

# The Analysis of Digitalized Soccer Game Recordings

Haruki Ambai s1270128

Supervised by Prof. Maxim Mozgovoy

## Abstract

The purpose of this study is to implement the data analysis tool of digitalized soccer game recordings. Data analysis is needed for sports analysts, AI researchers, and coaches, and there are datasets available. However, few tools can process and visualize them. Therefore, we have successfully implemented a data analysis tool to improve the effectiveness of the analysis. Finally, we successfully have attached new features and find a lot of profits to analyze them in the future. On the other hand, there are a few issues to fix for the more effective analyzing so we need to consider it carefully.

## 1 Introduction

In these days, data analysis, especially on soccer, has high demands for sports analysts, AI researchers, and coaches. One of the main reasons for this high demand is although there are a lot of available datasets, which is type of spatiotemporal data, but there are few analysis tools that process and visualize tools. Spatiotemporal data is a kind of the data that contains dynamic updates of spatial locations and/or extents along with time. A typical example of spatiotemporal data is a moving object (e. g., a car, an aircraft, or a pedestrian) whose location continuously changes. The providers of these soccer datasets, such as [StatsPerform.com](https://www.statsperform.com), [DataStadium.co.jp](https://www.datastadium.co.jp), and [Chyronhego.com](https://www.chyronhego.com) are known well. Therefore, we are working on implementing a data analysis tool (Figure 1) for digitalized soccer game recordings.

This tool is composed of main panel, filter radio buttons and step forward/backward buttons and slider for all frames. The main panel is the biggest part of this GUI contains UTC time and event type and jersey number and team side of the current frame's event on top. Under the information, there is the game field which all players on it with their jersey number. In the right side, several kinds of filter radio buttons for events and teams and movements. The events and teams filter can be selected multiple, but movements filter must be selected only one. The play forward/backward buttons can skip to specific or filtered frames by clicking it. Finally, slider, at the bottom of GUI, can move here to there so it is not depends on events or other filter. We can easily control the frame one by one in mouse and left/right button on keyboard.

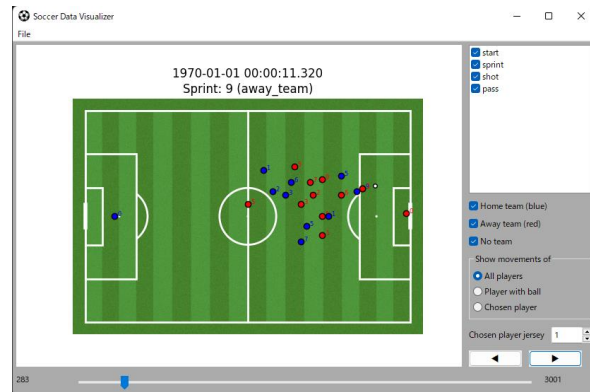


Figure 1. Our soccer data analysis tool

The data converter and viewer are composed of this tool. (Figure 2)

The data converter processes three types of dataset formats Data Stadium, STATS, and Google Football files [1] (plain text or zip-compressed) into compressed Signality files [2]. The Signality file is usually a collection of frames without event information and Google Football data format is mainly used for soccer AI research, and it needs event data. Then events can be reconstructed if they are missing. In addition, this consist of independent "channels". There is no apparent requirement for which channels must be present and how they are stored, so we will presume that the output Signality file is a zip archive with two-channel files, the track and event channel. The track channel contains mainly tracking objects in soccer games such as a ball, all players are divided into home and away teams. For example, the ball has the position (x, y, z coordinates) and possess team flags and the possessing player's jersey number. The event channel contains all event information such as passes, shots, and sprints. For example, passes have the UTC time of start and end and have player information from and to a player. Especially, we consider the Google Football data set because there are different data providers and we have to read all formats by each user's demands. In this data, we need to add more markup like pass recognition. There are methods [3] to achieve that, and we integrate them into our own data analysis system.

The data viewer read the Signality file generated by converter, so this part does not depend on file formats, just aimed for Signality format. Filtering it followed by given conditions from users. For example, if you would like to watch passes and shots event of all home team players have done, remove the radio checks without

two events name and home team. This feature enables us to smooth frame skipping.

In this paper, we describe the following four new features.

1. Adding player direction
2. Passes, shots, movements lines
3. PWB (Player with Ball) filter
4. Frame playback/pause buttons

These are new features of soccer data analysis tools. (1) is adding players' direction to the head of the player. Google Football format already has direction elements, but the rest of the formats are not. Then, we need to solve this problem. (2) means these events are shown by one line while events are happening. Movements are the player's data element such as sprint, highspeed run, medium-speed run, acceleration, and deceleration. These elements are extracted from each file format. (3) means chasing the players to possess the ball. This is currently not working well though it should be working so we must fix it. (4) is about adding new frame playback/pause buttons in the GUI. Frame forward/back buttons are already contained in this tool. However, if we want to skip to 30 steps forward, we need to click 30 times in that situation, and it is quite hard. Then, we implemented play/back button which is skipping in specific interval such as 0.1 seconds and pause in desirable frame. We describe the way of implementing one by one and explain the logic of them.

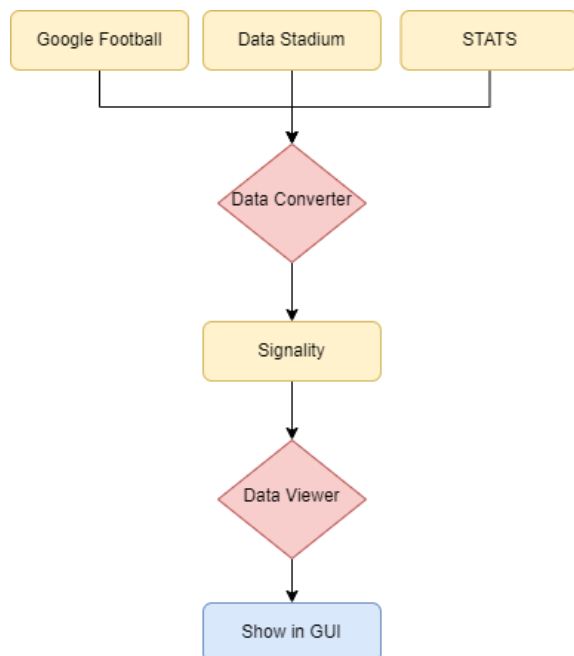


Figure 2. The flow of data processing

## 2 Background

There are different data providers, and we need a tool that can read all these formats and supply the information depending on each user's demands. Detecting the events of a soccer game is one of the quite hard tasks in soccer analytics. Especially, the spatiotemporal dataset may only include time and player tracking information. This data needs some additional markup like pass recognition, etc., and there are methods to achieve that [4], and we integrate them into our data analysis system. We have recognized the pass event from these data by implementing the rule-based algorithm.

Visualizing them is also an important aspect of soccer analysis on the spatiotemporal dataset. In the limited game information, the visualized lines of the passes or the player's movement are intuitive and provide many profits to sports analysts. For example, if you would like to immediately understand which team had made much more opportunities for goals in the game by watching the steps of the events of the shot, text information is not enough to satisfy this. However, the combination of text and visualized information makes it much easier and faster to recognize which team has made this decision.

Tracing the players who have the ball in the current frame and the next frame is so desirable for analysts. This feature enables us to concentrate on the current player's event and it would provide us with a different view of data analysis. In addition, smooth steps of each frame are required in this or other complicated scenarios. Therefore, we decided to implement a filter that only visualizes the player with the ball and the play/pause button for smooth steps of events.

## 3 Data Types

We have experimented with three types of datasets, Data Stadium, STATS, and Google Football. All the Data Stadium and STAS are the spatiotemporal. We will describe them in detail.

**Data Stadium:** This provides sports data for teams, players, and fans. These recordings accurately capture the course of the game, and even significant pauses within the game are not removed.

**STATS:** This is distributed as a small set of .pkl files, each containing several independent game sequences. Thus, we extract these sequences into individual pickled files before they can be used with DataConverter.

**Google Football:** This is the output of the Google Research Football game. The whole game information player tracking and event data elements are contained in detail.

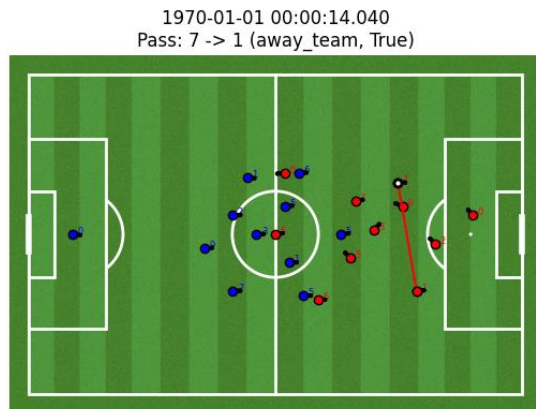


Figure 3. The pass of away team can be seen.

We convert them into the Signality format[1] to make visualizing easier.

**Signality:** In general, they provide the service of Soccer API which extracted tracking and event data from static videos took by utilizing existing camera setup. The reason why we accept this format is that this is flexible, extensible (based on JSON) and contains more important information. For example, in DS/STATS there is even no information about discretization frequency such as milliseconds pass between two subsequent frames. In this article, latter situation is the main subject.

## 4 Method

### 4.1 Data Converter

The data converter process three data format to Signality format so we must edit this file to add new type of element such as direction. In current task, we must add new direction element but only Google Football files contain that so we must create a direction element by calculate between current and next frame.

### 4.2 Data Viewer

All new features should edit this part of tool because we would like to test the movement of players and events visually.

#### 4.2.1 Adding Player Direction

In 2.1, we have created new direction element in processed data. Then, set the black dot to the player's direction of movement using that new element.

#### 4.2.2 Passes, Shots, and Movements Lines

These events name is already shown in the top of the GUI. However, it is difficult to recognize in a moment. Then, we added lines while these events are happening by the team color blue or red.

Events lines such as passes are shown when it happens. The event data of passes contains the

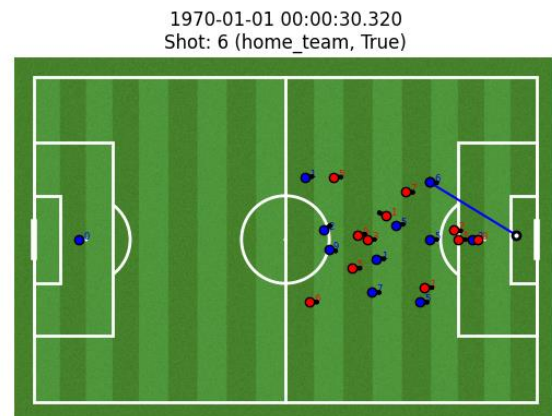


Figure 4. The shot of home team can be seen.

sender and receiver jersey numbers so we could detect the receiver's coordinates. Shots only have an event starting coordinate so the destination of the shot line would be the gall keeper's position.

Specifically, movement lines should be drawn when it starts and end. When movements are happens, tool start to find the ending UTC time of this movement. Then, use it as a key to find that player's position at that time. Finally, draw a line from the current position to the future position where it ends.

#### 4.2.3 PWB Filter

This is the filter for player to extract the only player with the ball (PWB) and next player like passed or took ball from the opponent player of current PWB.

Currently, our approach is finding the jersey number who's concerned of events and pass that number to the main view part. However, it is not working like that way. Then, we changed approach from this to finding nearest player to the ball.

In this method, first, we recognize current PWB in all frames by calculating the nearest player for

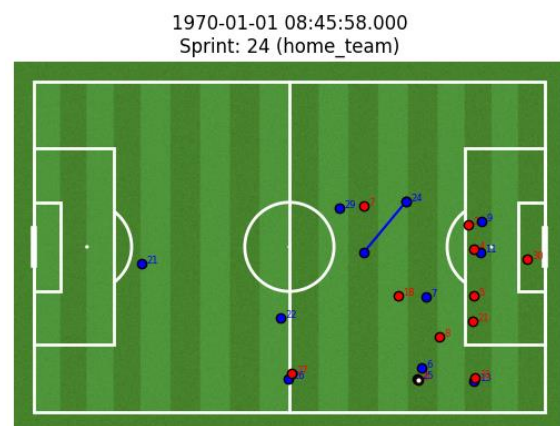
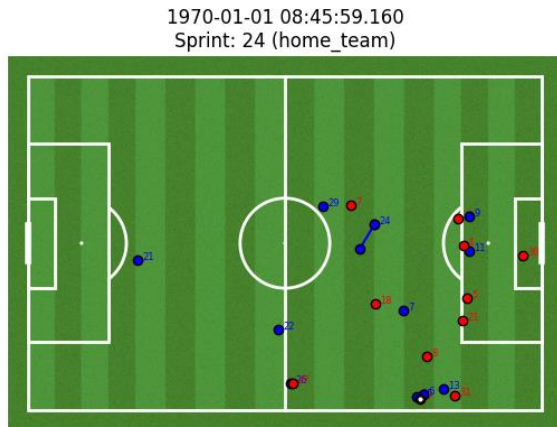


Figure 5(a). The beginning of sprint event is drawn.



**Figure 5(b).** The middle of sprint event is drawn, both of away and home teams. And then, set current PWB in the frame. These are shown in top middle of the GUI.

#### 4.2.4 Frame Playback/Pause Buttons

Play back button should work as smoothly so we set the time interval reference variable attribute in the code as class element so that it can be accessed anywhere in the class. Thanks for this, we could implement start and stop button in short code. The play icon would turn into pause icon when it is pressed. Again, pressing it returns the icon to the play icon.

## 5 Results

All tasks have been successfully done and made specific events clear to analyze soccer game data recording.

### 5.1.1 Adding Player Direction and Passes, Shots, Movements Lines

In the direction and the line of passes and shots are shown in Figure 3, 4. They are shown individually when it happens in the most of cases well.

The movements line would be kept drawing among the frames those happened. (Figure. 5(a), 5(b)) We can see the whole movements event on moving the slide bar between start and end frames. This would not be seen by clicking event skip forward button because that is the specification of that button ignoring long specific term event such as sprint.

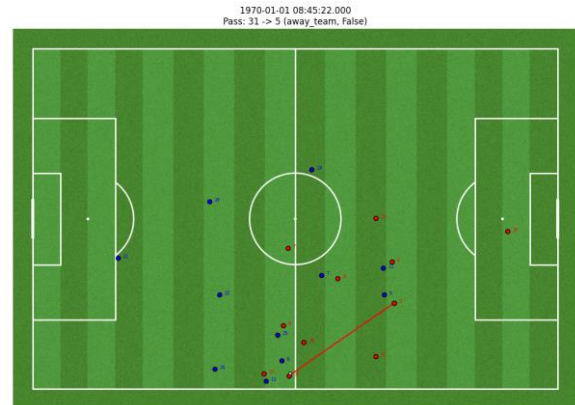
### 5.1.2 PWB Filter

This works well and filter the player's event who have ball in the current frame (Figure 6.(a), 6.(b)). It also works well when PWB have lost ball and turned into the ball possession to opponent team. For example, when away team took the ball

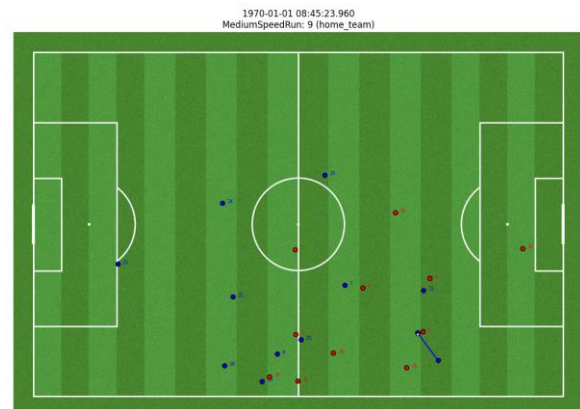
from current PWB in home team, that away team player would be new PWB in current frame.

### 5.1.3 Frame Playback/Pause Buttons

We have added button between the skip forward and backward button which controls playing and pausing of event frames. Play icon turns into pause icon when it is pressed and again, it back turns into play icon. The threshold is the 1 seconds/frame. This makes understanding the game flow and analysis easier and clearer. Especially, with the



**Figure 6(a).** This is the frame of the away team is starting of the passing



**Figure 6(b).** The Home team took the ball from the away team and PWB is changed can be seen.

function in 3.1.2, it provide quite powerful assistance for follow the route of ball in the game and also events happened with that ball.

### 5.1.4 Possible errors

Due to the following errors of dataset, there are some limitations.

1. The team flag of the ball is always null
2. The same jersey number in the same team.
3. A player passing to himself.
4. Coordinated are in integers, it is desirable to set it as float



1 happens in some cases then it makes showing which team possess ball in the frame. 2 causes to fail player recognition. Other similar one 3 causes crash the tool might be related to the lack of the dataset. 4 is not errors but it is important to make accuracy of the data analysis higher. Some of these are responsible for data providers might be in initial setting problem but not all of them. Therefore, we need to research more deeply on these whether we are able to fix these or it is responsible for provider in the future.

## 6 Discussion

These results of implementation made analyzing soccer games more effective and intuitive. The visualization of several critical events to analyze such as passes, shots, and movements accelerated the intuitive analysis of a soccer game. In addition, the combination of the PWB filter and playing frames in several intervals expands the possibility of understanding the whole game's flow.

Although these features are powerful updates for one frame, but not for analyzing the total result status. For example, the total number of shots in the whole game from one team and how many shots are reached to the goalkeeper and scored provide us with several chances to analyze based on the total result status.

In this case, we need to create a control panel that shows stats on several events. For example, a pie chart for the number of scored shots on total shots. We can also provide the same format chart for other types of critical events. We can analyze the relationship or correlation between the number of critical events that could make more opportunities for scores and the rate of winning. One issue came up with the placement of the chat, which is a serious issue for making the part of the soccer field visualization narrower. To solve this problem, separate the two pages between the game event visualization page and the total status of the whole game results.

## 7 Conclusion

We have improved our soccer data analysis tool by adding four new features. The features in the following list are working well and made experience of analyzing more effective and useful:

- Adding player direction.
- Passes, shots, movements lines.
- PWB (Player with Ball) filter.
- Frame playback/pause buttons.

However, there are some limitations in this tool which causes crash this tool and should be fixed in the future such as:

- The same jersey number in the same team.
- A player passing to himself.

## References

- [1] K. Kurach, A. Raichuk, P. Stańczyk, M. Zając, O. Bachem, L. Espeholt, C. Riquelme, D. Vincent, M. Michalski, O. Bousquet, S. Gelly "Google Research Football: A Novel Reinforcement Learning Environment", *AAAI*, vol. 34, no. 04, pp. 4501-4510, Apr. 2020.
- [2] Signality, "Data stream format for soccer provided by Signality tracking system", <https://help.signality.com/livedata/>
- [3] G. Mola Bogdan and M. Mozgovoy, "Towards Case-Based Reasoning with k-d Trees for a Computer Game of Soccer," 2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS), Shenyang, China, 2019, pp. 570-572, doi: 10.1109/IUCC/DSCI/SmartCNS.2019.00120.
- [4] V. Khaustov and M. Mozgovoy, "Recognizing Events in Spatiotemporal Soccer Data," *Applied Sciences*, vol. 10, no. 22, p. 8046, Nov. 2020, doi: 10.3390/app10228046.