Game States in Soccer

Liam Creedon and David DeScherer

Mentored by Anton Dahbura

Johns Hopkins University | Whiting School of Engineering | Baltimore, MD

Design Day 2022

JOHNS HOPKINS
WHITING SCHOOL
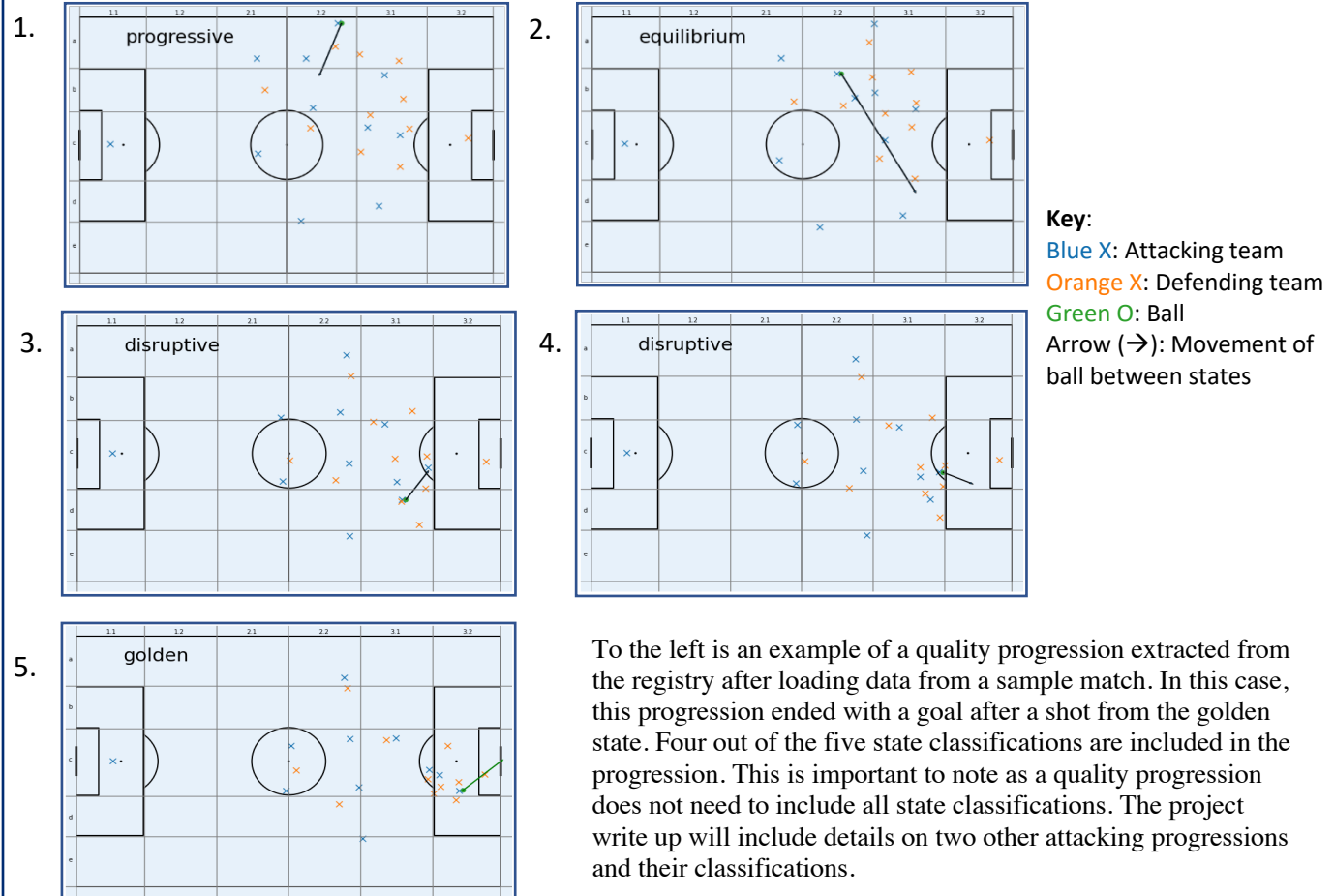of ENGINEERING

Overview and Objectives

This project takes event and tracking datasets from soccer matches, and from the data points, captures comprehensive and discrete game states through an extract-transform-load (ETL) pipeline. From here, database will be populated with a collection of json items representing states and events, maintaining events as connectors between states, establishing a graph-like relational hierarchy. This registry can be queried to explore different sequences of play, helping teams identify progressions that will exploit weaknesses in the defense, allowing the team to find threatening attacking opportunities.

Materials and Methods

- Dataset – Metrica Sports Sample Data
- Game States Github
  - Data model comprised of:
    - State.py
    - Event.py
    - Player.py
    - Grid.py
    - StateClassifier.py
    - DBConnector.py
  - Jupyter notebook ingestion scripts
    - kloppy python package
- Data dictionary
- Game States Registry – Azure Cosmos DB

Data is read into the data model using events and tracking data from samples provided by Metrica. States are taken from the time frames at the start and end of a given event. The StateClassifier gives each state a label based on the technical state definitions, which are then loaded into the database along with the event connecting the states.

Results

1.

progressive

2.

equilibrium

Key:
Blue X: Attacking team
Orange X: Defending team
Green O: Ball
Arrow (→): Movement of ball between states

3.

disruptive

4.

disruptive

5.

golden

To the left is an example of a quality progression extracted from the registry after loading data from a sample match. In this case, this progression ended with a goal after a shot from the golden state. Four out of the five state classifications are included in the progression. This is important to note as a quality progression does not need to include all state classifications. The project write up will include details on two other attacking progressions and their classifications.

Next Steps and Extensions

The primary focus for the project in the future will be applying analytics to the data that has been gathered. The analytics will primarily focus on evaluating the quality of progressions of state. Definitions of key metrics and quality of a final state and a progression are defined in the paper. Examples of possible metrics to be used include xG (expected goals) and xT (expected Threat), but since analytics ins soccer is an ever-evolving field, there is a possibility that our group can determine new ways to analyze the quality of a possession.