\sum_i^n v_i}{maxapv})$$

maxapv *is the max average pushing velocity seted by hand*

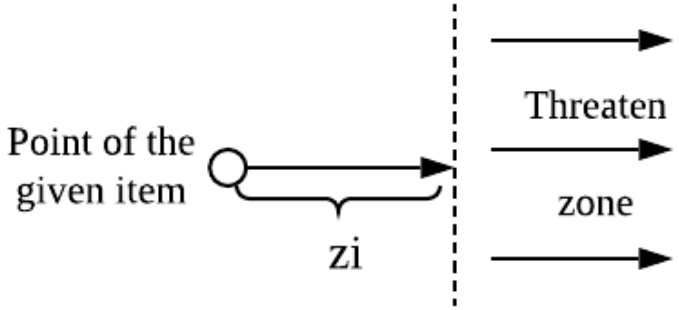


3.2 Evaluation Metrics
– Average Distance from Threaten Zone

$$f_{adtz} = \frac{-clip\left(\frac{1}{n}\sum_i^n z_i\right)}{maxadtz}$$

$\boldsymbol{maxadtz}$ *is the max average distance from treaten zone that we think it should be safe*

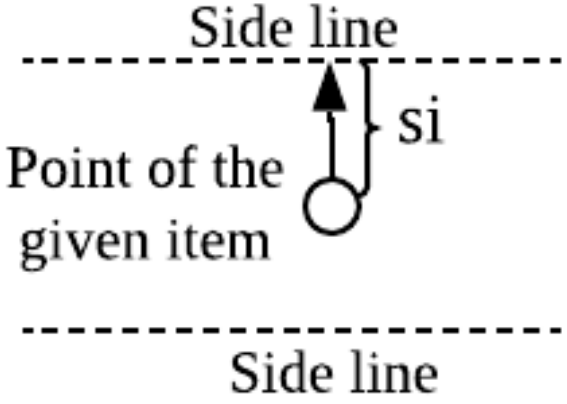
$\boldsymbol{clip(x)}$ *is a function that clip the value of* x *into* $[-maxadtz, 0]$



3.2 Evaluation Metrics
– Average Minimum Distance from the Side

$$f_{mds} = 1 - \frac{\frac{1}{n}\sum_{i}^{n} s_i}{maxmds}$$

$\boldsymbol{maxmds}$ *is the max distance from the side that we think it should be safe*



3.2 Evaluation Metrics – Total Score

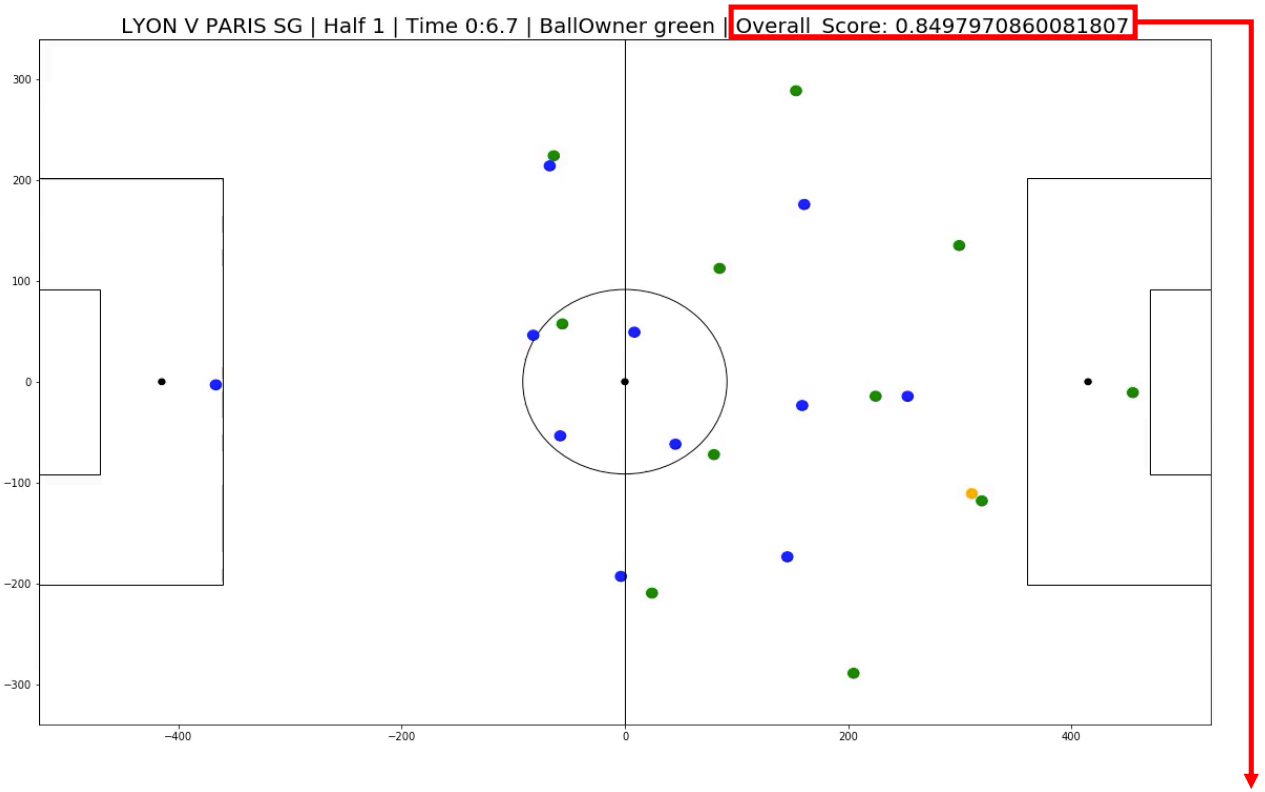
- Linear Weighted Score

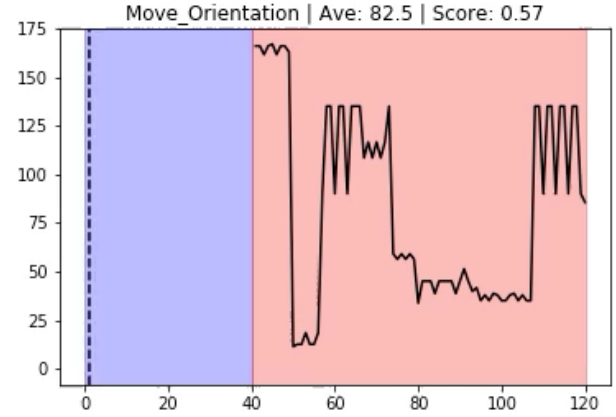
$$f_{total} = w_{amo} * f_{amo} + w_{pd} * f_{pd} + w_{apv} * f_{apv} + w_{adtz} * f_{zdtz} + w_{mds} * f_{mds}$$

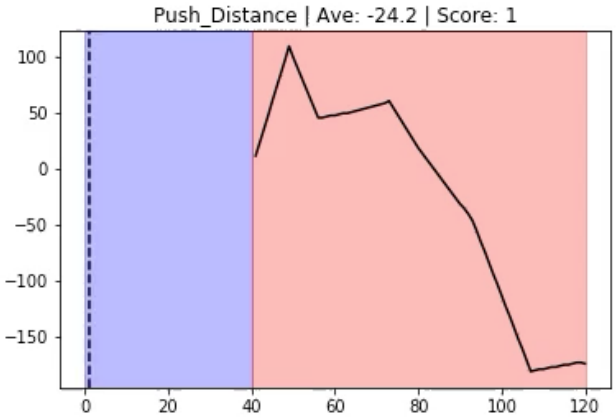
Each score can be used as a label to do train meaningful models.

- Because of short time I can only make it a future work

3.2 Evaluation Metrics – Example

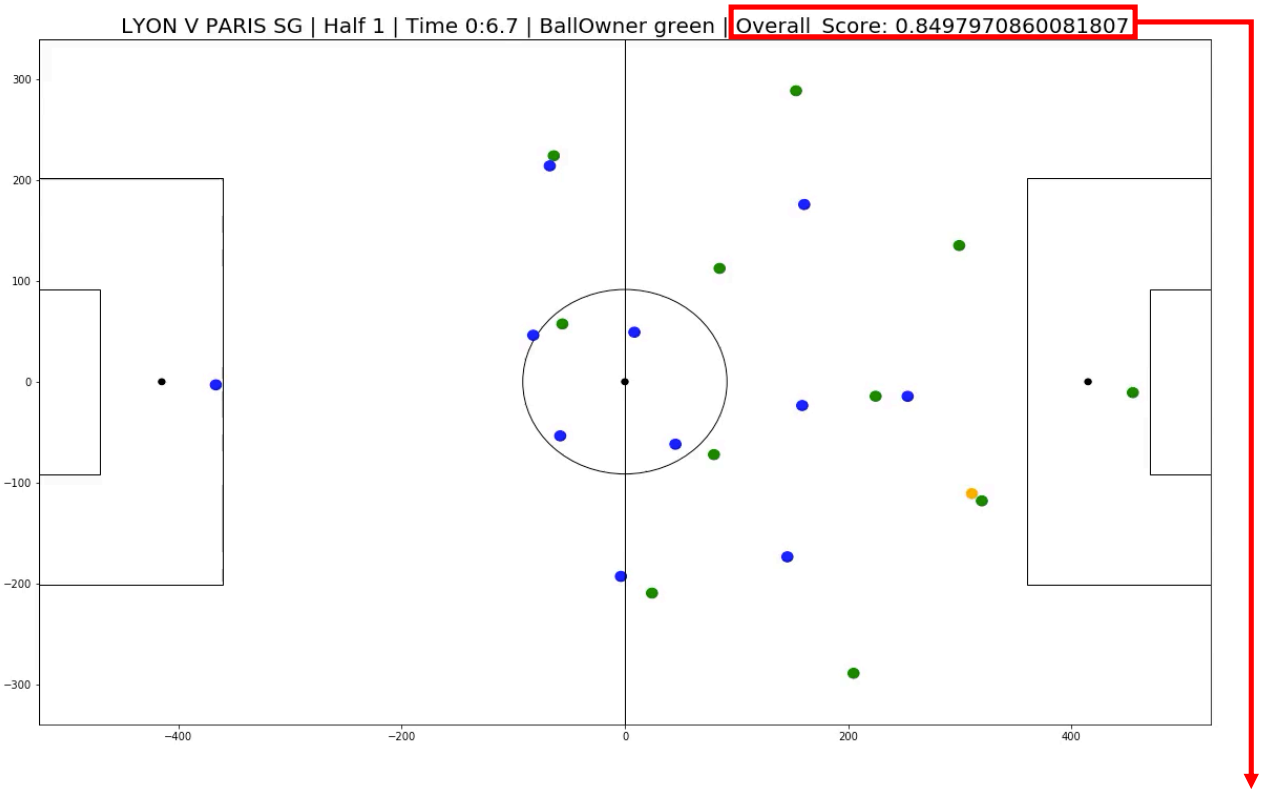


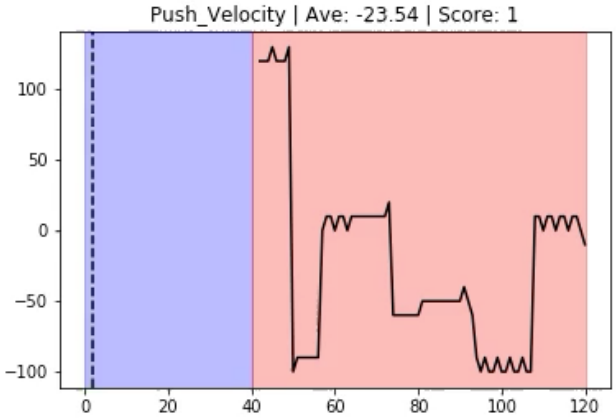


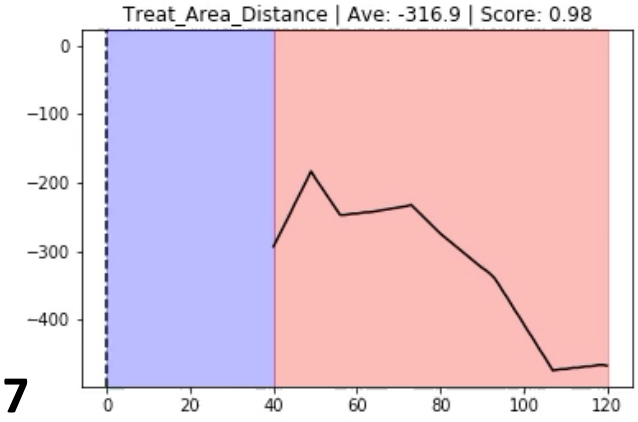


Overall Score: 0.849797

3.2 Evaluation Metrics – Example

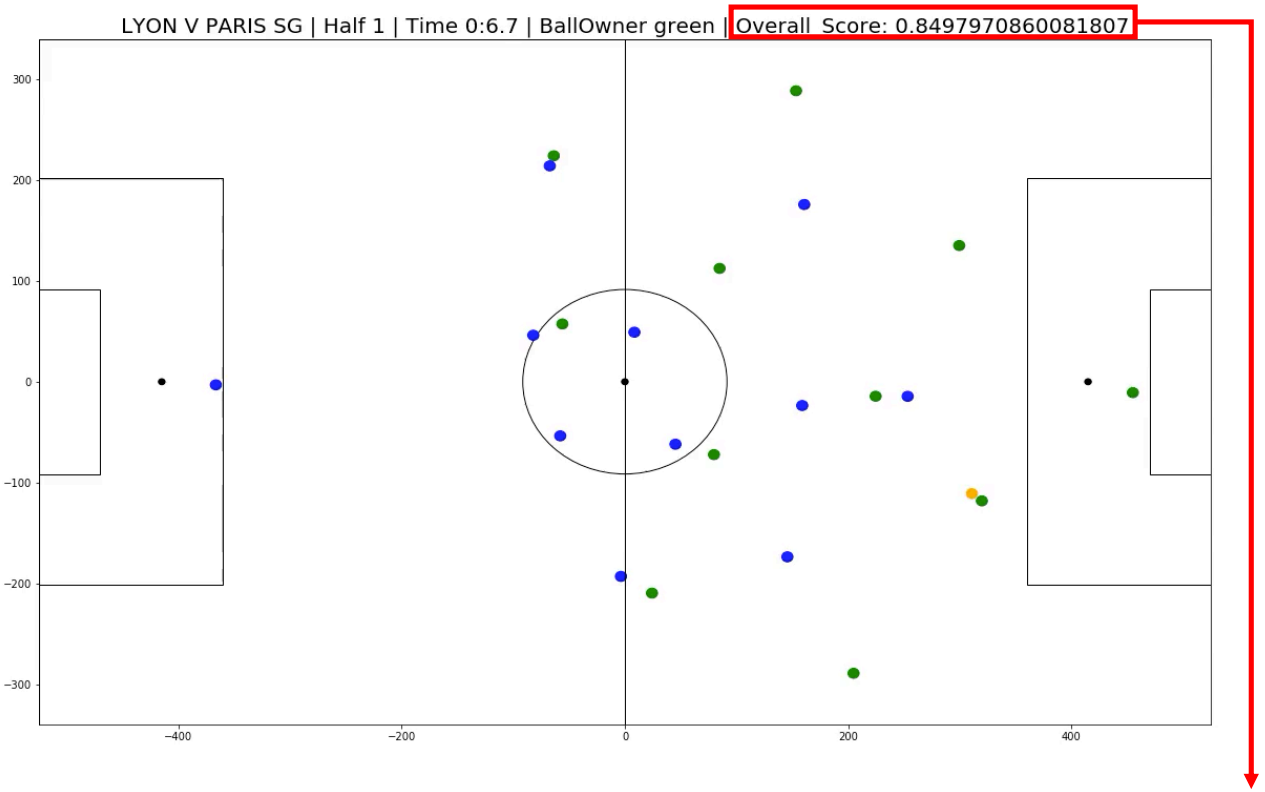


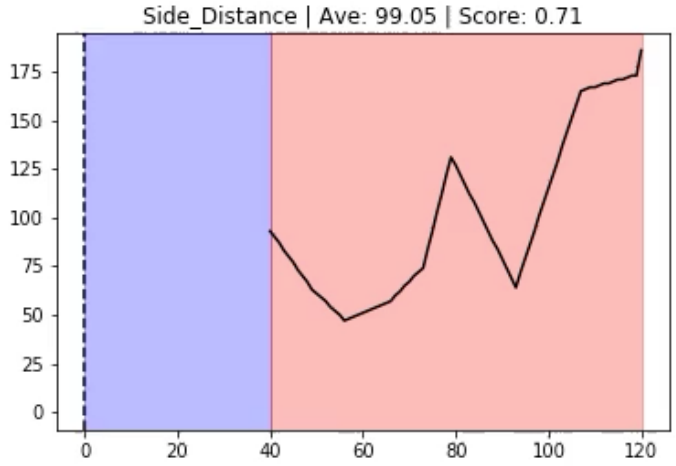




Overall Score: 0.849797

3.2 Evaluation Metrics – Example

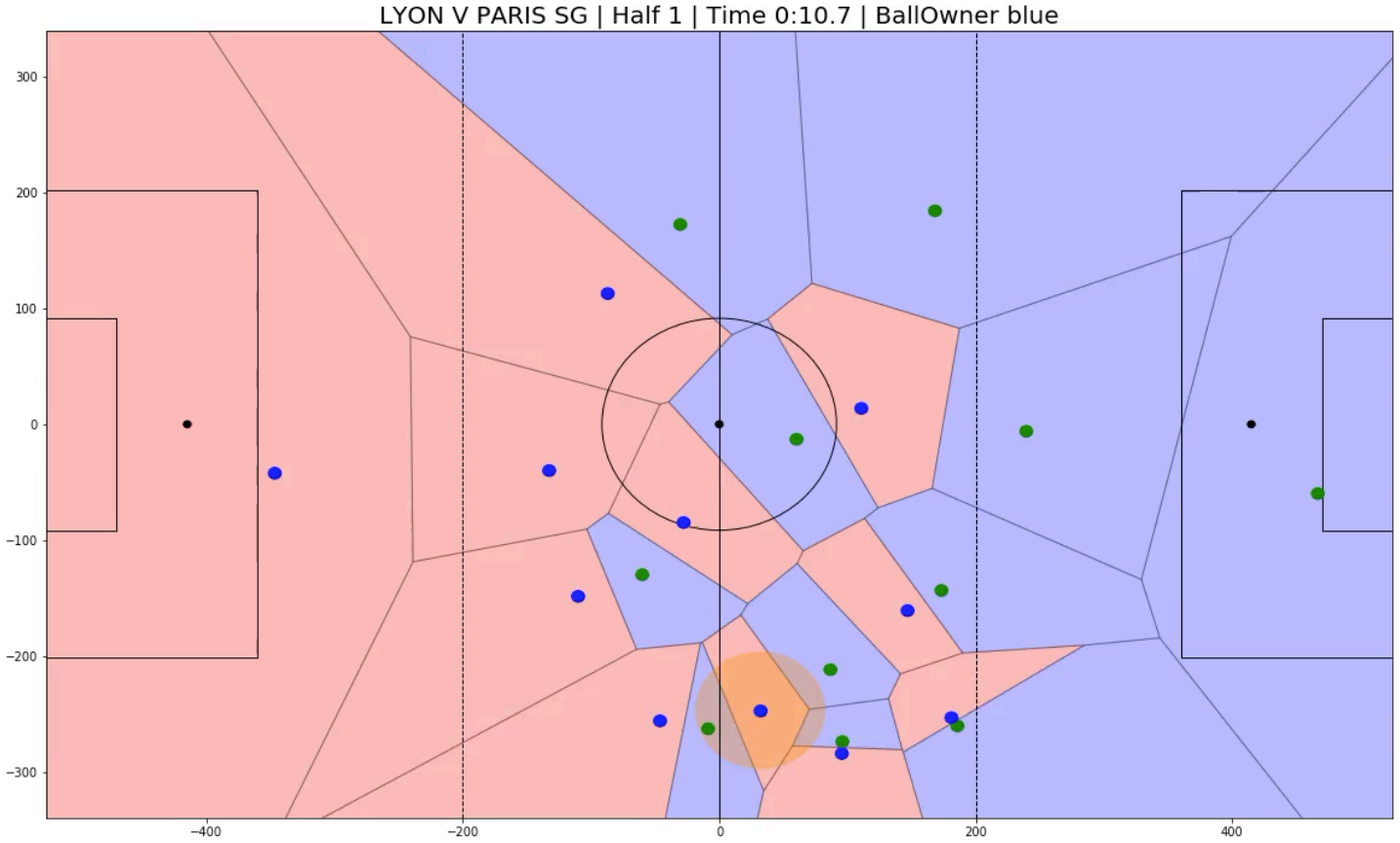




Overall Score: 0.849797

4. Identify the Pattern of Genepresing

4.1 Visualize the Pattern – Example



Voronoi Diagram

Space:

Red is for ball keeper

Blue is for defender

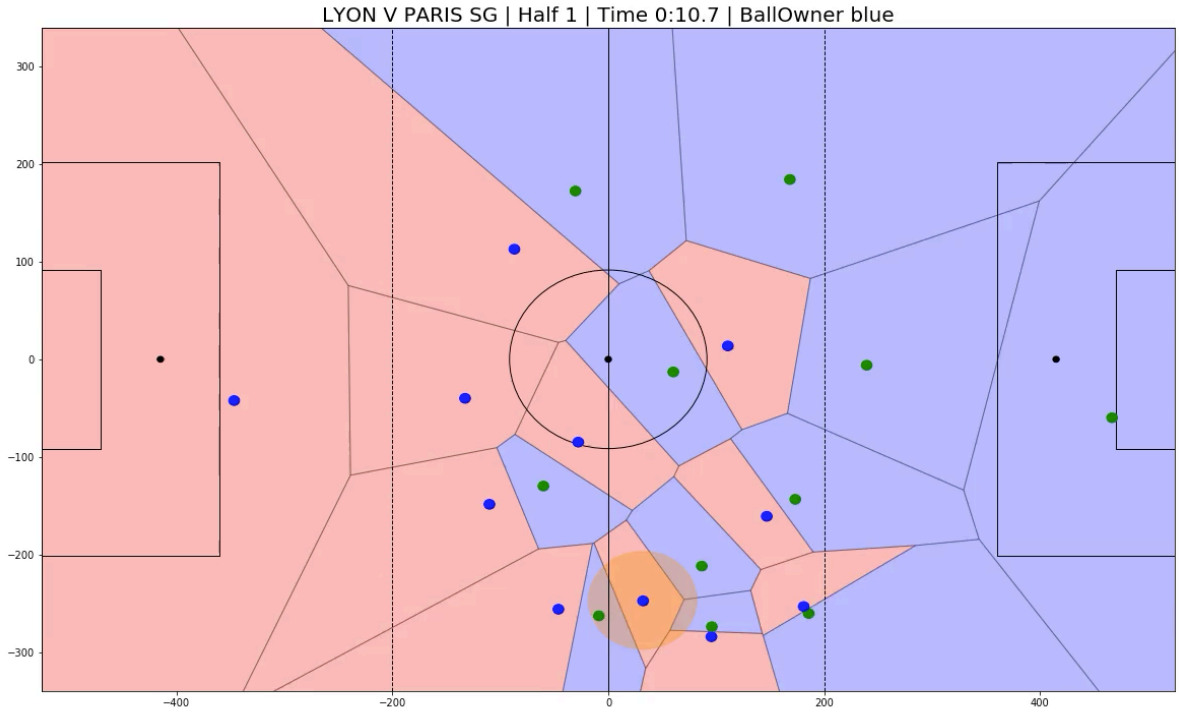
4. Identify the Pattern of Genepresing

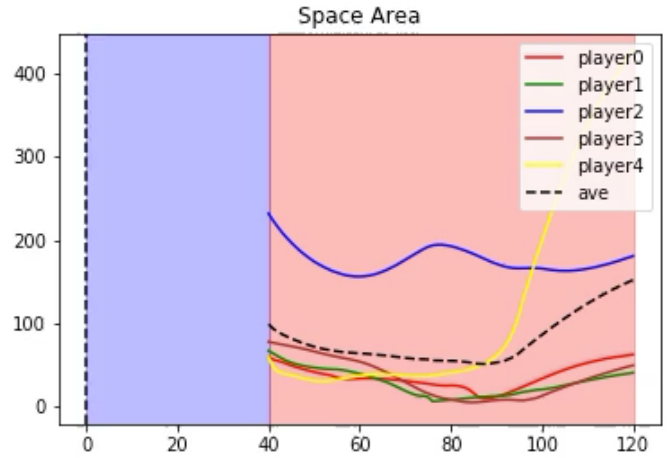
4.2 Identify Patterns through Diagrams

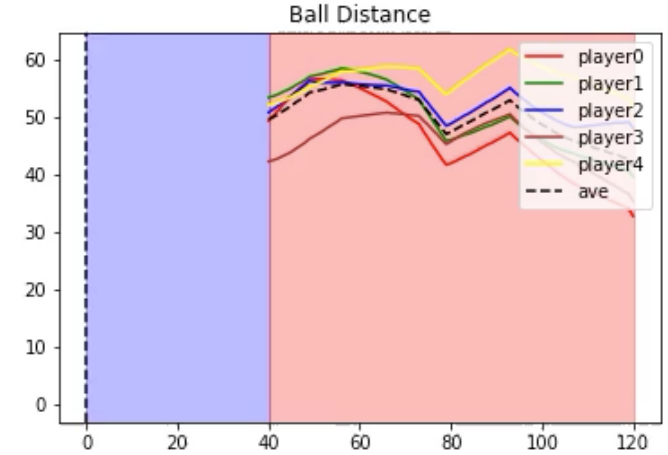
- How the spaces of the top 5 nearest ball opponent players changes
- How the distances from the ball of the top 5 nearest ball opponent players changes
- Patterns of genepressing can be seen from the diagrams and their statistics

4. Identify the Pattern of Genepresing

4.2 Identify Patterns through Diagrams

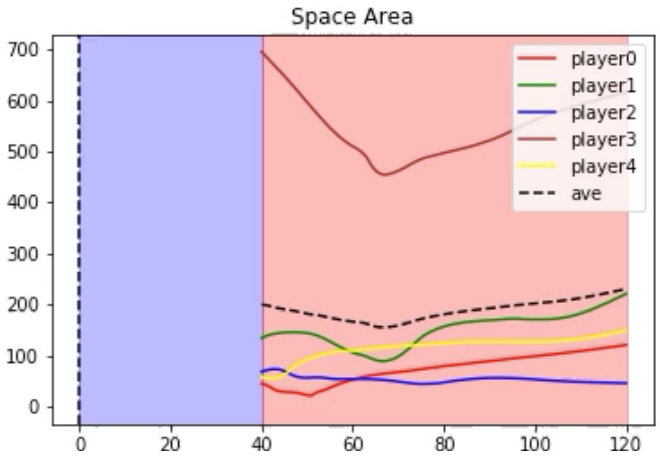


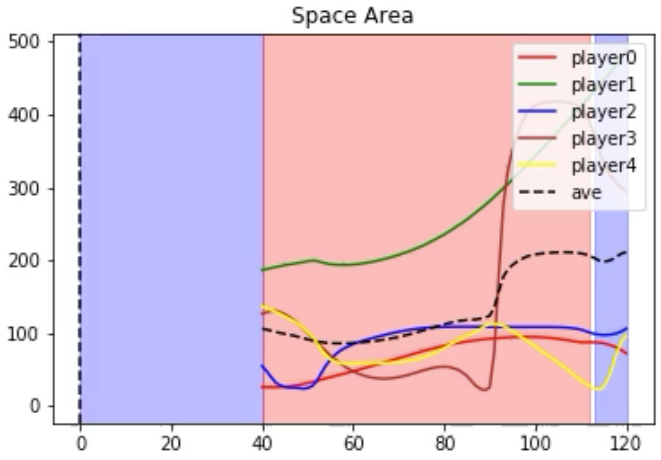


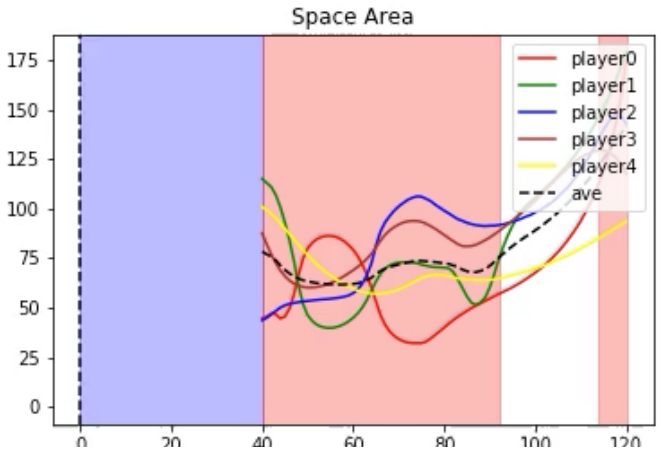


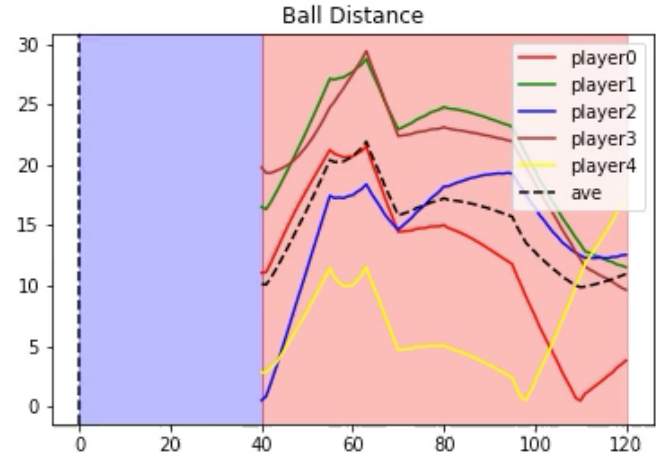
4. Identify the Pattern of Genepresing

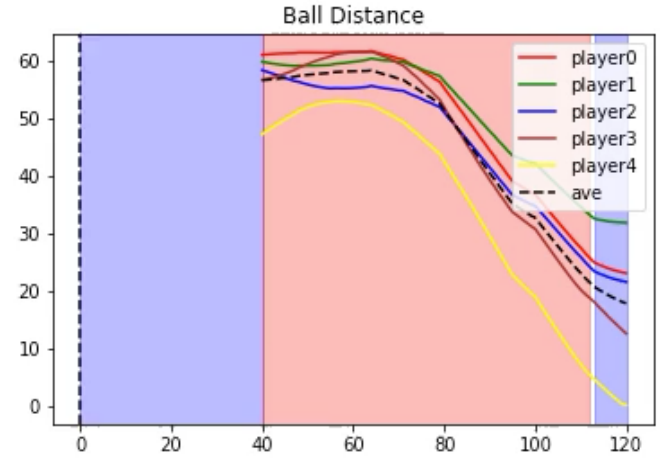
4.2 Identify Patterns through Diagrams

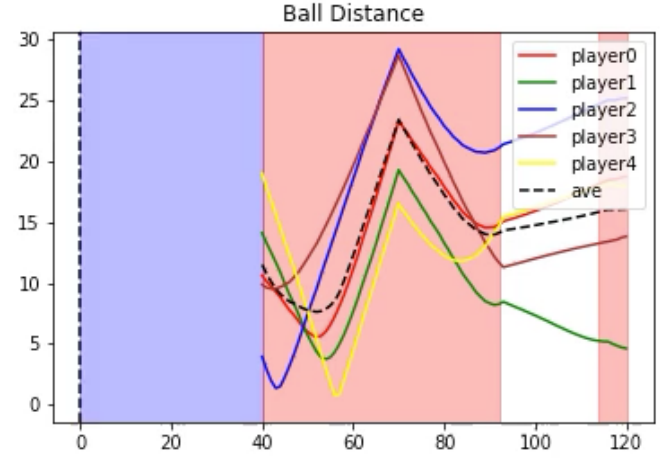








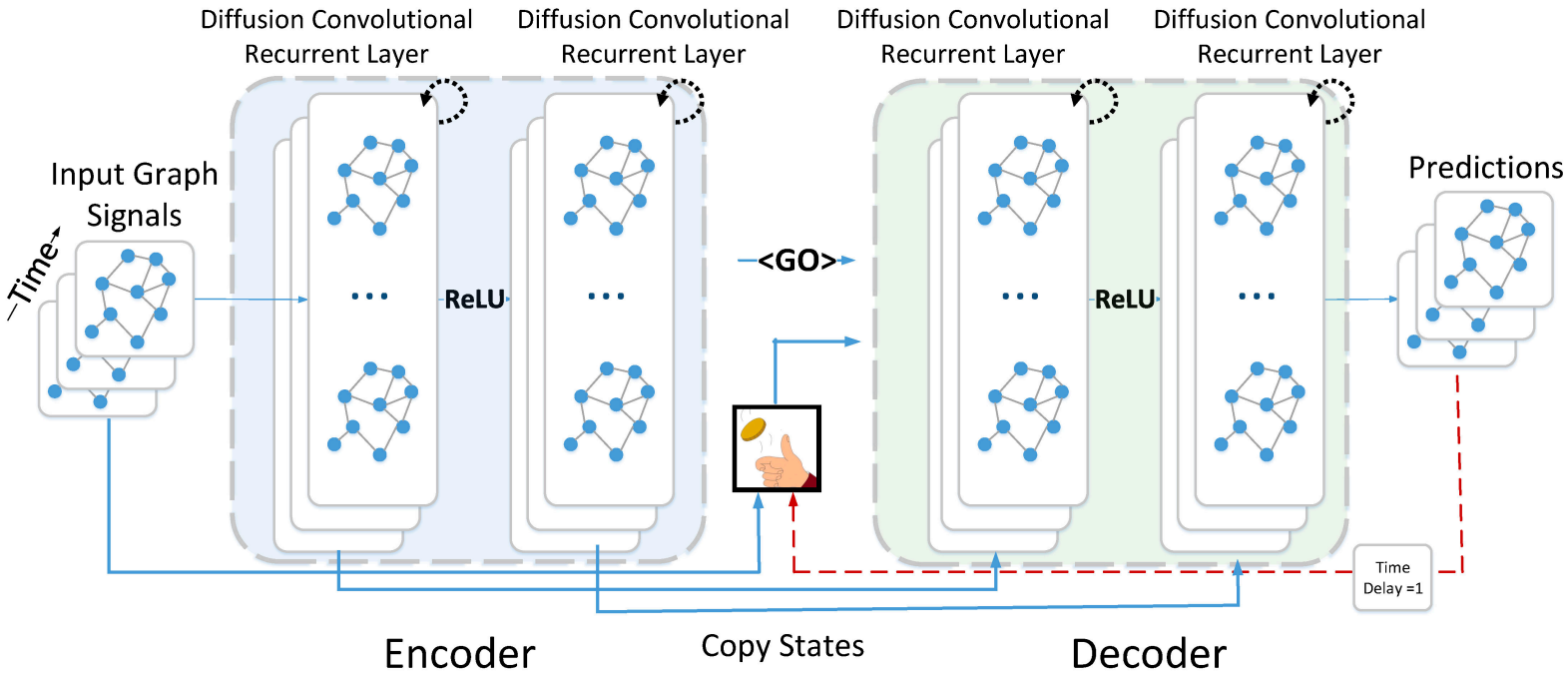




5. Future Work

5.1 How can we apply Deep Learning?

- Model multi-trajectory temporal spatial data -> Refer to Urban Computing
- GNN for the player + RNN for the time -> Learn a pattern embedding

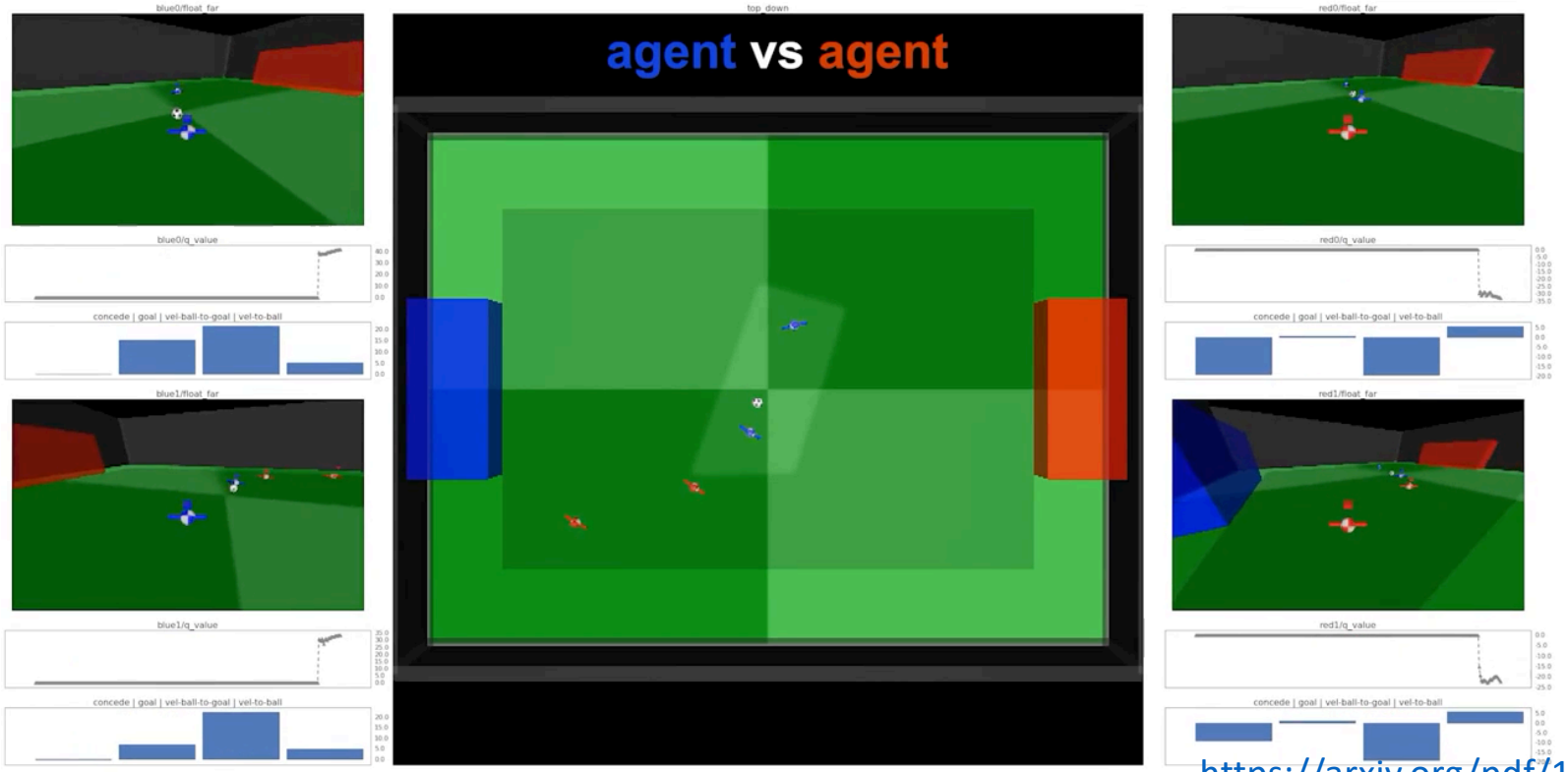


5. Future Work

5.2 How to use Reinforcement Learning?

- We can make agents to learn different policy of genepressing

Example 1:<EMERGENT COORDINATION THROUGH COMPETITION> - Deepmind



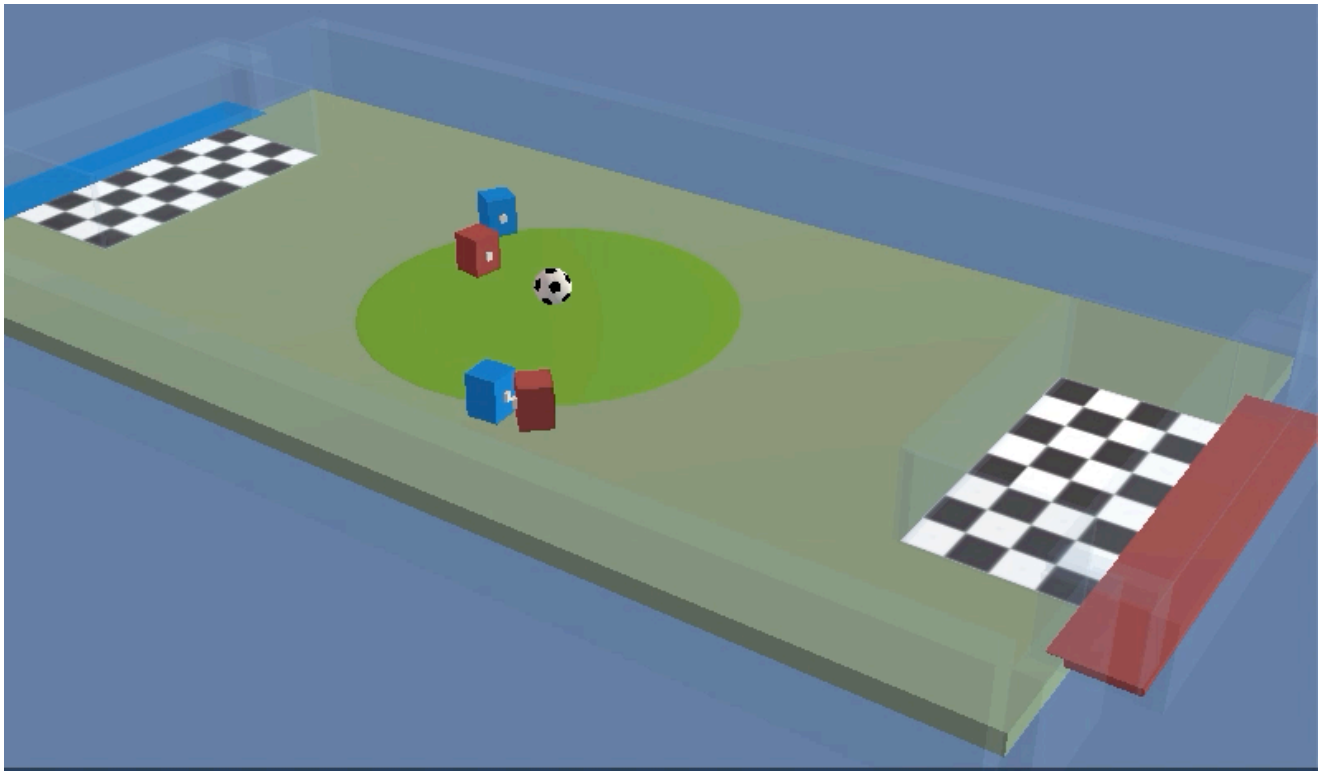
https://arxiv.org/pdf/1902.07151.pdf

5. Future Work

5.2 How to use Reinforcement Learning?

- We can make agents to learn different policy of genepressing

Example 2: Use the communication between players to learn policies - Ours



FM? FIFA?


Thanks for listening!
I dream to use AI for football, and make football more intelligent!

Minghuan LIU

Apex Data & Knowledge Management Lab

Shanghai Jiao Tong University


minghuanliu.com