Inferring Player Location in Sports Matches: Multi-Agent Spatial Imputation from Limited Observations

Gregory Everett
University of Southampton, UK
gae1g17@soton.ac.uk

Ryan J. Beal
Sentient Sports, UK
ryan.beal@sentientsports.com

Tim Matthews
Sentient Sports, UK
tim.matthews@sentientsports.com

Joseph Early
University of Southampton, UK
j.a.early@soton.ac.uk

Timothy J. Norman
University of Southampton, UK
t.j.norman@soton.ac.uk

Sarvapali D. Ramchurn
University of Southampton, UK
sdr1@soton.ac.uk

ABSTRACT

Understanding agent behaviour in Multi-Agent Systems (MAS) is an important problem in domains such as autonomous driving, disaster response, and sports analytics. Existing MAS problems typically use uniform timesteps with observations for all agents. In this work, we analyse the problem of agent location imputation, specifically posed in environments with non-uniform timesteps and limited agent observability (~95% missing values). Our approach uses Long Short-Term Memory and Graph Neural Network components to learn temporal and inter-agent patterns to predict the location of all agents at every timestep. We apply this to the domain of football (soccer) by imputing the location of all players in a game from sparse event data (e.g., shots and passes). Our model estimates player locations to within ~6.9m; a ~62% reduction in error from the best performing baseline. This approach facilitates downstream analysis tasks such as player physical metrics, player coverage, and team pitch control. Existing solutions to these tasks often require optical tracking data, which is expensive to obtain and only available to elite clubs. By imputing player locations from easy to obtain event data, we increase the accessibility of downstream tasks.

KEYWORDS

Multi-Agent Systems; Football; Imputation; Agent Behaviour

1 INTRODUCTION

Many Multi-Agent System (MAS) applications involve agents interacting in both space and time to achieve their individual and collective goals. Predicting the behaviour of such agents is an important task in many areas such as path prediction for autonomous vehicles [25], predicting locations of civilians in disaster response scenarios using phone data or drone footage [22, 27], and sports analytics [18]. However, situations may arise where the observability of the environment is limited. For example, an autonomous vehicle

may have an obstructed view and lose sight of nearby pedestrians, or drone footage may have limited observability of an area when searching for injured civilians. In these situations, reasonable estimations of agent locations need to be made in order to improve the efficacy of response systems. In this work, we build a multi-agent time-series imputation model, named *Agent Imputer*, which learns typical spatiotemporal interactions between agents to estimate the location of agents when system observability is limited. We apply this to the domain of Association Football (soccer; referred to as football in the remainder of this work) by predicting the location of all players on a pitch when only observing the location and action (e.g., a pass or shot) of the ball-carrying agent.

Football represents an interesting domain for modelling teams as dynamic MAS, with each player having individual roles and behaviours. Whilst all players exhibit different behaviours, the system is centred around the ball. Therefore, in this work, we use data regarding the ball and ball-carrier to estimate all player locations when an event occurs. We build a model with Long Short-Term Memory (LSTM) and Graph Neural Network (GNN) components to learn the temporal and inter-agent relationships between players in a football game to make generalized predictions on unobserved player locations during a game.

Analytical data in football has many uses. These include modelling team tactics, calculating physical metrics (e.g., distance covered), and measuring pitch dominance of teams [7, 12, 23]. It can also be used by football clubs to perform player evaluation, scouting, and opposition analysis. In this work, we examine how sparse football event data of on-the-ball actions can be used to impute knowledge of player positions. This facilitates the desired downstream analysis tasks without requiring expensive tracking data. Our contributions are as follows:

(1) We propose a novel *Agent Imputer* model that predicts agent locations when constrained to limited system observability. This model learns both temporal and inter-agent interactions using LSTM and GNN components to make estimations of agent positioning applied to football.

(2) We use event and tracking data from real-world football games to train and test our model, and find that it predicts agent positioning with a mean Euclidean distance error of ~6.9m, outperforming numerous baselines by ~62%.

(3) We apply our model predictions to real-world downstream applications in the football domain, such as calculating estimations of player physical metrics, pitch dominance, and player coverage heatmaps.

The rest of this paper is structured as follows. Section 2 discusses related literature and data in football, and Section 3 formally defines the problem. Section 4 then introduces our novel *Agent Imputer* model. Section 5 contains our evaluation, and Section 6 presents downstream applications. In Section 7, we discuss model outcomes and future work. Finally, Section 8 concludes.

2 BACKGROUND

In this section, we discuss the task of predicting agent behaviour in MAS (Section 2.1), then cover existing uses of MAS in sports (Section 2.2), and finally motivate the need to impute player tracking by discussing available football data sources (Section 2.3).

2.1 Predicting Agent Spatiotemporal Behaviour

There are a number of works in different MAS domains that predict the spatiotemporal behaviour of agents within dynamic multi-agent environments. These domains include collision avoidance for autonomous vehicles [25, 28], pedestrian trajectory predictions in crowds [1, 10, 15], location habits of human populations [16, 17], and sports analytics [11, 18, 26]. We categorise existing work into two use-cases: predicting future behaviour, and imputing agent locations when behaviour is observed intermittently.

Predicting Future Agent Behaviour. Given the previous behaviour of agents in a MAS, it can be useful to predict their future behaviour. A variety of approaches applied to different problems exist in the literature, such as LSTMs for estimating human trajectories within crowds [1, 10], variational recurrent neural networks for predicting future player trajectories in basketball [26], and imitation learning to predict future trajectories of players in football [11].

Imputing Agent Behaviour. In contrast to predicting future behaviour, imputation focuses on filling in missing information about agent behaviour using past and future observations. This is especially useful when behaviour is observed intermittently. An existing use case in football is the imputation of agent trajectories when players go out of camera view, which was achieved using GNNs and variational autoencoders [18]. Both future behaviour prediction and imputation can also be achieved with a single model [20].

2.2 Multi-Agent Systems in Sports

Existing work in the sports domain has modelled teams as MAS to better understand team behaviour. This includes modelling networks of interactions between players to evaluate teamwork [6], characterising team behaviour by mapping play sequences to game tactics [14], and classifying defensive actions to provide a framework for team tactical analysis [21]. In contrast to modelling teams, several studies have modelled individual player behaviours in the context of MAS. This includes using imitation learning to identify multiple agent policies within a coordinated team structure [11] and using LSTMs to predict trajectories of basketball players [9].

In contrast to existing work, instead of creating a model with a specific downstream sports analysis task in mind, we propose a model to impute missing player data, and then show how this facilitates a wide variety of applications that add value to those who can only access sparse sports data sources.

2.3 Data in Football

The two main spatiotemporal data sources in football are tracking data and event data. We compare both types of data below.

Tracking Data. The current gold standard in football analytics, optical tracking data records the location of every player and the ball ~20 times per second [2]. This equates to ~108, 000 locations for a single player over the course of a game. Access to tracking data facilitates many industrial downstream use-cases for clubs to optimise their tactical setup or recruitment policies by analysing off-ball player movement and team performance [7, 23]. However, collecting this data requires expensive equipment to be installed within stadiums, and as such, it is currently mostly restricted to the top European leagues. Clubs without access to tracking data are unable to perform downstream analysis to the same extent as elite level clubs, leading to a wider gap in performance.

Event Data. In contrast to high frequency tracking data, event data only records information on game events (e.g., passes and shots) along with the event location and the player involved [19]. A typical game involves ~3750 events, or ~170 events per player on average. This data is useful for on-ball statistics, but as it does not gather player information when they are off the ball, it cannot be readily used for many off-ball downstream tasks. However, this data is far cheaper and more easily attainable — there is an abundance of event data from many providers, covering 80+ global leagues.

In this paper, we propose a model that uses event data to impute off-ball player locations, and thus give snapshots of all player locations for all events during a game. This bridges the gap between event and tracking data, and facilitates the desired downstream analysis for clubs unable to access tracking data. Figure 1 gives a visual representation of the difference between event and tracking data, and how we use our model to estimate tracking data.
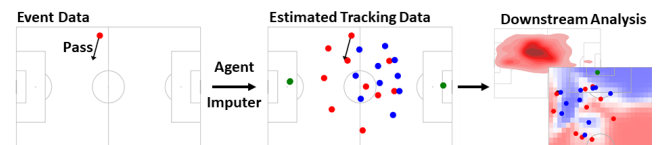

Figure 1: Comparison between event and tracking data for a single timeframe, and how our model estimates tracking data to facilitate downstream analysis tools.

In contrast to the data used in existing work, event data in football leads to sparser agent observations with non-uniform timesteps. This occurs as the behaviour for an agent is only gathered when the player performs an on-the-ball action (e.g., pass, shot, tackle). This often leads to a cluster of agent events followed by an extended period with no information as the player is no longer on the ball. Furthermore, as only a single player can perform an on-the-ball event at a time, data is only gathered for a single agent at each timestep. As such, event data fails to capture important off ball information, such as attacking runs into space or collective defensive positioning, which is crucial when analysing team performance.

To our knowledge, no previous work has studied the problem we approach in this work: imputing agent behaviour within a MAS when information occurs at non-uniform timesteps and for a single agent at a time. In the next section, we formally define this problem.

3 PROBLEM FORMULATION

Our MAS is a collection of N agents, $\mathbf{A} = [a_1, \ldots, a_N]$. In our target domain of football, $N = 22$ (two teams of 11 on-field players). The time-series is a sequence of T events $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_T]$. Note this time-series is non-uniform, as each element $\mathbf{e_t} \in \mathbf{E}$ is recorded when an agent performs an on-ball action (e.g., a pass, dribble or shot), leading to varying gaps between each timestep.

For each event $\mathbf{e_t} \in \mathbf{E}$, we have a set of observations $\Phi_t = [\phi_t^1, \ldots, \phi_t^N]$, where ϕ_t^n is the observation of agent n at timestep t. This gives a complete set of observations over time of $\Phi = [\Phi_1, \ldots, \Phi_T]$. However, in our configuration, only one value in each $\Phi_t \in \Phi$ is known, i.e., there are $N - 1$ missing values for each Φ_t, and $T(N - 1)$ missing values in total. This occurs as we only make an observation of one agent at each timestep — in our target domain of football, this is the position of the on-the-ball agent. As the observed agent changes over time, we construct a one-hot encoded mask $\mathbf{M}$, where $\mathbf{M}_t^n = 1$ if agent n is observed at timestep t, and is 0 otherwise. This $(T \times N)$ binary matrix fully captures the information regarding known and unknown observations across the time-series problem.

In this configuration, the goal of an imputation model is to predict values for the unknown observations. Formally, for each $\mathbf{e_t} \in \mathbf{E}$, the model makes a prediction $\hat{\phi}_t^n$ for every $n \in [1, \ldots, N]$. This leads to a complete set of predicted observations $\hat{\Phi} = [\hat{\Phi}_1, \ldots, \hat{\Phi}_T]$, where $\hat{\Phi}_t = [\hat{\phi}_t^1, \ldots, \hat{\phi}_t^N]$. Note, in this setup, the model is also predicting observations for cases that it has already observed (i.e., players on the ball). However, when we apply downstream analysis (Section 6), we instead use the actual locations for already observed agents.

Relating the above to our target domain of football, the predicted observations $\hat{\Phi}$ are the estimated locations of all players (both on-ball and off-ball) for every event $\mathbf{e_t} \in \mathbf{E}$. The known observations are derived from the location of the events that occur. So, for event $\mathbf{e_t} \in \mathbf{E}$ with an on-ball agent $a_n \in A$, the assigned observation ϕ_t^n is $e_t^{x,y}$, where $e_t^{x,y}$ is the x, y position at which e_t occurred.

To summarise, the set of known observations Φ (containing missing data) is a $(T \times N \times 2)$ tensor, and the set of imputed observations $\hat{\Phi}$ is also a $(T \times N \times 2)$ tensor. Note, each observation is two-dimensional as we are using agent x, y positions — in other domains, these observations could be a different size. In the next section, we explain our model to solve this imputation problem.

4 AGENT IMPUTER MODEL

The problem we describe in Section 3 is highly complex as over 95% of the values in Φ are missing. We aim to extract useful information on agents' spatial and temporal behavioural patterns from event data, and apply it to a model which can consider how agents move over time and in relation to other agents. In this section, we outline the feature engineering (Section 4.1), model architecture (Section 4.2), and training process (Section 4.3) used in our approach. For implementation details, see Appendix A.

4.1 Feature Engineering

The features available at a given timestep are strictly derived from on-the-ball event data, consisting of the event location, player on the ball, time at which the event occurred, and other information such as the current scoreline. This provides little to no information

on the spatiotemporal context of an off-ball agent, such as when their location was last observed (i.e, when they were last on the ball). Therefore, we perform feature engineering on the event data to generate our own feature set which captures a more comprehensive view of each agent and the general movement of play. We create these features for each agent, so they are a combination of agent-specific and global features. For a timestep t with event $\mathbf{e}_t \in \mathbf{E}$, and observed agent $a_n \in A$, we compute the following features:

Agent-Specific Features

`prevAgentTime,prevAgentX,prevAgentY`: time since the agent was last observed, and their location at that time. Formally, this uses the most recent previous timestep where the agent was on the ball: t' s.t. $\mathbf{M}_{t'}^n = 1$ and $t' \leq t$, which can be found by iterating backwards in time from t. Note $t' = t$ if the agent is currently on the ball. Given t', (prevAgentX, prevAgentY) $= e_{t'}^{x,y}$ and prevAgentTime $= t - t'$. If the agent is yet to be observed, we impute these features with the values at the first timestep where they are observed.

`nextAgentTime,nextAgentX,nextAgentY`: time until the agent is next observed, and their location at that time. Similarly to the previous time and location features, this uses the soonest future timestep (including the current timestep) where the agent is on the ball: t' s.t. $\mathbf{M}_{t'}^n = 1$ and $t' \geq t$, which can be found by iterating forwards in time from t. Given t', (nextAgentX, nextAgentY) $= e_{t'}^{x,y}$ and nextAgentTime $= t' - t$. If the agent has no future observations, we impute these features using the last time they are observed.

`avAgentX,avAgentY`: mean location of the agent across the entire game, i.e., the mean position of all events where $\mathbf{M}^n = 1$.

`agentRole`: the agent's role in the team (e.g., central defender or goalkeeper). Different data providers label roles differently — for our data source, there are 16 possible roles, see Appendix B.2.

`agentSide`: binary indicator of whether an agent is on the same team as the current ball-carrying agent.

`agentObserved`: binary indicator of whether the agent is the one performing the current on-ball action.

`goalDiff`: the difference in score between the teams (number of goals for this agent's team - number of goals for the other team).

Global Features

`eventX,eventY`: location of the current event (i.e. $e_t^{x,y}$).

`eventType`: the type of the current event (e.g., a pass or shot). For more details, see Appendix B.1.

These 15 features, for each agent at each timestep, represent a transformation of the original event data. This transformation means the data is more readily usable by machine learning methods. Note that agent positions are relative to the agent's own goal-line, irrespective of direction of play (as opposed to absolute position on the pitch). In the next subsection, we detail our *Agent Imputer* model.

4.2 Model Architecture

Our *Agent Imputer* model estimates the set of agent locations $\hat{\Phi}$ using the engineered features from the set of events $\mathbf{E}$. The goal of this model is to capture agent movement over time, the relationship between on-ball events and off-ball agent locations, and how agents interact as a team within the MAS. In football, this is often defined as team structure. We build an architecture which learns across time and agents within the MAS. Below, we explain this model
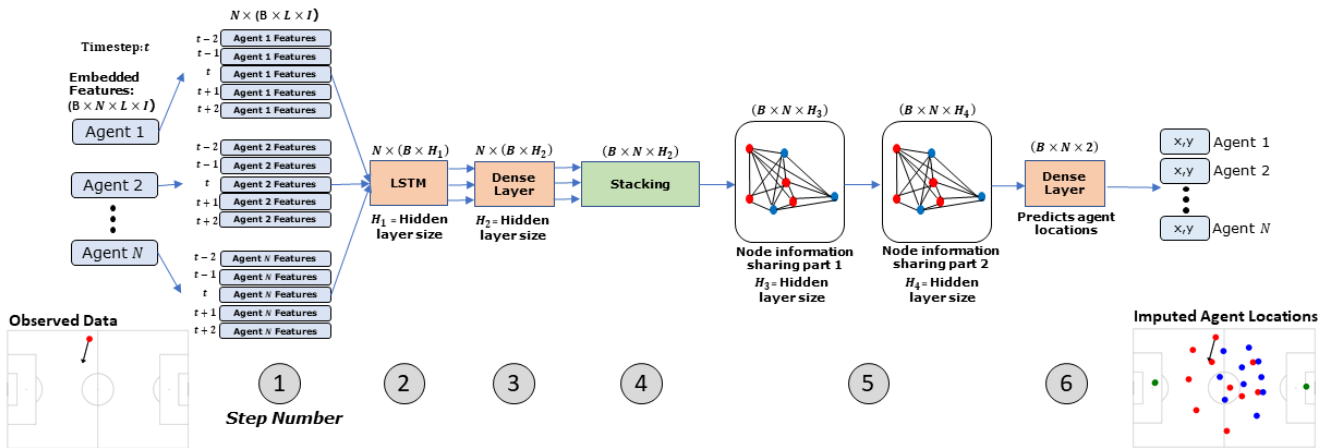
Figure 2: Agent Imputer model architecture. All agent location estimations for an event are made using agent feature information described in Section 4.1, with batch size B, number of agents N, sequence length L, and number of features I.

using a step-by-step process, with corresponding elements of the architecture marked in Figure 2.

4.2.1 **Step 1: Input Formatting**. To capture the ball and agent locations across time, we input a window of event data. We extract the feature set defined in Section 4.1 for a sequence of $L = 5$ events. This sequence is centred around a particular timestep t, i.e., we compute the features for timesteps $\{t - 2, \ldots, t + 2\}$.[1] Each categorical feature (`agentRole`, `agentSide`, `agentObserved`, `goalDiff`, `eventType`) is converted to an embedding (see Appendix A.2 for details). Our imputation model takes all the agent feature sets as a single simultaneous input, such that the model learns the spatial structure of the whole system (as opposed to processing each agent independently). Therefore, the input data has shape $(B \times N \times L \times I)$, where B is the batch size, N is the number of agents, L is the sequence length, and I is the post-embedding dimensionality of our feature set including both the numerical features and embedded categorical features. In this work, $I = 24$ and $B = 128$.

4.2.2 **Step 2: LSTM Component**. Each agent is independently passed into a shared bidirectional LSTM component, i.e., the input data is split into N segments of size $(B \times L \times I)$. The LSTM is then able to learn the temporal relationship between the engineered features and agent location. Importantly, the LSTM is shared across all agents. This overcomes issues with the sparsity of agent observations, as the LSTM is able to learn agent positioning through common movement patterns for agents with similar roles. Due to the irregular time intervals between timesteps, it is important to augment the LSTM architecture to deal with non-uniformity. Therefore, we use a Time-Aware LSTM [3] which adjusts cell memory to alter the discount rate of previous or future actions in the sequence based on the difference in time from the current event. In this work, an LSTM with a single hidden layer of size $H_1 = 100$ is used.

4.2.3 **Step 3: Dense Layer with ReLU Activation**. Outputs from the LSTM model for each agent at the current timestep t (i.e., the middle of the input sequence) are extracted. This output is passed through a dense layer with a ReLU activation function to give an

output of size $H_2 = 50$, resulting in a latent representation of the time-aggregated features for each agent.

4.2.4 **Step 4: Stacked Agent Embeddings**. The temporally aggregated latent representations are stacked together, giving an output tensor of size $(B \times N \times H_2)$. This data is now a suitable input for a graph neural network (GNN), with each tensor row representing agent information that become node features for the GNN.

4.2.5 **Step 5: GNN Component**. The LSTM component of the network models the temporal aspect of agent behaviour. We now look to also model the inter-agent relationships within the MAS, and how these impact the behaviour of each agent. We construct a fully connected graph with the temporally-aggregated agent representations as node features. The GNN uses this graph structure to allow information sharing across all agents. The GNN architecture consists of two message passing layers with feature sizes of $H_3 = 64$ and $H_4 = 32$, using the SAGEConv operator [8]. This updates node features by using a mean aggregation scheme to learn information about agent neighbourhoods (i.e., agent interactions).

4.2.6 **Step 6: Dense Layer with ReLU Activation**. The final output of the GNN component is of size $(B \times N \times H_4)$, i.e., latent representations of size 32 for each agent, which captures both temporal and inter-agent interaction information. These final representations are then passed through a single dense layer with a ReLU activation function to make the final position (x, y) predictions for each agent. Therefore, the final model output is of size $(B \times N \times 2)$.

4.3 Model Training

We use tracking data as target variables during training, and as ground truth labels to test the predictive accuracy of our *Agent Imputer* model. We use the mean Euclidean distance between the predicted positions and true positions as the loss function to train the model. The model was trained for 150 epochs with a batch size of 128. The AdamW [13] optimiser was used with an initial learning rate of 0.002. These hyper-parameters and the exact model architecture (e.g., hidden layer sizes) were found through trial and error, i.e., no formal hyper-parameter tuning was carried out.

---

[1]If timesteps before or after the current timestep don't exist due to being at the start or end of a game, the data for the current timestep is used instead.

Table 1: Predictive performance of models averaged across five folds, along with 95% confidence intervals. We give performance in the X, Y, and Euclidean (XY) directions. All errors are in metres. Bold results indicate the best performance (lowest error).

Models	Train			Test		
	X Error	Y Error	XY Error	X Error	Y Error	XY Error
Baseline 1	14.20 ± 0.02	9.69 ± 0.02	18.81 ± 0.05	14.22 ± 0.24	9.70 ± 0.14	18.82 ± 0.22
Baseline 2	14.10 ± 0.02	9.79 ± 0.02	18.91 ± 0.05	14.04 ± 0.48	9.78 ± 0.17	18.80 ± 0.50
Baseline 3	13.37 ± 0.05	9.63 ± 0.01	18.15 ± 0.05	13.27 ± 0.44	9.57 ± 0.14	18.01 ± 0.46
XGBoost	5.88 ± 0.01	5.14 ± 0.02	8.67 ± 0.02	6.42 ± 0.35	5.50 ± 0.09	9.26 ± 0.26
Time-Aware LSTM	4.46 ± 0.02	4.26 ± 0.02	6.89 ± 0.02	4.48 ± 0.08	4.49 ± 0.1	7.09 ± 0.06
GNN	5.28 ± 0.05	5.08 ± 0.1	8.18 ± 0.09	5.42 ± 0.35	5.13 ± 0.14	8.32 ± 0.25
Agent Imputer	**4.06 ± 0.02**	**4.12 ± 0.02**	**6.47 ± 0.02**	**4.29 ± 0.09**	**4.41 ± 0.11**	**6.88 ± 0.1**

5 EMPIRICAL EVALUATION

In this section, we describe the data used to train and evaluate our model (Section 5.1), introduce the baselines we compare to (Section 5.2), and give our results (Section 5.3).

5.1 Datasets

We train and evaluate our model using 34 games of event and tracking football data collected from K League 1[2] and supplied to us by Bepro Group Ltd. These are gold standard industry datasets that allow us to rigorously evaluate our approach. Each game provides event sequences, which we use as model input, and tracking data, which we use as training targets and for evaluation. In total, there are ~64,000 events and ~1.4 million tracking locations. We use a 31/3 (~91.2%/8.8%) train/test split and five-fold cross validation to evaluate our model. All geometric data is scaled to a standard football field size of 105x68 metres. We provide further details on the dataset in Appendix B.1.

5.2 Baselines

We use a number of baselines to evaluate whether our model has improved performance compared to other imputation methods. We first propose three naïve imputation baselines. Baseline 1 simply predicts agent location using the average on-ball location of the agent during the match. Baseline 2 predicts the centroid between the last and next observed location of the agent, and Baseline 3 predicts a time-scaled position on the straight-line trajectory between the last and next observed location of the agent. We also compare our *Agent Imputer* model to other machine learning models with the same input feature set, including an XGBoost regression model, a Time-Aware LSTM model (i.e., the same model as in Figure 2 without step 5), and a GNN model (i.e., steps 4, 5 and 6 of Figure 2, using the middle of the original input sequences as input features). These last two baselines allow us to compare how the constituent parts of our *Agent Imputer* model contribute to performance.

5.3 Results

Four experiments are used to evaluate the performance of our *Agent Imputer* model and relevant baselines. We evaluate player position prediction (Section 5.3.1), performance over time (Section 5.3.2), performance over different positions (Section 5.3.3), and how performance varies with observation (Section 5.3.4).

[2]The top men's professional football division in South Korea

5.3.1 Position Prediction. We evaluate the predictive performance of the model and baselines in Table 1. Note the X direction is along the length of the pitch (105m) and the Y direction is along the width of the pitch (68m), and that we do not normalise the error in these different directions by pitch size. We find that the *Agent Imputer* model predicts agent location with highest accuracy (lowest distance error). We also find that the Time-Aware LSTM model outperforms the GNN model, suggesting the essential part of our *Agent Imputer* is the LSTM component. However, we still see an improvement in performance by including the GNN component to the *Agent Imputer* model, demonstrating the value of modelling agent interactions. We note a 61.8% decrease in error using the *Agent Imputer* compared to the best performing naïve baseline. We also find the *Agent Imputer* model has roughly equal error in both X and Y directions.

5.3.2 Predictability over Time. We evaluate predictive accuracy of the model across different periods of a match. In football, the game has two 45 minute halves, so we evaluate how model accuracy changes during these periods. We compute this using a rolling average of model error; see Figure 3. Additional time at the end of halves is accounted for — first half added time is merged with early second half predictions, and values beyond 90 minutes facilitate second half added time.
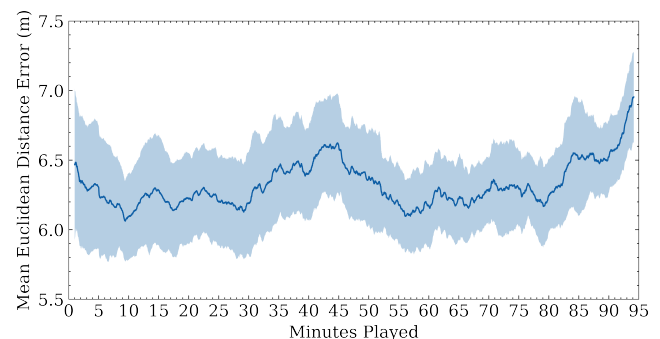

Figure 3: Predictive performance of the *Agent Imputer* model across time using a rolling average of Euclidean error (5 minute windows with a 1 second step size) on the combined training and testing data (all 34 games). Shading shows the standard deviation across means for each game.
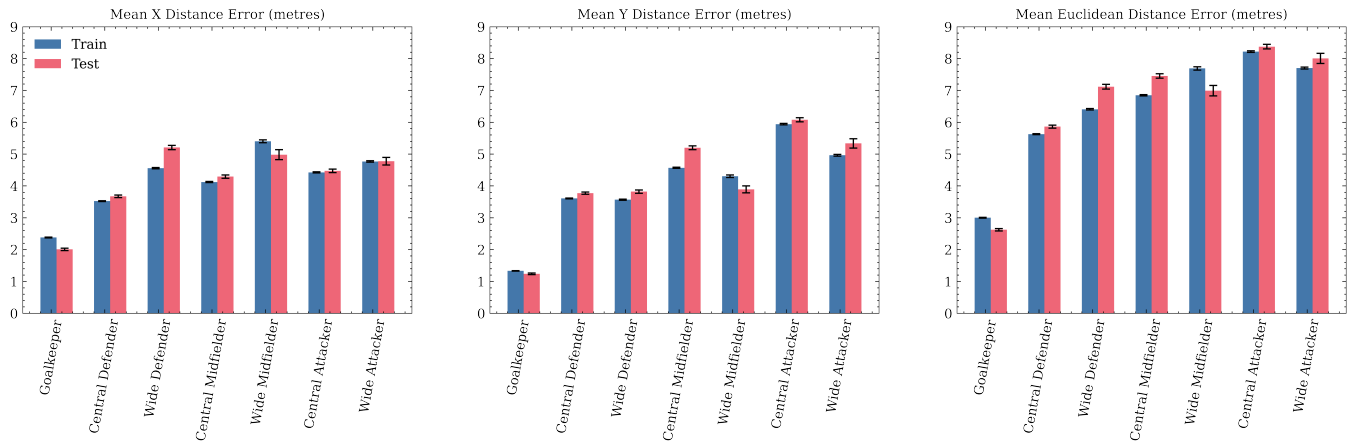
Figure 4: Mean distance error of player location estimations for each agent role using the *Agent Imputer* model.

From these results, we suggest that the most unpredictable moments of a game are at the end of both halves. Due to the error margin, we perform t-tests to investigate the hypothesis that the mean of the underlying sample distribution in the middle of a half is lower than the underlying distribution at the end of a half. When doing this for the the first half (all events between 20-25 minutes compared to all events between 42.5-47.5 minutes), this is found to be significant (p<0.01). The difference in the second half (comparing all events between 65-70 minutes with all events between 87.5-92.5 minutes), is also found to be significant (p<0.01).

This supports our theory that the game is more unpredictable at the end of halves, which corroborates the popular intuition that players get tired and teams become less structured at the end of halves. Furthermore, teams are looking to get back into a game or defend a lead, which leads to more chaotic and unpredictable periods. This highlights match phases at which team structure varies, and this type of analysis may also uncover situations where players are moving in 'unusual' ways. This type of analysis could be extended to identify links between unusual movement patterns and team performance, and used as a possible tool to proactively identify performance change during a match.

5.3.3 Predictability over Roles. We evaluate the performance of our *Agent Imputer* model across agent roles in Figure 4. This gives insight into how an agent's goals and responsibilities within a team affects the predictability of their behaviour. For simplicity, in this experiment and downstream applications (Section 6), we group roles into wide and central positions, reducing the total number of positions from 16 to 7 (see Appendix B.2). Goalkeepers are expectedly the most predictable role, as their range of movement is usually limited to their own box. We also find that defenders are more predictable than attackers, highlighting that defensive agents behave in a more structured way than attacking agents.

Interesting findings can also be drawn from the X and Y distance errors. A general trend is that the model finds it harder to predict the X location of wide outfield players in comparison to central outfield players (an average of 4.99m vs 4.15m), which is expected as wide players cover a lot of ground along the wings of the pitch. However, the model generally predicts the Y location of wide players better than central players (4.35m vs 5.02m), which is also expected as

wider players usually stick to a single side of the pitch. These comparisons help identify where the model could be improved, and could be used to help convey uncertainty in downstream analysis.

5.3.4 Predictive Performance over Observation Offset. In Figure 5, we evaluate the predictive performance of our proposed models with respect to "time since last observed" — the amount of time that has transpired since a particular agent was last involved in an event. We also do the same for "time until next observed" — the amount of time until the agent is next involved in an event. We find that the performance of the XGBoost and GNN model rapidly decrease as time increases. In comparison, the *Agent Imputer* and Time-Aware LSTM models show a slower decay in performance, as well as a lower plateau. This demonstrates the necessity of modelling temporal aspects of predictions. However, the time-aware models still decrease in performance as time increases. Similarly to the error by player role (Section 5.3.3), this may be useful in measuring the uncertainty of model predictions, i.e., there is greater certainty in predictions for agents that have been on the ball more recently.
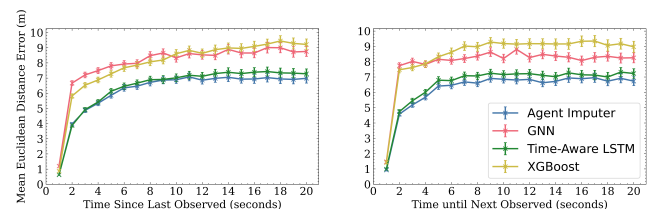

Figure 5: Model performance compared to the last (left) and next (right) time the agent was observed. We use one second intervals with a one second window, e.g., at 2s, we evaluate for all events between 1-2 seconds.

6 MODEL APPLICATIONS

In this section, we demonstrate downstream analysis that can be performed using outputs from our *Agent Imputer* model. In the football domain, these types of analyses are often implemented by elite level clubs using tracking data. However, we demonstrate that lower league clubs with limited resources could use our model to perform similar processes to the elite level clubs without requiring

expensive and hard-to-obtain tracking data. We apply our model to analysing player physical metrics (Section 6.1), pitch control (Section 6.2), and positional heatmaps (Section 6.3).

6.1 Player Physical Metrics

We show how predicted player locations from our *Agent Imputer* model can be used to estimate player physical metrics. We focus on distance covered, which is calculated by summing the Euclidean distances between a player's predicted locations throughout a match. We initially found that the predicted distance covered is consistently higher than the actual distance covered, showing an average overestimate of 11.5% for *Agent Imputer*, 16.8% for the Time-Aware LSTM, 31.7% for the GNN, and 29.7% for XGBoost.

This overestimation bias can be attributed to the large number of events that occur soon after another event — 29.9% of events happen within one second of the previous event. Typically, these are duels between a player on each team, or a player receiving the ball and then quickly passing. In these instances, the players' predicted positions shift markedly, i.e., moving faster than the maximum player speed. This implies the models are not considering realistic player trajectories between events which occur in quick succession. Fixing this model issue is future work. For now, we use a post-processing step that combines events which occur within a second of each other, using the model output for the initial event as the player location when grouping. This leads to more accurate distance covered estimations. Results for the predicted distance covered averaged across different player roles are shown in Table 2.[3]

Table 2: Player distance covered results. Distance is averaged over the test data and given in kilometres. Absolute error is averaged over each game for each player role.

Role	Pred. Dist.	True Dist.	Abs. % Error
Goalkeeper	3.23	3.12	4.05 ± 2.52
Central Defender	8.26	8.42	4.31 ± 1.38
Wide Defender	8.97	9.08	4.99 ± 1.89
Central Midfielder	9.37	9.48	3.66 ± 1.26
Wide Midfielder	8.73	8.97	2.79 ± 0.76
Central Attacker	8.34	8.35	6.03 ± 2.32
Wide Attacker	8.26	8.70	5.33 ± 0.76

6.2 Pitch Control Analysis

To test a model application which considers agent locations relative to each other within the system, we calculate pitch control of teams. Pitch control [7, 23, 24] is a popular downstream analysis tool within football analytics that uses player locations and trajectories to compute the area in which a team dominates. It splits the pitch into a grid of zones and computes the probability of the attacking team controlling the ball if it arrived in that zone. For example, if a defending player is closer to the zone than the nearest attacker, it

---

[3]Distance is normalized as if all players play for 90 minutes (i.e., accounting for extra time and substitutions). Players that played for less than 20 minutes were excluded from these results. Results averaged across all games for players in a certain role.

is unlikely the attacking team will control the ball in that zone. We perform similar analysis to that used in [18] — we compute pitch control using *Agent Imputer* outputs and compare with the same computed from actual player locations in tracking data. Table 3 shows the mean absolute error of pitch control for each model, and shows that the *Agent Imputer* performs best.

Table 3: Pitch control performance compared to the ground truth. Calculated over test data.

Model	Mean Average Error
Baseline 1	0.272 ± 0.002
Baseline 2	0.304 ± 0.002
Baseline 3	0.305 ± 0.002
XGBoost Regression	0.150 ± 0.001
GNN	0.149 ± 0.001
Time-Aware LSTM	0.135 ± 0.001
Agent Imputer	**0.130 ± 0.001**

We provide example of pitch control outputs in Figure 6. These visualisations can be used by clubs to review their team's dominance in particular pitch areas in different game scenarios. This plot also highlights the differences in predicted player locations between the different models. It can be seen that, for the defending team, players are further apart and a structured horizontal defensive line is shown in the *Agent Imputer* predictions. Whereas for the other models, some defenders can be seen to be stacked vertically - a very unlikely prediction given the usual defensive setup and strategy employed by football teams. This suggests that the GNN module is correctly learning how agents spatially interact.
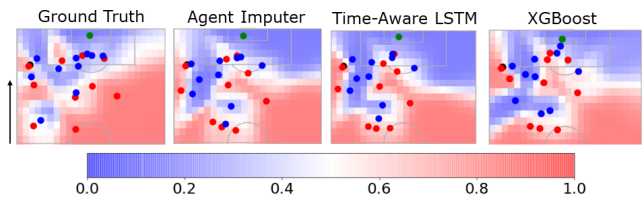

Figure 6: Pitch control diagrams comparing three models with ground truth. Red regions are in control of the attacking team, and blue regions are in control of the defending team. We only show the attacking half as the remainder of the pitch is entirely under the attacking team's control (red). Arrow refers to attack direction.

6.3 Player Heatmap Analysis

In this section, we use predicted player locations from the *Agent Imputer* model to generate player heatmaps over an entire game. This provides coaches with useful information on the areas most frequently covered by their players. Note, while it is possible to generate heatmaps from event data, due to the nature of the data, they will be missing ∼95% of player positions. Figure 7 presents some comparisons between heatmaps of a central attacker and central defender using imputed positions.
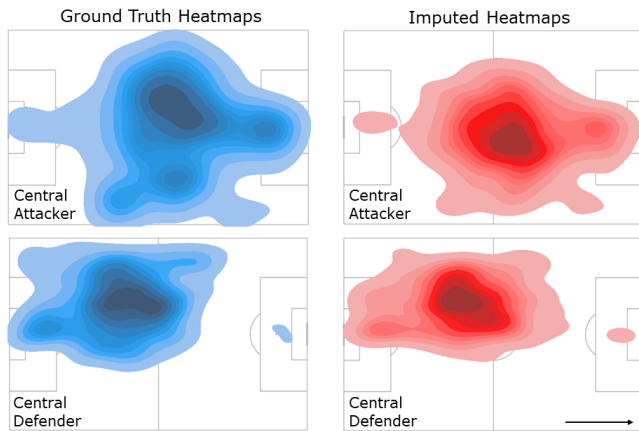
Figure 7: Example *Agent Imputer* heatmaps compared with ground truth heatmaps. Darkest regions indicate most frequently occupied areas. Players are playing from left to right. We give further plots for all player positions in Appendix C.

The imputed heatmaps show many similarities to the ground truth heatmaps. Note the model has picked up interesting features, such as the small distinct region for the central defender near the opponent's goal. This most likely stems from set-pieces (such as a corner), demonstrating the *Agent Imputer* has learnt features of the game state and how this influences player position. Furthermore, the central defender heatmap is offset to the left-hand wing of the pitch. In comparison, the central attacker heatmap which is more central. This shows the model has learnt role-specific information — typically there are two or three central defenders, allowing them to focus on particular sides of the pitch, whereas a central attacker is likely more free to roam in an unstructured manner.

7 DISCUSSION

In this section, we further discuss our findings (Section 7.1) and highlight limitations and areas of future work (Section 7.2).

7.1 Analysis of Results

As shown by our results (Section 5.3), our model is able to impute agent locations using very sparse data, and do so with greater accuracy than naïve baselines and other machine learning models. The increase in performance compared to XGBoost suggests that the temporal and inter-agent dynamics captured by the LSTM and GNN components offer value in learning agent location.

Our *Agent Imputer* model also enables easier access to insightful use-cases in football, such as player physical metrics, player heatmaps, and team pitch dominance analysis (see Section 6). These use-cases could be implemented by lower league clubs with fewer resources to perform similar player evaluation, scouting, and game analysis processes to elite level clubs. This would help close the performance gap between these teams.

As an example, the player heatmaps are shown to be strongly correlated to the ground truth heatmaps exhibited in-game. This can provide value to an opposition analytics scout who could monitor an opponent and get their players to exploit areas that are frequently left uncovered by the opposition team. It may also allow scouts to find players who occupy dangerous space more often.

The predictive ability of our model demonstrates that positioning in football is somewhat predictable. This is likely due to tactical positioning, where teams will be set up to defend and attack in consistent shapes. Work that considers the game theoretic implications of team tactics may introduce further insight into the predictability of teams and how this varies over time [4, 5]. Furthermore, players in certain roles will be given instructions on where to position themselves in certain patterns of play. These patterns can be learned by our model to predict future occurrences of such positioning.

7.2 Future Work and Limitations

Our *Agent Imputer* approach could also be applied to other MAS with limited observability, such as searching for injured civilians in disaster response systems or predicting human movement from phone location data [16, 17]. Change of domain will inevitably lead to some differences in problem structure, such as the bounds of the environment or the sparsity of the dataset. These differences must be considered when applying our method to other domains. Another research direction would be to make the model probabilistic so that uncertainty in agent locations can be quantified.

Our model performs well at estimating location. However, as discussed in Section 5.3.2, extracted trajectories from these estimations can sometimes be unrealistic when events happen in quick succession (although we note this issue is worse in the non-temporal models we compared our approach to). Future work could consider predicting the trajectory of an agent over a number of timesteps simultaneously to mitigate this issue.

Further studies could also compare our proposed *Agent Imputer* model against state-of-the-art models that use tracking rather than event data (e.g., [11, 18]). In this work, we have only compared to other models that use event data. Comparison to tracking data models would facilitate better comparison of the effectiveness of our imputation approach in downstream analysis tasks.

Finally, we used a relatively small number of games in this work (34 matches in total) as it is difficult to obtain tracking data. Future work could extend our dataset to include a larger number of games. This would help the model learn a wider range of game scenarios beyond the ones that occur within our current dataset. Furthermore, this would facilitate the development of more specific models, for example, focusing on a particular player or team.

8 CONCLUSION

This paper presents a novel *Agent Imputer* model to address a multi-agent behavioural prediction problem in dynamic environments with limited system observability. We apply this task to football by imputing off-ball player locations using only on-ball data. We find that our model can impute player location to within ~6.9 metres, outperforming multiple baseline imputation models. We perform deeper analysis of model performance, examining accuracy over different game times, player roles, and observation rates. We also present football analytics applications facilitated by our model, which allow lower league clubs to perform similar processes to elite level clubs without requiring access to expensive data. This novel work could be applied to other real world domains involving limited agent visibility, such as disaster response or tracking daily human movement from phone data.

ACKNOWLEDGMENTS

REFERENCES

[1] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. 2016. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 961–971.

[2] Sian Barris and Chris Button. 2008. A review of vision-based motion analysis in sport. *Sports Medicine* 38, 12 (2008), 1025–1043.

[3] Inci M Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K Jain, and Jiayu Zhou. 2017. Patient subtyping via time-aware LSTM networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 65–74.

[4] Ryan Beal, Georgios Chalkiadakis, Timothy J Norman, and Sarvapali D Ramchurn. 2020. Optimising Game Tactics for Football. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. 141–149.

[5] Ryan Beal, Georgios Chalkiadakis, Timothy J Norman, and Sarvapali D Ramchurn. 2021. Optimising Long-Term Outcomes using Real-World Fluent Objectives: An Application to Football. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*. 196–204.

[6] Ryan Beal, Narayan Changder, Timothy Norman, and Sarvapali Ramchurn. 2020. Learning the value of teamwork to form efficient teams. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 7063–7070.

[7] Javier Fernández, Luke Bornn, and Dan Cervone. 2019. Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer. In *13th MIT Sloan Sports Analytics Conference*.

[8] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems* 30 (2017).

[9] Sandro Hauri, Nemanja Djuric, Vladan Radosavljevic, and Slobodan Vucetic. 2021. Multi-Modal Trajectory Prediction of NBA Players. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 1640–1649.

[10] Boris Ivanovic and Marco Pavone. 2019. The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2375–2384.

[11] Hoang M Le, Yisong Yue, Peter Carr, and Patrick Lucey. 2017. Coordinated multi-agent imitation learning. In *International Conference on Machine Learning*. PMLR, 1995–2003.

[12] Daniel Link, Steffen Lang, and Philipp Seidenschwarz. 2016. Real time quantification of dangerousity in football using spatiotemporal tracking data. *PloS one* 11, 12 (2016), e0168768.

[13] Ilya Loshchilov and Frank Hutter. 2018. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*.

[14] Patrick Lucey, Alina Bialkowski, Peter Carr, Eric Foote, and Iain Matthews. 2012. Characterizing multi-agent team behavior from partial team tracings: Evidence from the english premier league. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 26. 1387–1393.

[15] Francesco Marchetti, Federico Becattini, Lorenzo Seidenari, and Alberto Del Bimbo. 2020. Multiple trajectory prediction of moving agents with memory augmented networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).

[16] James McInerney, Alex Rogers, and Nicholas R Jennings. 2012. Improving location prediction services for new users with probabilistic latent semantic analysis. In *Proceedings of the 2012 ACM conference on ubiquitous computing*. 906–910.

[17] James McInerney, Jiangchuan Zheng, Alex Rogers, and Nicholas R Jennings. 2013. Modelling heterogeneous location habits in human populations for location prediction under data sparsity. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. 469–478.

[18] Shayegan Omidshafiei, Daniel Hennes, Marta Garnelo, Zhe Wang, Adria Recasens, Eugene Tarassov, Yi Yang, Romuald Elie, Jerome T Connor, Paul Muller, et al. 2022. Multiagent off-screen behavior prediction in football. *Scientific reports* 12, 1 (2022), 1–13.

[19] Luca Pappalardo, Paolo Cintia, Alessio Rossi, Emanuele Massucco, Paolo Ferragina, Dino Pedreschi, and Fosca Giannotti. 2019. A public data set of spatio-temporal match events in soccer competitions. *Scientific data* 6, 1 (2019), 1–15.

[20] Mengshi Qi, Jie Qin, Yu Wu, and Yi Yang. 2020. Imitative non-autoregressive modeling for trajectory forecasting and imputation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12736–12745.

[21] Dominik Raabe, Reinhard Nabben, and Daniel Memmert. 2022. Graph representations for the analysis of multi-agent spatiotemporal sports data. *Applied Intelligence* (2022), 1–21.

[22] Sarvapali D Ramchurn, Trung Dong Huynh, Feng Wu, Yukki Ikuno, Jack Flann, Luc Moreau, Joel E Fischer, Wenchao Jiang, Tom Rodden, Edwin Simpson, et al. 2016. A disaster response system based on human-agent collectives. *Journal of Artificial Intelligence Research* 57 (2016), 661–708.

[23] William Spearman. 2018. Beyond expected goals. In *Proceedings of the 12th MIT sloan sports analytics conference*. 1–17.

[24] William Spearman, Austin Basye, Greg Dick, Ryan Hotovy, and Paul Pop. 2017. Physics-based modeling of pass probabilities in soccer. In *Proceeding of the 11th MIT Sloan Sports Analytics Conference*.

[25] NN Sriram, Buyu Liu, Francesco Pittaluga, and Manmohan Chandraker. 2020. Smart: Simultaneous multi-agent recurrent trajectory prediction. In *European Conference on Computer Vision*. Springer, 463–479.

[26] Chen Sun, Per Karlsson, Jiajun Wu, Joshua B Tenenbaum, and Kevin Murphy. 2018. Stochastic Prediction of Multi-Agent Interactions from Partial Observations. In *International Conference on Learning Representations*.

[27] Jing Wang, Jianping Cai, Xiaohang Yue, and Nallan C Suresh. 2021. Prepositioning and real-time disaster response operations: Optimization with mobile phone location data. *Transportation research part E: logistics and transportation review* 150 (2021), 102344.

[28] Xu Xie, Chi Zhang, Yixin Zhu, Ying Nian Wu, and Song-Chun Zhu. 2021. Congestion-aware multi-agent trajectory prediction for collision avoidance. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 13693–13700.