Spatiotemporal Estimation of Player Positioning using Soccer Event Data

Soccer Track

1. Introduction

Player tracking data has the potential to drive value for clubs and present new research opportunities in soccer analytics (e.g., physical metrics and pitch control [1]). However, this data is extremely expensive due to the advanced data collection process, meaning it is unaffordable to the vast majority of clubs. Therefore, in this paper, we present a model to impute snapshots of player tracking data from event-based data which is far cheaper and more widespread (Figure 1). This model consists of a graph network of long short-term memory (LSTM) models and hence captures both the spatial and temporal structure of player positioning. Finally we apply imputed player positions to off-ball analyses such as pitch control and player physical metrics. In summary, we make the following novel contributions:

1. A machine learning model that imputes off-ball player locations using only event data.
2. Evaluation of model accuracy on real-world data, with Euclidean displacement error of 7.68 meters; a 56.8% and 16.3% improvement upon an 'average position' baseline and XGBoost model respectively.
3. We demonstrate that various analyses typically applied to tracking data can also derive useful insights when applied to imputed data.

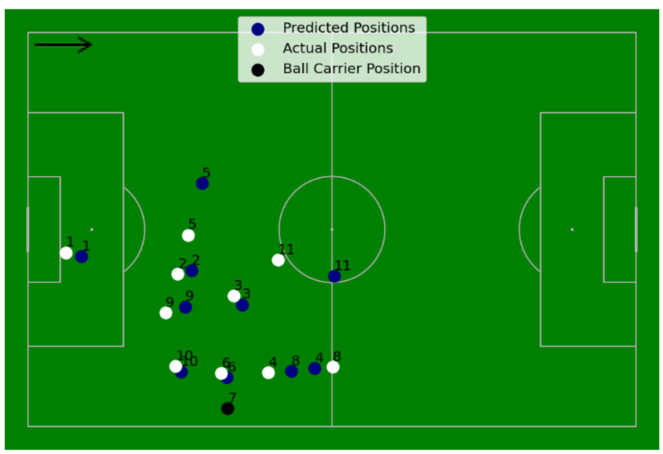

Figure 1- Example event prediction showing imputed off-ball player locations compared to actual player locations for a single team.

2. Methods

Using event and tracking data for 34 matches from the K League 1, we engineer a feature set based on the observed event data. We use the tracking data as ground truth for player positions, and use a train-test split of 32 and 2 games respectively. By extending other imputation models [2], we pass the feature set into a graph network of time-aware LSTM models to impute all player locations for a single event. Finally, we use the imputed player locations to perform off-ball analyses such as physical metrics, player heatmaps and pitch control.

3. Results

The model imputes player position with an average displacement error of 7.68 meters. This improves upon models such as XGBoost (9.17m) and a baseline model which calculates time-based estimates of location using last and next seen location(17.76m). A Spearman correlation of 0.922 is found between the predicted and actual average speed of players during matches. Furthermore, similarity is found between predicted and actual heatmaps of players (Figure 2).

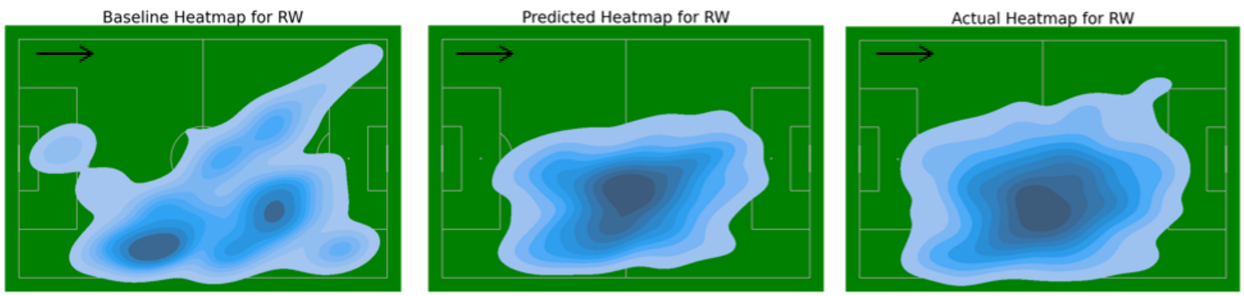

Figure 2- Example match heatmaps for a right winger (RW) using baseline and graph model predictions, compared against actual locations. Darkest regions indicate most frequently occupied areas.

4. Conclusion

We present an advanced AI modeling approach to extract estimated player locations throughout matches. Our work impacts the sports industry by extending the research and analytics potential of lower league clubs and sports researchers who only have access to event data. Our model allows analysts and coaches to supplement their recruitment and post-match analysis workflow with only event data, reducing the impact of data restrictions in stopping clubs from following similar data science processes as the elite level clubs.

References

[1] Spearman, W., Basye, A., Dick, G., Hotovy, R. and Pop, P., 2017, March. Physics-based modeling of pass probabilities in soccer. In Proceeding of the 11th MIT Sloan Sports Analytics Conference.
[2] Omidshafiei, S., Hennes, D., Garnelo, M., Wang, Z., Recasens, A., Tarassov, E., Yang, Y., Elie, R., Connor, J.T., Muller, P. and Mackraz, N., 2022. Multiagent off-screen behavior prediction in football. Scientific reports, 12(1), pp.1-13.

Word Count: 499