A spatiotemporal Model Predicts the Dynamics of a Basketball Game.

A.Shana
University of Kufa
Adelshanna3000@gmail.com
Riyadh Alsaeedi
University of Wasit
ralsaeedi@uowasit.edu.iq

Abstract—Our work as a spatiotemporal model predictive dynamic (STMPD) on a basketball game capable of high-level decision making . Our motivated and skilled experts solve complex decision-making problems at every time point during a game. The conditional intensity of a temporal point process can be learned based on a network, and the likelihood is maximised during training. **A** Fine-grained player locations started with the simplest case of modelling, where certain players tend to shoot the ball. They discretize the court into cells fine enough to capture all significant spatial variations in model for predicting dynamics in-game, e.g., shooting policy, passing, and who relies on Markov transition probabilities to propagate action through a possession. We propose a method for learning factored Markov decision problems . We use a method that captures the multi-modal behaviour of the players' truth ground and estimation trajectory Prediction is a method for predicting multimodal trajectory by learning a model that assigns probabilities through conducting temporal classification (resolution) (CTC) architectures by generalising away from actual time stamps to recreate the predicted trajectory. Our proposed method of contribution is as follows : i) The method is built on an LSTM-based architecture predicting multiple trajectories and our probabilities, and then we use training with a multimodal loss function when updating the best trajectories . ii) a discriminative learning method for automatically training models to predict near-term game events based on current game conditions . iii) learning factored Markov decision processes (FMDP) problems from domain exploration and expert assistance, ensuring convergence to near-optimal behaviour even after the agent has started to learn critical success factors. IV) An algorithm 1,2, detects changes in possession, pass, and shot players that have been incrementally learned for all components of FMDP, and algorithm 3, which guarantees convergence to near-optimal behaviour even when the agent begins unaware of factors critical to success. A group of players shooting probability to enhance a probabilistic matrix factorization decision method to predict score results in the NBA.

Keywords -Adaptive Event Detection, Spatial and Temporal Model , Algorithm detecting possession updates, Algorithm Detecting passes and shots Player , Predicting Dynamic s

I.INTRODUCTION

In recent years, artificial intelligence and computer vision have started revolutionising how particularly important in invasion sports performance and results are being analysed and trained, which includes the use of fine-grained player tracking data during sporting events, with exact player and ball motion prediction. Refer to [12] for a previous application of statistical techniques to analyses technical and tactically relevant events, the outcome of shots on target, and the ratio of points to ball possession to determine variables supporting successful team performance. It has recently been utilizing multi-agent alignment to players where bottom-up and top-down group representations of trajectory prediction are utilised

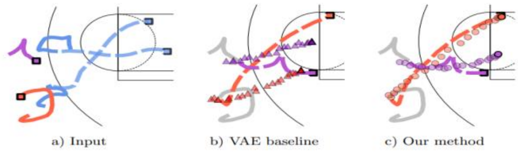

Fig. 1: a) Given a 2D trajectory history of moving agents (solid lines), and the future motion of a subset of the agents (blue dashed lines); our prediction task b) is to generate the most likely motion of the other agents (orange, purple dashed lines) .

On the other hand, there are use cases that can be useful for accurate motion agent prediction, which extends to other applications [25], for example, improving game broadcasting on television by optimizing camera movement [13], [14]. Prediction of human trajectories can be used to enhance tracking accuracy by[10][16] [15]. In order to train a model to predict multiple trajectories and their probabilities, we rely on an adaptation of the multi-modal loss function presented in [40].

II. Related Work

Multi-modal trajectory predictions are used to generate multi-modal use training with a multimodal similar loss function for pedestrian trajectories. This is difficult in a number of scientific areas [58]. Researchers have worked on this problem to improve the combined performance and the analogies between the coarsened possession process [4]. The main reason is that it is an area of [17] [1] [2] that affects learning fine-grained spatial models for dynamic sports play prediction of the game. A dataset of fine-grained, personalized, adversarial multi-agent tracking data which can be made available for research purposes [9], [10], [16]. We proposed an iterative method to calculate the probability distribution vector of a transition probability tensor (TPT) . MDPs framework with inherent non-stationarity the shot clock makes assessing shot choice optimality solution for the problem for basketball plays dependent TPT, as studied by [32]. It is an area of their research . Sports analytics systems rely on a variety of data sources for event detection[63] [68]. They We have described an unsupervised learning framework based on a time-varying Poisson process model that can account for anomalous events [26]. Event detection from time series data demonstrated that the maximum likelihood estimate for is equivalent whether it maximises over the product of independent Poisson densities , according to [6] ,[29], [30]. hey used hierarchical spatial prediction modelling specifications for large data sets [47]. A particle swarm optimization(PSO) based algorithm for object tracking Object locations in each frame are determined from a Gaussian distribution to cover the PSO. This takes place in order to concentrate the particles near the true object state. [50] [51] [52]. Our work describes online learning, which employs adaptively updated proximal functions to control the gradient steps of the stochastic optimization algorithm, which significantly simplifies setting a learning rate and computing descent second order gradient vectors that are sparse by constructing approximations to the functions Hessian [59]. The problem of planning under uncertainty in observation space is very large. Using a dynamic of probabilistic matrix factorization (PMF) for Bayesian variant networks to generate spatial and temporal abstractions capable of high-level decision making ,where transition and reward functions of an factored MDP are defined [33] , [53]. They have derived a belief-dependent reward that the accumulated rewards correspond exactly to the performance criterion [54] and problem of inferring the transition , used by [55]. Our experiments show our agent learns optimal behaviour on small and large problems while MDP with unawareness is defined as (MDPUs).We propose a system where an agent makes explicit attempts to include its unawareness with sequential decision problems [57]. Conditional random fields (CRF) provide a unique combination of properties for discriminatively trained models for sequence segmentation and labelling [65]. We propose a method for testing and comparing shot policies that computes for the dynamic regularity of a basketball play. Our goal is to learn the state-value function. The expected possession value (EPV) called the expected return from a state and is based on a policy that defines the probability to solve the problem of estimating the long-term reward that a team in possession of the ball could expect according to the game situation state at any adaptive event time With soccer spatiotemporal data, traces of player trajectories can be discrete actions that a player can take at any time [24][25],[61] . [5] to incorporate spatial priors and non-negative matrix factorization technique (NMF), and the constraint makes our approach similar to related approaches based on NMF [19] [20] [37], in order to best decision making for NBA shot selection. Learning factors correspond to an interpretable representation that can recover the adjacency matrix of a given network with block stochastic gradient descent of NBA players, by [19] and [21].We have knowledge that how players make their decisions and (NBA) players are movement trajectories during game We understand how players make their decisions and (NBA) players are movement trajectories during a game who are highly motivated and skilled experts that solve complex decision-making problems at every time point.

A .Adaptive Event Detection

A simple approach for detecting a pass, so, can be based on detecting two consecutive basic continuous-time events for predicting basketball possession, a general framework for continuous-time within-play valuation in football using player-tracking data [61]. [6] proposed an iterative likelihood-based method for segmenting a time series into piecewise homogeneous regions. Referring to [7], in the data analysis, the Bayesian modelling intensity and reversible jump Markov chain Monte Carlo (RJMCMC) methods can be used as dynamic programming algorithms for finding the best fitting and piecewise constant intensity functions can be utilised to model continuous intensities. This paper investigates the problem of segmenting sets of low-level time-stamped events into time-periods of relatively constant intensity. This is a parallel hidden Markov model for each trajectory under study. Simple heuristics for pruning the number of the potential change points of the functions provide approach to modelling by indicating the expected number of event sequences using time-dependent intensity functions. Each upper level state is modelled using a Gaussian mixture model, a combination of Poisson models and Bayesian estimation methods, the duration of time spent. The model proposed is derived from the Markov modulated Poisson processes used by [8]for analysis of [22]. Detecting soccer video streams and detecting the events directly from [24], [23]. A survey conducted by [25] lists numerous tasks related to the spatiotemporal analysis of team sports. However, as of yet, only a few attempts have been made to deal with ball-event processing. Furthermore, these studies concentrate on categorizing and labelling known events [3], [67] predicting future events [18], [27].Certain approaches to automatic event detection have been proposed only in recent introduced paper by [28], [68]. Both of these factors complicate the data mining problem of locating and analysing these anomalous events. We describe an unsupervised learning framework based on a time-varying Poisson process model that can account for anomalous events.In [26], consider a Multinomial-Poisson transformation as y = $(\mathcal{Y}_1, \mathcal{Y}_2,\ \mathcal{Y}_n)$ Poisson random variables are independent with means λ = $(\lambda_{1,}\ \lambda_{2,},\ldots,\ \lambda_{n,})$. The joint distribution of e factorizes into the product of a multinomial distribution and a Poisson distribution over n = $\sum_{j=1}^{M} \mathcal{Y}_j$and f(y)=$\prod_{i=1}^{M} f(y_i)$ and Poisson(n| Λ)Multinomial($\mathcal{Y}_1|\alpha$,n**)** Λ=$\sum_{j=1}^{M} \lambda_{j\,,}\ \alpha = (\alpha_1, \alpha_2, \ldots \alpha_M)$. $\alpha_i = \frac{\lambda_{i\,,}}{\sum_{j=1}^{M} \lambda_{j\,,}}$ (1) **.**For the constraint n = Λ , demonstrated that the maximum likelihood estimate is equivalent whether it maximises over the multinomial density/ the independent Poisson densities is product . (according to [29], [30]). The connection to modelling Markov transitions is obvious; assuming Poisson, we estimate the dependent transition probabilities as follows: $\mathcal{P}_{ij}$ = $\lambda_{ij}/\sum_{j=1}^{k} \lambda_{ij}$ (2) according to [6].

B. Fine-Grained Dynamic Spatial Models

They adapt their approach to represent the spatial structure that governs shots that can be selected among basketball players. A spatial of player shooting probability to enhance a probabilistic matrix factorization method to predict score results in the NBA. [64] has been significantly based on the statistical approach to the temporal point process concept.

On the other hand, its effectiveness in capturing the underlying temporal dynamics as well as the correlation within sequential pattern activity data ,according to [60], also extends their system dependent based on the estimation application for predicting the timing. [34] enhanced any application for the estimation of the defensive effect of NBA centers on shot sequence and shot efficiency. Both of the spatial/temporal functions and a nonparametric prior have been used to model neural data, in[35].**A** spatially random point process in some space is represented Poisson Processes that can be derived from the entire X, for which the number of points defined as A that end up in some set A ⊆ X is Poisson distributed. They use an inhomogeneous Poisson process on domain X. That is, used a model of the set of spatial points, $x_1, , ..., x_N$ with $x_n \in$ X, as a Poisson process with a non-negative intensity function λ(x) : X → R+. Fitting the a log Gaussian Cox process(LGCPs) for each player's set of points, x_n, the likelihood of the point process is discretely approached as:

$p(x_n|\lambda_n(.)) \approx \prod_{v=1}^{V} p(x_{n,v}|\Delta A\lambda_{n.v})$ (3)

where, symbols , $\lambda_{n.v}(.)$ is the exact intensity function, λ_{nn} is the discretized intensity function (vector), and ΔA is the area of allowing us to treat each tile independently that is implying one from Progressive. , uniform intensity $\lambda_{n.v}$.Explicitly representing the Gaussian random field z_n, the posterior as follows :$p(z_n|x_n) \propto$

$p(x_n|z_n)p(z_n)=\prod_{v=1}^{V} e^{-\lambda_{n.v}} \frac{\lambda_{n.v}{}^{x_{n.v}}}{x_{n.v}!} \mathcal{N}(z_n|0,k)$ (4)

$\lambda_{n=\exp(z_n+z_0)}$,where the prior cross z_n is a mean zero normal with covariance $K_{n.v}$ = k(x_v; x_u), determined by Equation (4), and z_n is a bias term that parameterizes the mean rate of the Poisson process. Samples of the posterior $p(\lambda_n|x_n)$ can be done by transforming samples of $z_n|x_n$. They instead employ elliptical slice sampling. [36] to approaches defined as the posterior for each player of λ_n, and it has subsequently store the posterior mean . The optimization problem has also been solved by the tactics from [37] and [20], comparing them to highlight the difference between the resulting basis vectors. Consider the latent factors **(L)** representing the locations on the court that players tend to shoot from having been given the regularization term as:$R_L(L) = \sum_{\ell \neq \grave{\ell}} K_{\ell,\grave{\ell}} || L_\ell - L_{\grave{\ell}} ||^2$ (5) , where $K_{\ell,\grave{\ell}}$ governs for the degree of similarity between ℓ and $\grave{\ell}$. And Equation (5) has the effect that is pushed to the spatial parameters to vary smoothly. The similarity coefficients k can be used to select any kernel function using as follow: RBF kernel: $K_{\ell,\grave{\ell}} = exp\left[\frac{|| L_\ell - L_{\grave{\ell}} ||^2}{\sigma}\right]$ (6)

where σ is a tunable parameter governing the degree of smoothness. To regularise all of their spatially strong factors, they usually (t)adopt a symmetric approach. This approach carries convergence to generative models that perform a Gaussian process prior to [66], [35], [33], Non-Negativity Constraints: their flexible space of model parameters C requires that all of the latent factor components be non-negative. They reasoned that the constraint makes their approach similar to related approaches based on non-negative matrix factorization, which are used to extract explainable models by [19], [20], and [21]. As such, non-negative latent factor models have also been used for analysis of shot charts of NBA players, as summarised in figure 3. In [5], factorized point process intensities are defined.

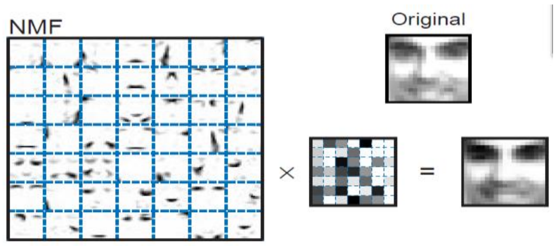

Figure 3: Nonnegative matrix factorization (NMF) learns a parts-based representation of faces .

They model player locations as positively influencing the probability of an event happening, and defensible locations as negating the influence. We have constructed a general inference framework wherein one can estimate continuous Markov transition densities in two steps: First, assume that sequential elements of the Markov chain (x_{t-1}, x_t) are realizations from a Poisson point process note that x_{t-1} and x_t can be vector valued and estimate the corresponding point process intensity function $\Lambda((x_{t-1}, x_t))$. Second, calculate the conditional transition density fixed x_{t-1}as follows: $\mathcal{P}(x_t, x_{t-1}) = \frac{\Lambda((x_{t-1},x_t))}{\int \Lambda((x_{t-1},\xi))d\xi}$ (7)

Our estimate of the transition density is a valid density, and variates from it can be easily simulated using any method that allows for sampling from a nonstandard distribution, i.e., rejection sampling, in [42]. Such as One of the simplest is the kernel density estimator introduced by [43] [63], where it is proposed that the intensity function be estimated as follows: $\tilde{\Lambda}$(s)≡ $\frac{1}{P_b(s)} \sum_{i=1}^{m} k_b\,(s - s_i)$ (8) where k_b (·) is a kernel function with bandwidth $\boldsymbol{\rho}_b(s)$, and is an edge correction that scales the intensity function to integrate to the appropriate count. While this kernel estimator is useful, in order to incorporate additional structural information, a more sophisticated approach is necessary; LGCP provides an elegant solution. It is a Poisson point process with (s) =$exp^{z(s))}$, where z(s), denotes the Gaussian process .We are interested in modelling the movement of the ball around the court. The movement of the ball is assumed to be a first-order Markov chain, with shots, fouls, and turnovers serving as absorbing states. A spatial model can be constructed by convolving a very simple process with a kernel point at the computational expense; a method that replaces the latent Gaussian process with a basis function expansion reliant on a smoothing kernel is described. Convoluting model for flexible space and space-time models [44] , The hidden-state at time *t* is used to predict the distribution of the trajectory position $(\hat{}x,\hat{}y)_{t+1}^{i}$at the next time-step *t*+1, they assume a bivariate Gaussian distribution parametrized by the mean $(\mu)_{t+1}^{i} = (\mu x, \mu y)_{t+1}^{i}$, standard deviation $(\sigma)_{t+1}^{i} = (\sigma x, \sigma y)_{t+1}^{i}$, and correlation coefficient $\boldsymbol{\rho}_{t+1}^{i}$, [9]. They had simplified computation structural information (LGCPs) are doubly stochastic, including integrated nested Laplace approximation (INLA), which is designed for latent Gaussian, is a computationally effective alternative to MCMC for Bayesian inference . INLA models, flexible class of models from linear mixed to spatial and spatio-temporal models [45], nearest-neighbors Gaussian processes, [46], and predictive processes, [47]. . We graphically depict our point process data, LGCP representation reconstruction in Figure 4.

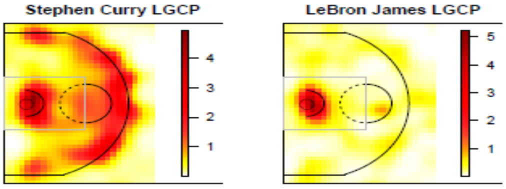

Fig 4. NBA player representations LGCP surfaces

In this paper, A target receiver and a pass result (successful or failed) are Some passes comprise the pass event of the player speed trajectory line in their model. Possession can be summarised by a player attempting to pass the ball to a teammate, who can record the occurrence of a pass event Fine-Grained Dynamic Spatial Models in [1] and [2]. The ball sometimes leaves the neighbourhood of a ball-possessing player; the ball goes out of bounds to players that can receive the ball into the possession of another player, either a teammate or not. Determining whether the ball is located where it can be intercepted by a certain player can take the process of a z-coordinate of the ball. The ball speed and trajectory near a player is a sharp modification, and it is the best indication of pass reception. However, passes that do significantly notice ball movement, which is required for undefined passes to modify moving in the direction, single-prediction models as multiple hypothesis prediction (MHP) models, and an associated meta loss function and optimization procedure to train in [41]. It is enough to treat an event that ends with a ball passing closer than a certain distance from the nearest goal as a shot target, in [5]. Human trajectory prediction in crowded spaces is addressed in [9] and spatial graph networks for pedestrian trajectory prediction [10]. [16]. temporal point process is a stochastic model that helps to capture the arrival pattern of a series of events. These include conducting temporal classification (CTC) architectures. [11] provides an example of a formalised CTC model conducting classification by generalising our coming from actual time stamps, but prediction methods progress from temporal values. A temporal point process is a stochastic model used to capture the arrival pattern of a series of events. The "conditional intensity" characterizes temporal point processes, which are tested in a variety of areas. [60], [38] for efficient prediction of Gaussian-process-modulated renewal processes with framework application to medical event data [65] describes the conditional intensities of several temporal point process models based on the Poisson Process. A temporal point process is a stochastic model used to capture the arrival pattern of a series of events. The "conditional intensity" characterizes temporal point processes, which are tested in a variety of areas. [60], [38] for efficient prediction of Gaussian-process-modulated renewal processes with framework application to medical event data. [65] describes the conditional intensities of several temporal point process models based on the Poisson Process. An unsuccessful indication of a pass event will be the ball being sent out of bounds. Such passes have to be recognised as shots on goal. Variables in the tracking data indicate when an action, such as a dribble or shot, occurs. A failed dribble indicates the player has lost control of the ball, and the other team rewards possession**.** The conditional intensity of a temporal point process (TPP) can be learned based on a network, and **TPP** is a stochastic model that helps capture the arrival pattern of a series of events, and the likelihood is maximised during training. These include conducting temporal classification (CTC) architectures. [11] provides an example of a formalised CTC model conducting classification by generalising our coming from actual time stamps, but training prediction methods progress from temporal values. They are interested in using supervised learning scenarios in training a predictor. f_θ : X → Y, parameterized by θ ∈ R_n,such that the expected error. $\frac{1}{N}\sum_{i=1}^{N}\mathcal{L}(f\,\theta(x_i),\ y_i)$ (9) is minimized, where it is assumed that the training samples follow $\mathcal{P}$(x, y). Here, L can be any loss function, e.g. the classical ℓ_2-loss. $\mathcal{L}_2(u,v) = \frac{1}{2}||u - v||_2^2$ (10) according to [41]. This can be seen in section (IV. B)

C. Multiresolution possession Model

Concepts similar expected possession value (EPV), where are modeled conditional on observed progress, have had statistical treatment in-game win probability in baseball .The present process model for a basketball possession requires some formalism by describing the full found **path planning** of a specific possession players. This enables us to represent the free space of all possible basketball possessions players derived from a stochastic process model for the evolution of a basketball possession. defined as X(ω)∈{ 0,2,3}as point score. Z is a high dimensional space that includes (x; y) coordinates for all players on the court,(x; y; z) coordinates for the ball can be possession path ω , define as Z (ω) .The optical tracking time-varying series generated by this possession players that $Z_t(\omega) \in$ Z, t > 0, is a catch of the tracking data exactly time t seconds from the begin of the possession time (t = 0), time–varying Poisson processes according to [26].The except can be defined as Ω that reperesnted as a sample space of possession paths, their model Z(ω) as a stochastic process, and likewise, denote by $Z_t(\omega)$ for all t > 0 as a random variable in Z. Z(ω) gives best filtration of the locations and trajectories of all players F_t^z = $\sigma(Z_s^{-1}$: 0≤ s≤ t), which represents all information can be the optical tracking data for the first t seconds of a model E[Y_i|F($X_{t,i}$)], the expected yards gained by the ball-carrier from their current position on the field conditional on the team possession for expected possession value (EPV) is the expected value of the number of points scored for the possession (X) and can be provided for all available data up to time t . The current possession expectation for players E[X|F_t^z] can take an integral over the distribution of paths(z). They used T(ω) denfine as the time at which a possession path ω , all the possession's point process is a deterministic function of the full resolution data at this time, X(ω) = h($Z_{T_\omega}(\omega)$). Let P(D_i|F($X_{t,i}$) be a probability mass function for the decision made by the quarterback on play i, a dropback, conditional on the locations and trajectories of all players and the ball over the course of play i up until time t, in [61] . Predicting the motion of offensive players over a short time window is driven by the players' dynamics (velocity, acceleration, etc.). Let the location of offensive player $\{\ell \epsilon 1, .. L\}$ at time t be Z^{ℓ}`(t) = ($X^{\ell}$`(t), Y^{ℓ}`(t)) then model movement in each of the x and y coordinates using Estimate EPV is the expected value of the number of points scored for the possession (X) given all available data up to time t. (F_t^z) amounts to integrating over the joint distribution of (T;Z_T).

Definition

: The expected possession value, or EPV, at time t $\geq$ 0 during a possession is $\boldsymbol{v_t} = \mathbf{E}[\mathbf{X}|\boldsymbol{F_t^z}] = \int_{\Omega}^{\infty} \boldsymbol{x(\omega)p}(\mathbf{d}\boldsymbol{\omega}|\boldsymbol{F_t^Z})$ .Where the Markov chain portion method requires populate the coarsened state space C a view of the data P($C_{\delta t}$ |F_z^t) , it is necessary to forecast the full-resolution data for only a short period of time relative to D. historical trajectory definition , as δt ≤ T, which our EPV estimate only depends on the coarsened state and enhancement have similarities to the coarsened possession process. To make the Markovian assumption plausible, the coarsened state space C with summaries of the full resolution data so that transitions between these states represent events in a basketball possession. with summaries of the full resolution shown in Figure 1 for an illustration.

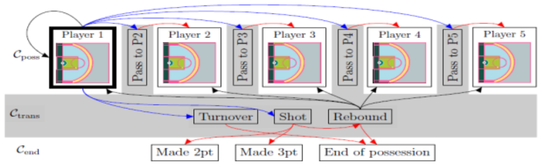

Figure 1: Schematic of the coarsened possession process C, have states rectangles state transitions arrows as s. The unshaded states in row1 compose Cposs. distinct ball handlers are grouped when Player 1 through 5, Ctrans, the rectangles in the row(3), C end [4].

As a step towards , We propose a method that captures the multi-modal behavior of players where they might consider multiple trajectories . The method is create an LSTM -based architecture predicting multiple trajectories and their probabilities, trained by a multimodal loss function that updates the best trajectories [9]. The locations of the players and the ball through NBA game with high accuracy and temporal resolution to reconstruct the trajectories of all players and the ball through an entire basketball game, which allows us to extract 2-D locations. $\ell_t^{\mathrm{P}} = [x_t^{\mathrm{P}}, y_t^{\mathrm{P}}]$ of player defined as P at time t, with p ∈ {1, . , 10},where x-coordinate The x-coordinate represents the length of the field, while the y-coordinate represents the width, with the general at the upper left corner. We are employing an ordered sequence of defining as previous L + 1 time steps that can be used to generate the p-th player's historical trajectory as $h_{\mathrm{t}}^{\mathrm{P}} = [`\ell_{\mathrm{t-L}}^{\mathrm{P}}, \ldots, \ell_{\mathrm{t}}^{\mathrm{P}}]$, where time steps are equally spaced at an interval at Δt. In the same time , they can be generated a historical trajectory of the ball as $h_{\mathrm{t}}^{\mathrm{b}} = [`\ell_{\mathrm{t-L}}^{\mathrm{b}}, \ldots, \ell_{\mathrm{t}}^{\mathrm{b}}]$. Let's assume that the group of players represents the team on offence and defense. We are interested in predicting the future trajectory of an offensive player, represented as a vector.$\tau_{\mathrm{t}}^{\mathrm{P}} = [`\ell_{\mathrm{t+1}}^{\mathrm{P}}, \ldots, `\ell_{\mathrm{t+H}}^{\mathrm{P}}]$, where H is the number of horizon for predict the trajectory of the time steps . Assume that the player of interest is on the offensive player for whom we are predicting trajectory is defined as player index P. The raw tracking data can be processed to build a labelled data set. D = {(u_t^{P} , τ_t^{P}), t = 1, . . . , T, P = 1, . . . , 5}, indicates that one data point is required for each time step and offensive player . t is the total number of time steps, input vector u_t^p = { h_t^{P} , $h_t^{-\mathrm{P}}$, h_t^{h} , s_t} is a set of historical player and ball trajectories, where h_t^{P} indicates history of the player of interest, $h_t^{-\mathrm{P}}$indicates histories of all other players, h_t^{h} and s_t is the shot clock denoted by the time in seconds remaining until the shot clock expires, in [70] concepts of the current location at time t changed trajectory to a velocity vector P by using a direct mapping of velocities to locations, $v_t^p = [v_{t+1}^p, \ldots \ldots, v_{t+H}^p]$ (11) where h∈ $[1, \ldots \ldots, H]$ $used\ for\ horizon, v_{t+H}^p = [v_{x,t+H}^p, v_{y,t+H}^p] = \left[\frac{x_{t+h}^p - x_{t+h+1}^p}{\Delta}, \frac{y_{t+h}^p - y_{t+h+1}^p}{\Delta}\right]$ (12) [9].

They can be displacement error (FDE) and average displacement error (ADE).

FDE = $\frac{1}{5T}\sum_{t=1}^{T}\sum_{p=1}^{5} || \ell_H^p - \acute{\ell}_H^P||$……………..(13)

ADE = $\frac{1}{5HT}\sum_{t=1}^{T}\sum_{p=1}^{5}\sum_{h=1}^{H} || \ell_t^p - \acute{\ell}_t^P||$……..(14)

As well, FDE considers the location error at the end of the prediction horizon H, but ADE averages location errors during the entire trajectory. MSE errors, defined as in equation (47) (see in section (IV-B) by using a Ground Truth trajectory prediction player (GTPP) that differs from the (FDE) and (ADE) that measure trajectory prediction errors, has been published by [41]

III. METHODOLOGY

A. The Framework Dynamic *Markov Decision Process*

Our framework for incorporating transition tensors , how to format an average chain from more than one independent Markov chain variable with overlapping state spaces ,in [32], probability policies are used to compute the nonstationary of these probabilities in time**,** and Markov decision processes (MDP) are used in more than one case in modern reinforcement learning problems to determine the interactions between an agent and the environment. In this research, we restrict our attention to finite MDPs, according to [32], which can be represented based on a tuple: <S;A; P(.);R(.)>, where S represents a discrete and finite set of states, A represents a finite set of actions the agent can take, P(.) denotes the transition probabilities between states, and R(.) denotes reward the agent receives for any given state/action . The agent operates in the environment according to a policy,π (.), which denotes as the probabilities that govern the agent's choice of action based on the current state of the environment. π (.) is only represented as the system control of the agent. defines these functions as following:

p(s, a,s')=P[$s_{n+1} = s', s_{n+1} = s, A_{n=a}$]

R(s, a)= E[$R_{n+1}, s_{n+1} = s, A_{n=a}$] (15)

$\pi(s.a)$ =P[$A_{n=a}|s_n = s$

(s_n; A_n; and R_{n+1}) defined as represent random variables , and n,n+1 italicized lowercase letters represent realized values of the corresponding random variables. Specifically , s_n , represents the agent's state in step n of an episode from the MDP, A_n represents the $A_{n=a}$ action and R_{n+1} is the sub sequent reward can be provided that is both of state\action . their concept for basketball ,they defined an episode as the sequence of events comprising a single play from begin to end. refer to [31] for large expansive provide to reinforcement learning and Markov decision processes. **Shot policy model π(.)** that is a given an arbitrary play from team and defined as n_{TPT} = 8, Our model of the probability that a given action an arbitrary of the play is a shot as a function of the ball carriers can be a shot given the current state of the MDP, our court region, defensive pressure, and the shot clock time interval:

$$\pi(s,a) =' shoot' \left| s_n^{(x,y,z)}, t_n, \theta \right) = exp\left(\theta_{t_n}^{(x,y,z)}\right) \quad (16)$$,

where A_n is a Bernoulli random variable for whether the n^{th} action of a play is a shot, $s^{(x;y;z)}$ denoted in [4], at any given moment, the state of a team's MDP is used the identity of the ball carrier, their court region, and an indicator of their defensive pressure and contested, t_n represents the interval of the shot clock at the n^{th} moment of the play, $\theta_{t_n}^{(x,y,z)}$ defines as player x's extract to shoot the ball y's region y when in court under defensive pressure z, and exp (.) is the inverse log it function: $\frac{\exp(.)}{1+\exp(.)}$. defines as that θ is performance for multi-level hierarchical priors that a large parameter matrix of dimension | S_{team}|-by-n_{TPT}. For a 15-player roster using 3-second shot clock intervals for the TPT

$$:\begin{cases} a ----- \theta^{(x,y,z) \sim N_8(\beta^{(G(x),y,z)}, \Sigma\theta} \\ b ---- \beta^{(g,y,z) \sim N_8(\gamma^{(y,z)}), \Sigma\beta} \\ c ------- \gamma^{(y,z)} \sim N_8(0, \Sigma\beta) \end{cases} \quad (17)$$

Predicting trajectory is difficult because an agent's target position is unknown, whereas G(x) returns the position type g of a player point(x), guard, and center. The first layer of the **hierarchical prior (LHP)** has a structure that virtually shrinks player x's the 8-dimensional state exactly for shot propensity paramete$\theta^{(x;y;z)}$ is given a multivariate normal prior with mean vector $\beta^{(G(x);y;z)}$, defines as the average shot propensity parameter vector for all players and estimated propensity who have the same position under all players $\{x_1'; x_2', \dots\dots, x_{Ng}'\}$ for 1, N for position g whom G(x_i') =G(x)). This effectively shrinkage player x's shooting analyzed shooting propensity toward players who can be share his position and the shrinkage is cleared when a player has less observed data, in [32]. The second (**LHP**) has virtually structure to (b-8) but here the position-exactly parameter vector (g; y; z) is given multivariate normal prior with mean vector $\gamma^{(y;z)}$, which defines The average proclivity to shoot in the court region y when under defensive pressure z, regardless of who is playing in position for the last layer of the (LHP), is almost identical to (c-8) The average proclivity to shoot in the court region y when under defensive pressure z, regardless of who is playing in position for the last layer of the (LHP), is almost identical to (c-8) where the 8-dimensional-vector as the prior mean, yielding a 0.5 prior probability of shooting in any region/defense given combination. While this is an unrealistic shot probability for all case states, their model is sensitive to all the values of this prior mean, given the amount of data. We have for each region/defense combination throughout the season, along with weakly informative priors on each level of the hierarchy. (a-8)-(c-8) is an AR(1) covariance matrix with variance and correlation parameters corresponding to the hierarchy's respective levels. (i.e. ρ_θ; ρ_β; ρ_γ, and σ_θ ; σ_β; σ_γ). [62].

***-Transition Probability Model** (TPM)* Conditional on a_n = `Not shoot', we model the probability that the play transitions to state $s^{(x',y',z')}$ as a function of the ball carrier's latent propensity to transition to state $s^{(x,y,z)}$ given current state $s_n^{(x',y',z')}$

$$\text{P(s,a, } s')\text{=P(}s_{n+1}\text{)= (} s^{(x',y',z')} \left| a_n s_n^{(x,y,z)}.\ t_n, \lambda \right) = \frac{\exp\left(\lambda_{t_n}^{((x,y,z),),(x',y',z')}\right)}{\sum_{i,j,k}\exp\left(\lambda_{t_n}^{((x,y,z),),(x',y',z')}\right)} \quad (18)$$

where s_{n+1} defines as random variable is a category and $\lambda_{t_n}^{((x,y,z),),(x',y',z')}$ defines as player x's propensity to transition to $s^{x'y'z'}$ when in court region y cross defensive pressure z. The symbols a_n, $s_n^{(x,y,z)}$, and t_n are all as defined as equation(14). The support of Sn+1 is used to the state space S_{team} , with one additional state Steam| + 1 the terminal state representing a turn over. λ is a massive 3-dimensional parameter array with dimensions |S_{team}| x (|Steam| + 1) x n_{TPT}. As with the models for π(.) and R(.), they are represented as multi-level hierarchical priors for λ refer to [4].

$$\begin{cases} \lambda^{(x,y,z),(x',y',\ z') \sim N_8(\zeta^{(G(x),y,z),(G(x'),y',\ z')}), \Sigma\lambda} \\ \zeta^{(g,y,z),(g,y',\ z') \sim N_8\left(\omega^{(y,z),(y',\ z')}\right), \Sigma\zeta} \\ \qquad \omega^{\left((y,z),(y',\ z')\right)} \sim N_8(0, \Sigma\omega) \end{cases} \quad (19)$$

In (19), $\zeta^{(g,y,z),(g,y',\ z')}$ denotes as the 8-dimensional average propensity of all players (i.e. all players $\{x_1'; x_2', \dots\dots, x_{Ng}'\}$ for whom H(x) = H(x_i')) with G(x) = G in court region y cross defensive pressure z to transition to all state with G(x0) = G', in court region y' cross defensive pressure z' over the 8 discrete intervals of the shot clock time. the position-specific parameter vector $\zeta^{(g,y,z),(g,y',\ z')}$ *that* given a multivariate normal prior with mean vector $\omega^{\left((y,z),(y',\ z')\right)}$, which can be used similarly term at the global region/defense level. As in the model for π(.), they use the 8-dimensional 0-vector as the prior mean for $\omega^{\left((y,z),(y',\ z')\right)}$ and all Σ is an AR(1) covariance matrix with variance parameters and correlation parameters corresponding to their respective levels of the hierarchy (i.e. i.e. $\rho_{\lambda\_}$; ρ_ζ; ρ_ω, and σ_λ ; σ_ζ; σ_ω). In broader **passing policy updates**(**PPU**) that include shooting as well as passing and dribbling. This includes altering the probabilities of non-terminating state transitions via the TPT [72].This allows us to define the **Reward function (R)** of the MDP completely in terms of a shot efficiency model prior to formally defining R (.), we propose a model for the probability that a shot is made. A model can be make-probability as a function of the shooter's skill and a region-specific additive effect if the shot was open:Make(s)=P$(M_{n=1} \left| s_n^{(x,y,z)} \mu, \xi \right) = expit(\mu^{(y,z)} + I(z_n =' open')x\xi^{y}$ (20) ,Where M_n is a Bernoulli random variable for whether the attempted shot at the moment n was made, $\mu^{(y,z)}$ denotes player x's contested shooting skill in court region y, I(.) is an indicator function of the defensive pressure z_n in moment n, and $\xi^{y}$is the effect in the court region y if the shot is uncontested. Note that in this model, defensive pressure is a region-specific additive effect. Rather than being built into the player-specific parameters. Detecting player-specific differences in how defensive pressure affects their shot-make probability is impracticable. It would take massive amounts of shot data to detect these differences with statistical confidence. The $\mu^{(x,y)}$ is given a normal prior with mean vector (H(x);y), which denotes the average shooting skill of all players who are share the same group as player x such as . all players $\{x_1'; x_2' \dots, x_{Ng}'\}$ for whom H(x_i') = H(x) in location y. Certainly that while P(.) and $\pi(.)$ denote by set of the player g using player positions such as center, forward, backward, point guard, this model employs new set of the player known as h, which is represented using regularization. The updated position of a player's shooting skills case because they don't have as clear a correspondence to their position. In the second layer of the hierarchical prior, is given a normal prior with mean vector *called* $\varphi^{(Y)}$, which defines the global average shooting skill from court region y. the final stage of the hierarchy is given a mean-0 normal prior with variance σ_φ^{2}'. This is represented as all the parameters in the model are provided as univariate priors, which differ from multivariate priors, since they model a player's shooting skill as being constant across the shot. They use half-Cauchy priors for the scale parameters in R(.): σ_μ; σ_φ; σ_ψ;'σ_ξ;half-Cauchy(0; 2:5): clock, by[62].MDP applications the transition $\pi(.)$ dynamics, P(.),can be corrected as being static, whereas R(.) is the reward function with time-independent assumed to vary temporally, and the shot clock is updated with each new episode of the process rather than continuing globally across temporal episodes. These are conventionally modelled statically to incorporate within-episode non-stationarity via a **transition tensor** matrix

B. Ball trajectory

A ball is thrown at a certain angle to the ground, its trajectory is parabolic, if it is thrown vertically up and down. The parabolic shape is a sequence based on Newton's laws. We proposed that the ball is launched at height h above the ground at angle θ and velocity v, the horizontal velocity component $vcos\theta$ does not change as there is no horizontal force, so we can be written for the x coordinate of the ball at time t define as Newton's first Law.

It is assumed that x = 0 during the ball is released at t = 0. Vertical movement happens under gravity with the gravitational acceleration g pulling the ball to the ground, so as a constant sequence of Newton's Second Law the y coordinate at time t can be written as following :x=$vcos\theta$ and $y = vtsin\theta - \frac{1}{2}gt^2$ (24), where $vsin\theta$ is the vertical component of the initial velocity and it is assumed that y(0) = h. $y = -x^2 \frac{g}{vcos\theta} + xtan\theta + h$ (25) where Equation 24 that can derive a velocity-angle relationship given the target point (x; y). the height of the

reward function RF (.) is simply the scaled make-probabilities for each state, if a shot is taken (scaled by 2 or 3, depending on the court-region), and 0 in the case that a shot isn't given , if a turnover occurs: R(s ,a)=$\begin{cases} 3xMake(s), y \in \{3-pointer\}, a_n = 'shoot' \\ 2xMake(s), y \in \{2-pointer\}, a_n = 'shoot' \\ 0 \quad otherwise \end{cases}$ (21)

As with P(.) and π(.),use a multi-stage hierarchical prior for the player-specific parameters μ

$$\begin{cases} \mu^{(x,y)} \sim N_8(\psi^{(H(x),Y)}, \sigma_\mu^2) \\ \psi^{(h,y)} \sim N_8(\varphi^{(Y)}, \sigma_\psi^2) \\ \varphi^{(Y)} \sim N_8(0, \sigma_\varphi^2 \end{cases} \quad (22)$$

(TPM) as an approach as tensors with matrix slices to dynamic transition probabilities function a continuous with temporal covariate, and shot policies, which is the shot clock time, in approach, which represents a transition probability effect via a simple indexing function T() defined as T(c_n) =$\begin{cases} 1 \dots\dots\dots c_n \in [0,1] \\ 8 \dots\dots c_n \in [21,24] \end{cases}$ (23)

Term c_n represents the shot clock time at the n_{th} moment of a play. Can be able to going forward is that represented as realized values of T(c_n) as time t , by [32] [69] is shown in Figure 2.

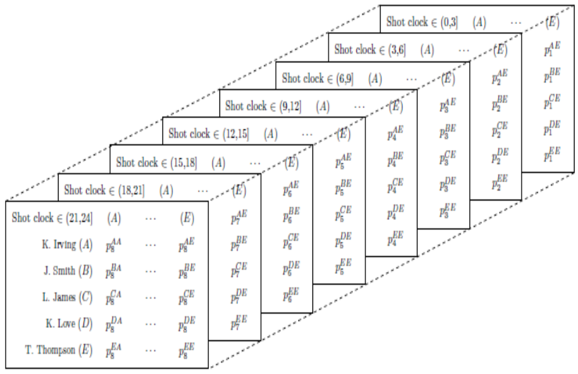

Fig 2: illustration's (row, column) state space of the TPT has condensed to 5-player states (200). Each slice represents an approach of the state-to-state transition probabilities during a second interval of the shot clock.

target is usually y= H, $v = \frac{x}{cos\theta}\sqrt{\frac{g\backslash 2}{xtan\theta + h - H}}$ (26) ,where Equation (26) clearly states that in order to bring the score, there must be a. Let r and R define the ball and. Otherwise, the back of the rim will be hit, and the ball may bounce away from the target. We assumed that the ball descends at an angle and neglected any modification in this angle while the ball is above the basket. $sin\alpha = \frac{r}{2R-r}$ (27) .To relate the angle of descent α to the minimum angle of throwing θmin ,as defined as:

tanα = $v_y \backslash v_x$, $tan\theta$min=$\frac{r}{2R-r}\left(1 - \frac{r}{R}\right) + \frac{2H-h}{d}$ (28)

A player of height h and releasing the ball at (h) is that standing a distance d from the basket at which minimum angle to throw .Since the raw ball trajectory data was computed trajectory from the start velocity by a significant amount, it was assumed that there were significant influences due to drag and lift forces on the ball , according to [39].

C. trajectory prediction the players' position sequences

Referring to [48], multiple visual features for tracking moving objects. Our proposed method consists of observation pattern-matching units and trajectory prediction the players' position sequences have been estimated from the corresponding video data set. We consider symmetries in the basketball games that exchange position. For each player's position sequence is determined by using a particle swarm undefined algorithm (PSO), in [50], has been used for its advantages in solving object tracking problems . In [51], a model-based, top-down solution to the problem of tracking the 3D position, orientation, and full articulation of the human body from markerless visual observations obtained by two synchronized RGBD cameras is presented by [49]. Inspired by recent advances to the problem of model-based hand tracking by algorithm PSO is a stochastic undefined technique that operates on a vector of real-value data. In that algorithm, PSO is used to search for the best a hypothetical solution to the problem, which is a candidate during a update process, it explores the search space impact by reversing a limited amount of information. resulting in the updated estimate of the object state can be determined

D. Fine-grained player locations.

The local decision model is a general approach that can be defined as variable y which have been using a multi-class Conditional random fields (CRF) also a limitation of maximum entropy Markov models (MEMMs) and other discriminative Markov models based on directed graphical models ,in [65]. Let Y (x) = {s, p1, p2, p3, p4,⊥ } define as the space may be intuitive can be given x, which corresponds to choice a shot, passing to one for each teammates, respectively. We can create model of the Intuitively, each $F(\dot{y}|x)$ can be translated as the log-odds of an event that is occurring in the game state defined as x. The distance between the ball handler and the basket Players are only predicting shooting closer to the basket more similarly. Players are more likely to pass the ball to teammates that are close to them. Therefore, one can create, for instance basic version of F(x,y)=

$$\begin{cases} F_s(x) \equiv \omega_s^T \phi_s(b, \ell_b) \\ F_p(i,x) \equiv \omega_p^T \phi_p(p_i, \ell_i, b, \ell_b) \quad (32) \\ F_{\perp}(y|x) \equiv \omega_{\perp} \quad \text{if } y = \perp \end{cases}$$

i) Model player shooting :We can use the simplest case of modelling, where we represent each player by using a Ks-dimensional latent factor. Certain players tend to shoot the ball. Similar to [5], discretize the court into cells fine enough to capture all significant spatial variations. We use notation and use it to refer to the factor of the cell that position belongs to score the tendency of player b to shoot at a location as L. Our use of latent factors follows from the assumption that the variability in how players shoot that can be well captured by a low-rank projection, which is similar to [8]. Since shot attempts for individual players are sparse. From the data, learn B and L consider two visions of the data. They can score the tendency of player b to shoot at location ` As $B_b^T \mathcal{L}_{\ell_b}$ Combining with the simple model in (32), our scoring function for shooting can be written as:$F_s^{\ell} = \omega_B^T \phi_s(b, \ell_b) + B_b^T \mathcal{L}_{\ell_b}$ (33)

through PSO algorithm by [52]. The current position x_i, velocity v_i $for\ each\ i_{th}$, and its best position p^{best} beside, the apples have entered to the best global position g^{best}, which has been done by using the swarm. The d_{th} components of the velocity and position player of each particle are updated by using the following*:*

$$v_{i,d}^{k+1} = x\left[v_{i,d}^k + c_1 r_{1,d}\left(p^{best} - x_{i,d}^k\right) + c_2 r_{2,d}\left(g^{best} - x_{i,d}^k\right)\right] \quad (29)$$

$$x_{i,d}^{k+1} = x_{i,d}^k + v_{i,d}^{k+1} \quad (30)$$

where c1, c2 are positive constants equal to 2.05,(r1,d),(r2,d) are taken from a uniform distribution in [0, 1] and *c* is a constriction factor. The selection of the particle position (p^{best}) and the best global position g_{best} are based on the fitness function value f (.).where ϕ_s and ϕ_p are feature mappings that characterize the distance between the ball handler and the basket and teammates, respectively and b is represented between 3-5 feet from the basket, and ω_p^T, wp and ω_s^T are parameters to be learned.

conditional probability of the label sequence data define as event y ∈ Y (x) as being log-linear and it's a response function . we denote the joint probability density over input variables and labels by p(x, y) = p(y|x)p(x), where p(y|x) denotes the conditional probability for the label y given the input x. F(y|x): P(y|x) =$\frac{1}{z(x|F)}$ exp{F(y|x)} (30),

where Z(x|F) denotes the standard partition function: Z(x|F) = $\sum_{\dot{y} \in Y(x)}$ exp{F($\dot{y}$|x)} (31).

ii) Modeling Passing: in our modeling ,where players are likely to receive passes, we consider two views of the data.First, how we model shooting?, we represent each player pi using a Kp-dimensional latent factor Pi, and each location using a Kp-dimensional factor M`. The score for player $p_i$ pi receiving a pass at location would then be $P_i^T$ $M_{\ell}$`.

$$F_P^{\ell}(i,x) = \omega_P^T \phi_P\left(P_{i,}, \ell_i, b, \ell_b\right) + P_i^T M_{\ell_i} + \mathbb{Q}_{1,\ell_b}^T \quad \mathbb{Q}_{\ell_i} \quad (34)$$

can be learning by P,M,$\mathbb{Q}$1, $\mathbb{Q}$2 $from\ data$

-MDEL position

In our motivating application, we consider the problem of predicting future game events (y) given the current game state (x). Our approach to using basketball as a assumes the following information is provided in every game: x:i)the identity b and location `b of the ball handler ii) the identities and locations $\{(p_m, \ell_m)\}_{m=1}^4$ m=1 of the ball handler's s teammates.iii) The locations $\{\tilde{}p_m, \tilde{}\ell_m)\}_{m=1}^5$m=1 of the five opponents iv) The amount of time b the ball handler has possessed the ball.We are interested in predicting near-term events from the ball handler as follows: a) Passing to teammate pi within the next t seconds and ii) Shooting the ball within the next t seconds b) Maintaining possession of the ball as multiple events occur within the next t seconds (e.g., the ball handler passes to any player, a teammate shoots immediately),given a training set S = $\{(y_i, x_i)\}_{i=1}^n$, the goal is to learn a model that can accurately predict the correct y given x. our approach is to model the response variable y using a multi-class conditional random field [65]. Let R(x, ℓ_1, ℓ_2) denote a function that computes the relative positions of the defenders with respect to locations ℓ_1 and ℓ_2.Specifically, R(x, ℓ_1, ℓ_2) performs the following operations: c) rotates the coordinate around its position ℓ_1 until the direction $\ell_1 - \ell_2$is the upward direction for the system .

d) its re-centers the coordinate around applies this transformation to the locations of each defender, and outputs our updated locations for the system. Consider Figure5, where the ball handler Parker is being defended by Williams.

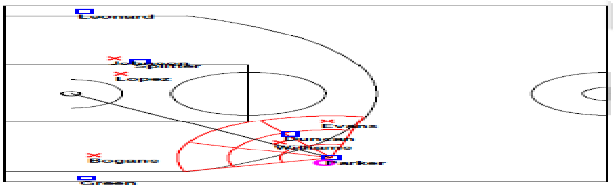

Fig. 5. coordinate systems of defenders are transformed to relative to the ball handler

-Modeling Shooting: We consider modelling the positions of defenders relative to the ball handler, which can suppress the probability of taking a shot. Let ˜D $_{\ell}$ˇ denote a Kd-dimensional latent factor corresponding to the relative location the ball handler, and let ˜L$_{\ell_b}$ denote a factor representing the position of the ball handler. The scoring function for how defenders impact shooting follows as: F_s^d(x) = $\sum_{\ell \in R_b(x)} [\mathcal{V}_{1\tilde{\ell}}\ \tilde{} D_{\tilde{\ell}}^{\tau}\ L_{\ell_b}] + \omega_{s,d}^{\tau} \phi_{s,d}$ (35) computes a set of distance features, such as the distance of the closest defender to the ball handler. Intuitively, the way which opponents should vary depending on location defend a player . such as , opponents tend to defend much more closely when the ball-handler is closer to the basket. Our final scoring function for shooting events can now be follows as

F_s (x)= $F_s^{\ell}(x) - F_s^d(x)$ (36) $F_s^{\ell}(x)$ defined in (33).

We consider modelling the positions of defenders relative to the ball handler, which can reduce the probability of a shot being taken. We merge spatial priors and non-negative matrix factorization in order to create a summary of basketball modelling shooting probability.. Let $\tilde{}D_{\tilde{\ell}}$ denote K_d- dimensional latent factor that is corresponding to the relative location $\tilde{\ell}$. The ball handler, and let $\tilde{}L_{\ell_b}$ defined as a factor representing the position of the ball handler. We've written into the scoring function for how defenders affect shooting: F_s^d(x) = $\sum_{\ell \in R_b(x)} [\mathcal{V}_{1\tilde{\ell}}\ \tilde{} D_{\tilde{\ell}}^{\tau}\ L_{\ell_b}] + \omega_{s,d}^{\tau} \phi_{s,d}$

IV. Adaptive to optimal policy

A. Trajectory Prediction Based on Ground Truth Estimation

We propose a method that captures the multi-modal behavior of players, where they might consider multiple trajectories and select the most advantageous one. Each cell contains all points that are closest to its generator the Voronoi of the label space Y = $\cup_{l=1}^{M} \bar{Y}_l$ which is induced by M generators $f_\theta^j(x_i)$, predictions j ∈ {1, . . . , M} for each training sample (x_i, y_i) by a forward pass though the network. Compute gradients for each Voronoi cell $f_\theta^j(x_i)$, yi), $\frac{\partial}{\partial\theta|y_i|} \sum_{y_i \in Y_i} \mathcal{L}(f_\theta^j(x_i), y_i)$ (43) , where $|Y_i|$ denotes .

(37) where R_b(x) ≡ R(x,ℓ_b , ℓ_0), $\mathcal{V}_{1\tilde{\ell}}$ is called a global parameter independent of the location of the ball handler, and $\phi_{s,d}$ (x). calculates a group of distant features, e.g. the distance of the closest defender to the ball handler. Prediction , in which opponents defending a player will be vary depending on location.

-Modeling Passing: We next consider modeling the positions of defenders relative a potential pass recipient teammate, which can suppress the probability of the ball handler passing that teammate the ball. We consider two sets of views or coordinate systems here. The first view models defenders relative to the direction of the pass, and the second view models defenders relative to the direction of the basket (which is similar to the defender model for shooting Where ˜C$1_{\tilde{\ell}}$, and ˜ $C2_{\tilde{\ell}}$, define as (the latent factors) representing the relative position $\tilde{\ell}$ˋ of the defender can be used for the two assumes. Let $\tilde{}M_{1\ell}$, and $\tilde{}M_{2\ell}$,ˋ define as the latent factors representing the location of the pass recipient team the ball on the court. We have the scoring function that showns how defenders effect a team from the players who can be receiving a pass as following :$F_p^d(i,x) = \sum_{\tilde{\ell} \in R_1(i,x)} \tilde{}\mathcal{V}1_{\tilde{\ell}} + \tilde{}C1_{\tilde{\ell}}\tilde{}M_{1\ell} + \sum_{\tilde{\ell} \in R_2(i,x)} \tilde{}\mathcal{V}2_{\tilde{\ell}} + \tilde{}C2_{\tilde{\ell}}\tilde{}M_{2\ell} + \omega_{p,d}^{\tau}\phi_d(i,x)$ (38)

where ˜$\mathcal{V}1_{\tilde{\ell}}$,˜ˋ and ˜$\mathcal{V}2_{\tilde{\ell}}$ˋ are global parameters, and ϕ_d a feature function that calculates a set of distance features, such as, the distance between the teammate and the defenders with the ratio of the passing distance to the defender's distance. We have got results in the scoring function for passing events:

$F_p(x) = F_p^{\ell}(x) - F_p^d(x)$ (39). Defined $F_p^{\ell}(x)$ *in* (34)

iii)Model Modeling Duration of Possession

How player behaviour variety depending on duration of possession? Such as, at some locations, a player may often choose to immediately shoot or pass upon receiving the ball. A a player may often choose to keep the ball for a few seconds before shooting or pass upon receiving the ball. Where $T_{s,\tau}$ $T_{p,\tau}$ and $T_{\perp,\tau}$ *is* denoted a latent factors corresponding to the ball handler's having had possession τ time steps prior to the current state. where V_ℓˋ is define a latent factor representing the location of the current ball handler. As well as, our the contribution can be appear duration of possession to shooting, passing that can be holding on to the ball.

, $F_p^{\tau} = T_{p,\tau b}{}^{\tau}V_{\ell b}$

$F_{\perp}^{\tau} = T_{\perp,\tau b}{}^{\tau}V_{\ell b}$ (40)

as: $F_s^{\tau} = T_{s,\tau b}{}^{\tau}V_{\ell b}$

the cardinality of Yi. We can be update step of $f_\theta^j(x_i)$, using the gradients per hypothesis j from before this state.$\int_x \sum_{j=1}^{M} \int_{y_i(X)} \mathcal{L}(f_\theta^j(x), y)\, p(x,y) dydx$ (44)

, the closeness is defined by the loss L. Thus, (52) divides the space into M Voronoi cells generated by the predicted hypotheses $f_\theta^j(x_i)$ and aggregates the loss from each. In a typical regression case L is chosen as the classical ℓ2-loss it operates on top of any given standard loss L:

$\mathbf{M}(f_\theta(x_i, y_i) = \sum_{j=1}^{M} \delta(y_i \in Y_i(\mathrm{xi}))\mathcal{L}(f_\theta^j(x_i), y_i)$ **(45)**

We use the Kronecker delta δ that returns 1 when its condition is true and 0 otherwise, in order to select the best hypothesis $f_\theta^j(x_i)$ for a given label y_i. to gradient descent methods used for training with back-propagation. One simple way to transform an existing network into a MHP model is to replicate the output layer M times . During training, each of these M predictions is compared to the ground truth label based on the original loss metric but weighted by δ as the meta loss suggests (Equation (45)). Similarly, during backpropagation ,δ provides a weight for the resulting gradients of the hypotheses. from the target labels y that all y lie in a single Voronoi cell k. In that case only the k-th generator f_θ^k (x) gets updated since $\delta(y_i \in Y_i(x_i))$ = 0, ∀j ≠k.. We relax the hard assignment using a weight 0 < α < 1. calculated as follows: $\delta_\epsilon(\alpha) = \begin{cases} 1-\epsilon & \text{if condition } \alpha \text{ is trure} \\ \frac{1}{M-1} & \text{otherwise} \end{cases}$ (46)

As well as , δ_ϵ is a relaxed [46] brining the most weight to the best matching trajectory, while A label y is now assigned to the closest hypothesis $f_\theta^k(x_i)$, weight $1-\epsilon$ and $\frac{1}{M-1}$ the remaining hypotheses. The term is $v = f_\theta^j((x_i), y_i)$ given a ground truth trajectory and $\grave{v} = f_\theta^j(x_i), y_i)$ given predicted trajectory $\dot{v}$, define the trajectory loss function $\mathcal{L}$ as follow:

$\mathcal{L}^{MSE}(f_\theta^j(x_i), y_i) = (v, \dot{v}) = \frac{1}{2H}||v - \dot{v}||$ (47) ,Where MSE is defined as a mean square error, which is known as the predicted velocity vector. We propose a variety loss function that encourages the network to produce diverse samples. For each scene ,we generate k possible output predictions (decode) by randomly sampling **z** from $\mathcal{N}(0, 1)$ and choosing the "best" prediction the coordinates ($\hat{x}_i^t$, $\hat{y}_i^t$) Social" context is generally provided as input to the LSTM cell , when the pooled context only once. in $\mathcal{L}_2$ sense as their prediction as $\mathcal{L}_{variety=\frac{min}{k}} |y_i - \hat{y}_i^k|_2$ .In order to train a model to predict multiple trajectories and their probabilities approach on an adaptation of the multi-modal loss function presented, in [40]. A similar loss function for Multimodal trajectory predictions is used by [58] to generate multi-modal pedestrian trajectories. The Multiple-Trajectory Prediction (MTP) as losing for time step t and player P, it comprises a linear [59].

$$\mathcal{L}^{MTP} f_\theta^j((x_i), y_i) = \sum_{m=1}^{M} \delta_\epsilon (m = m^*) \log \hat{P_m} + \alpha \mathcal{L}^{MSE}(v_t^p, \dot{v}_t^p) \quad (48),$$

Where is $\hat{P_m}$ an produce of a soft max, α is excessive -parameter used to drop one in order to possessing one of the classifies and trajectory losses, and m^* is the index of the winning mode that can be produced the trajectory closest to the ground truth, calculated according to a distance function :distance ()

$$, m^* = \underset{m \in [1,,,M]}{argmin}\ dist\left[v_t^p, x_{t,m}^p\right] \quad (49)$$

where, m^* simply defined as a path that is closer to the truth and has the lowest trajectory loss, multiple closeness mode measures to consider [40].

$$dist_{MSE}(v, \dot{v}_m) = \mathcal{L}^{MSE}(v, \dot{v}_m) \quad (50)$$

where other distance functions, as choice has a large effect on the model performance. Anyway, they considered the distance function with the smallest cross-all displacement error, defined as a location error at the last time step and calculated as: $dist_v(v, \dot{v}_m) = ||\ v_{t+H}\ , \dot{v}_{+H,m}\ ||_2^2$ (51) ,

The error of final player velocity, which can be translated as the player's, as shown in [58], as followed: $dist_v(v, \dot{v}_m) = ||\ v_{t+H}\ , \dot{v}_{+H,m}\ ||_2$ (52) **.**

B. Training

Training model by using gradient descent. To summarize, we only present derivations for two of latent factors. The partial derivatives can be derived similar ,l can begin with the factor's partial derivativeL_ℓ, which models whether the ball handler is similarly to shoot at location as followed:$\frac{\partial \mathcal{L}}{\partial L_\ell} = 2\lambda \sum_{\ell \neq \grave{\ell}} \left(L_\ell + K_{\ell,\grave{\ell}} ||L_\ell - L_{\grave{\ell}}\ ||^2\right) - \sum_{i \in I_\ell^b} \frac{\partial log P(y_i|x_i)}{\partial L_\ell}$ (53), where I_ℓ^{b}` defines as training data indices where the ball handler that can be denoted as ℓ. For latent factors that don't have spatial regularization , there doesn't specify for a similar term for $K_{\ell,\grave{\ell}}||L_\ell - L_{\grave{\ell}}\ ||^2$.where b_i define as the ball handler in the i-th training. The partial derivative of conditional probabilities is :

$$\frac{\partial log P(y_i|x_i)}{\partial L_\ell} = \frac{\partial log P(y_i|x_i)}{\partial L_\ell} = \left(\mathbf{1}_{[y_i = b_i]} - \boldsymbol{p}(\boldsymbol{b}_i|\boldsymbol{x})\right) \frac{\partial F_s(x)}{\partial L_\ell}, \quad \mathbf{(54)}\ ,$$

where $\frac{\partial F_s(x_i)}{\partial L_\ell} = \boldsymbol{\partial B_{bi}}$. first condition is an indicator function that givens value 1 if condition is true, and 0 otherwise. We used the last observe spaces as well as reward functions that All partial derivatives for parameters can be calculated in a shooting event, which can be derived similarly. We can derive from the partial derivatives of M_ℓ, which models whether a teammate at location ℓ is similarly to receive a pass:

$$\frac{\boldsymbol{\partial \mathcal{L}}}{\boldsymbol{\partial M_\ell}} = \mathbf{2\lambda} \sum_{\ell \neq \grave{\ell}} \left(\boldsymbol{M_\ell} + \boldsymbol{K_{\ell,\grave{\ell}}} ||\boldsymbol{M_\ell} - \boldsymbol{M_{\grave{\ell}}}\ ||^2\right) - \sum_{i \in I_\ell^P} \frac{\boldsymbol{\partial log P(y_i|x_i)}}{\boldsymbol{\partial M_\ell}} \quad (55)$$

Where I_ℓ^{P}` defines as the training data indices and a teammate can be defined as ℓ.Where p_i defines the teammate in the i_{th} training that *can be defined as* ℓ . We can write the partial derivations can be written each training as:

$$\frac{\partial log P(y_i|x_i)}{\partial M_\ell} = \left(1_{[y_i = p_i]} - p(p_i|x)\right) \frac{\partial F_p(p_i, x)}{\partial M_\ell}, (56)$$

Observed spaces as well as reward functions that: $\frac{\partial F_p(p_i,x)}{\partial M_\ell} = P_{pi}$. All partial derivatives for parameters that can be determined in a passing event can be derived similarly. The adaptive stochastic gradient descent method is enhanced by [59] .Learning criteria are described as being optimized during model fitting, which is defined as all of the parameters of their model depending on the number of factors. We formulate a discriminative learning objective to minimise a trade-off between the negative conditional log likelihood of the training data and a regularization term controlling the complexity of their model:

$$\underset{\Theta \in C}{argmin}\mathcal{L}(\Theta) \equiv -logP(y_i|x_i)\lambda R(\Theta) \quad (57)$$

where C denotes the feasible space of model parameters, denotes R(Θ) the regularization function, and λ is a hyper parameter that trades off between the two. As define R(Θ) using two components that decompose additively. The first component is the standard squared 2-norm that uses small parameter weights . The second component is a spatial regularization term that encourages the parameters of nearby cells to be similar to each other. Our learning objective which is non-convex is particularly susceptible to local optima, and running gradient descent algorithms from random initializations leads to bad solutions. As such, finding a good undefined is key to find a good initialization, we first train our latent factor models as a full rank matrix of parameters. For example, instead of modelling the player by location shooting factors B and L as separate factors, we instead directly estimate the full player-by-location matrix of parameters B=$B^{\perp}L^{\sigma}$.Let b_i denote the ball handler in the i-th training instance, and let I_ℓ^b denote the training indices where ball handler was in location ℓ.the partial derivative as defines:

$$\cdot\frac{\partial\mathcal{L}}{\partial B_{b\ell}} 2\lambda\sum_{\ell\neq\grave{\ell}}\widetilde{B_{b\ell}} + \left(K_{\ell,\grave{\ell}}||\widetilde{B_{b\ell}}-\widetilde{B_{b\grave{\ell}}}\ ||^2\right) - \sum_{i\in I_\ell^P}\frac{\partial logP(y_i|x_i)}{\partial\widetilde{B_{b\ell}}} \quad (58)$$

and

$$\sum_{i\in I_\ell^P}\frac{\partial logP(y_i|x_i)}{\partial\widetilde{B_{b\ell}}}=\left(1_{[y_i=b_i]} - p(b_i|x)\right)\frac{\partial F_p(p_i,x)}{\partial\widetilde{B_{b\ell}}} \quad (59) \ where \ \frac{\partial F_s(x_i)}{\partial\widetilde{B_{b\ell}}} = 1$$

We used partial derivatives for merging factors, which can be derived similarly. it was initially decoupled. Since each row of B with other rows of B, the interactions should be between players. The resulting unspecified problem is convex and can be solved optimally, at the expense of the learned model being statistically with confidence due to each player/location pair having some data points. Performed for non-negative matrix factorization to recover the best initialization of B and L, before learning $\bar{B}$. the approach as follow in [5], Their objective is to find a B and L in order to minimize the matrix divergence, which discourages large ratios between B and the recovered factors L, modification approach from [37] .

C. Factored Markov Decision Processes with Unawareness

We propose a system where an agent makes explicit attempts to include its unawareness time and deal with sequential decision. A method to learn factored Markov decision problems from both domain exploration and expert assistance ,which guarantees convergence to near-optimal behavior, even when the agent begins unaware time of location factors critical to success. The problem can be created because of the uncertainty in observation space is very large. Using a dynamic Bayesian network (DBN) to generate spatial and temporal abstractions capable of high-level decision making , where the transition and reward functions (RF) of an FMDP are defined [53] . The optimal policy is a tractable method to learn an FMDP through cross-dynamic programming , specified to join to learn a structured model of the transition and reward prediction functions for the event of a pass or shot, in [54] and the active learning problem of inferring the transition , used by [55]. It is only defined as MDP with unawareness defines as (MDPUs) . Their experiments show our agent learns optimal behavior on small and large problems that is beginning while unaware of more than one action ,propose a system where an agent makes explicit attempts to include its unawareness with sequential decision problems, in[56],[57].

$$V_\pi(s) = R(s) + \gamma\sum_{S\in s0}(P(s0|s;\ \pi(s))V\pi(s') \quad (60)$$

$$Q_\pi(s,\ a) = R(s) + \gamma\sum_{S\in s0}(P(s0|s,a)V\pi(s') \quad (61)$$

If T and R are known, we can compute $\pi +$ via value iteration , we can measure the expected loss in discounted return of following policy π versus π + using (69), which we refer to as the policy error. If the agent's policy is unknown, we can approximate the policy error using (70), where P(s0) is the probability of starting an episode in s0:$Err(\pi) = \sum_{S\in s_s} P(s_0)\left(V_{\pi+}(s_0) - V_{\pi+}(s0)\right)$ (62)

$$Err(t;\ t + k) = \sum_{S\in s0} P(s_0)(V_{\pi+}(s_0) - \frac{\sum_{i=t}^{t=t+k} G^i}{k} \ , (63)$$

G^i is discounted return for episode n

Appendix

Algorithm 1 for detecting possession updates

Require: ball location R(x, ℓ_1, ℓ_2) enters neighborhood of a non-ball-possessing player p_0

1:Possession Changed from shooting clock to pass or update using question (16-22)

2: concurrent $\ell_t^{b`}$ = [x_t^b , y_t^b], where ℓ *location*, *b ball* , *t time*

3: historical trajectory of the ball as h_t^b historical trajectory of path offensive player ,as a vector τ_t^pusing equation(11,12)

4: vector updated shot clock to pass by Transition Tensor MDP: using equation (23)

5:if vector updated (v_t)> k_b (·) is a kernel function with bandwidth $P_b(s)$, and is an edge correction that scales the intensity function to the appropriate count then using equation (8)

6: The updated estimation the position of a player's shooting skills case using equation (29)

return true

7: Previous Dir calculate ball direction (v_{t-1}, v_{t+1})

8: next Dir calculate ball direction (v_t, v_{t+1})

9: next Dir + previous Dir > Min True Change Angle Ball using equation (34-28)

10: Trajectory Prediction Based on Ground Truth Estimation using equation (43-47)

return true by used conditional probability density using equation(30-40)

11: if ball is still possessed using equation (5,6)

12:return true density estimation used non-negative latent factor models analysis of shot charts using equation (7,8)

13: return false doesn't change in ball possession using equation (9,10)

Algorithm 2. Detecting passes and shots Player .

1:function Detection Pass or Shot ()
2: if $Ball_{y_j}$ > 34.0 then . ball crossed the touchline for (Modeling Passing)
3:n ←0;
4;while $s_{n,}$ ≠Turnover do $t_{n,}$ T($c_{n,}$) ;
5 ; an Bernoulli variate from $\pi(.|\theta, \text{sn}, t_{n,})$;if an = Shot or no shot then
6 rn+1 ←R($s_{n,}a_n|\mu'\zeta'$)
7 break loop using equation (33)
8: return VERIFY FAILED PASS( ) rn+1 ← 0 then true using equation
(3,4) the discretized intensity function
If ground- fail trajectory predicted between two position time using equation (48)
9: if $Ball\,^{j}_{x}$ > 52.5 then ball reached the goal line where j represented base on time
10:if DISTANCE(Bally, 0) < Goal post Distance + Goal Length/2 then using equation (36,39)
11: else data P($C_{\delta t}$ |F_z^t) , the multi-resolution data for only a short period of time relative to definition , as δt ≤ T, which our EPV estimate only depends on the coarsened state .
12: return VERIFY FAILED the passing distance to the defender's distance using equation (51)
13: return shot on goal
14: if ball is within vicinity of another player p0 then
15:if Possession changed( ) is false then using equation (48)
16: return no event detected so we used predicting multiple trajectories and their probabilities, using equation (49)
17: if ball is within vicinity of another player p0 then
18: return successful pass using equation (40)
19: else un successful pass using equation (13,14)
20: return VERIFYSHOT( ) using equation (39,40)
21:functionVERIFY FAILEDPASS() using equation (65)
22:pass distance is longer than Min Failed Pass Length equation(32)
18: return unsuccessful pass . using equation (50) scoring function for passing
19: else using equation (56)
20: return no event detected using equation (49)
21: function VERIFYSHOT( ) using equation (58,59)
22: *ball trajectory line crosses the goal line using equation (27,28)
23: Goal post Distance from the goal post then
24: return shot on goal using equation(39,40)
25: else using equation (47)
26: return VERIFYFAILEDPASS( ) using equation (59)

Algorithm 3 Learning FMDPs with Unawareness

function LEARNFMDPU(A0, X0, T0, Q0, V0, s0)
2: for t = 1 , max Trials do
3: $(s_t; r_t)$ ← ε-GREEDY (s_t, Q_t, adv0:t-1) using question (69) and (70)
4: T_t; R_t Add s_t; r_t using below question

$$p(p^a_{a\,x'}) = \prod_{j\in(p^a_{a\,x'})} \frac{\beta(N^a_{1|j}+\alpha^a_{1|j},\dots\dots N^a_{m|j}+\alpha^a_{m|j}}{\beta(\alpha^a_{1|j},\dots\dots,\alpha^a_{m|j})} \quad (1)$$

$$p(p^a_{a\,x'}) = \rho^{|p^a_{a\,x'}|}(1-\rho)^{|x|-|p^a_{a\,x'}|} \quad (2)$$

$$P(DT^a_x|p^a_{a\,x'}\, D_{0:t}) \propto p(DT^a_x)p(D_{0:t}|DT^a_x) \quad (3)$$

$$p(D_{0:t}|DT^a_x) = \prod_{\ell\in Leaves\quad DT^a_x} \frac{\beta(N^a_{1|\ell}+\alpha^a_{1|\ell},\dots\dots N^a_{m|\ell}+\alpha^a_{m|\ell}}{\beta(\alpha^a_{1|\ell},\dots\dots,\alpha^a_{m|\ell})} \quad (4)$$

branch label (ℓ) of ball and players

$$E(\theta^a_{x=i\,\epsilon\,time}, p^a_{a\,x'=j}\,|D_{0:t}, DT^a_x) = \frac{N^a_{i.j}+\alpha^a_{i.j}}{N^a_{.j}+\alpha^a_{.j}} \quad (5)$$

& DT^a_x restricting node tests to members of PaaX0 is a directed acyclic graph with nodes $\{x_1, x_2\dots\dots, x_n, x'_1, \dots . x'_{2\dots}, x'_n\}$
5: if Update to R_t fails then
6: Z Ask expert ?$\lambda(x\epsilon x^+ \Lambda\, s_{t-k}[X] \neq s_t$ (6) multiple variables in multiple variables in X+ ,X whose assignments differ in s_{t-k} and s_t
7: scopet (R); X^t) Append Z to each
8: R_t ← Update via must be rechecked to see if a replacement test would yield a tree with better information gain
9: if t-$t' > \mu$
$Err(n', n) > \beta \vee m > k$
$\exists a' \in A^+, Q_{\pi+}(S_{m,n,a'}) > Q_\pi(S_{m,n,a_{m,n}})$ (7)
are true then Equation (7) ensures some minimum time μ has passed since the expert last gave advice. Equation (7) ensures the expert won't interrupt unless its estimate of the agent's policy error in (44) and (45) is above some threshold β,
10: advt ← Expert advice of form $Q_{\pi+}(\omega^s{}_{m,n}, a') > Q_{\pi+}(\omega^s{}_{m,n,}a_{m,n})$ (8)
11: if advt mentions action $a' \neq A^{t-1}$ then
12: $A^t \leftarrow A^{t-1} \cup [\alpha']$
13: $\mathcal{T}$t ← $\mathcal{T}$t-1 ∪ DBN_α made via $p(p^a_{a\,x'}) = \rho^{|p^a_{a\,x'}|}(1-\rho)^{|x|-|p^a_{a\,x'}|}$ (2)
14: if adv0:t-1 conflicts with advt then
15: Z Ask expert (6)
16: $\mathcal{X}^t \leftarrow \mathcal{X}^{t-1} \cup \{Z\}$ (9)
17: if $\mathcal{X}^t \neq \mathcal{X}^{t-1}$ then
18: $\mathcal{T}$t Update via (3,5,7,910,11)

$$\alpha^a_{x=i}|Y=j := \begin{bmatrix} \frac{K}{|v(z\cup Y)|} & if\ X=Z \\ \frac{K}{|v(z\cup Y)|}P(i,j[Y\backslash Z]dbn^t_a) & else \end{bmatrix} \quad (10)$$

$$p'(DT^a_{X'}|p^a_{a\,X'}) \propto \prod_{\ell\in Leaves\ (DT^a_{X'})} \beta(\alpha^a_{X'=1|\ell},\dots\dots\dots,\alpha^a_{X'=n|\ell}) \quad (11)$$

Equation (10) summarizes $D_{0:t}$ via inferences on the old best Dynamic Bayesian Network (DBNs), then encodes these inferences in the new parameters to generate spatial and temporal. The revised _-parameters ensure the new tree structure prior and expected parameters defined via (11)
19: $(V_t; Q_t) \leftarrow (R_t; T_t; V_{t-1})$ Increase Once our agent has a transition and reward tree, we can then use Incremental structured value iteration [56].

ACKNOWLEDGEMENTS

The authors would like to thank for the all help for us

V. Conclusion

We designed a rule-based algorithm for adaptive event detection as a spatiotemporal soccer dataset based on Algorithm 1. (DBN) to generate spatial and temporal abstractions capable of high-level decision making based on Algorithm 3. To learn factored Markov decision problems (FMDP) from both domain exploration and expert assistance, which guarantees convergence to near-optimal behaviour even if the agent has started to learn factors critical to success. We can investigate the problem of segmenting sets of low-level time-stamped events into time-periods by using a combination of Poisson models and Bayesian estimation methods and Markov decision processes (MDPs) with shot clock dependent transition probability tensors. To borrow strength across players and through time, Bayesian hierarchical models are employed in the modelling and parametrization of these tensors. Any changes to the list of supported events may necessitate a code update using A minimal change in trajectory direction is considered a reward function that is time-independent. The mean squared error (MSE), which is defined as the predicted velocity vector, then, in order to train a model to predict multiple trajectories and their probabilities.

REFERENCES

[1] Y. Yue, P. Lucey, P. Carr, A. Bialkowski and I. Matthews, "Learning Fine-Grained Spatial Models for Dynamic Sports Play Prediction, " in procc: *IEEE International Conference on Data Mining*, 2014,pp.670-679.

[2] B.Skinner and M. Goldman, *Optimal strategy in basketball In Handbook of Statistical Methods and Analyses in Sports*,2019 . [E-book]

[3] M. Horton, J. Gudmundsson, S. Chawla and J. Estephan , "Automated classification of passing in football, "in proc Intern Conference on Knowledge Discovery and Data Mining, Ho Chi Minh,Vietnam , 2015; pp.319–330.

[4] D.Cervone, A.DAmour, L.Bornn,. and K.Goldsberry ,A multiresolution stochastic process model for predicting basketball possession outcomes, Journal of the American Statistical Association , vol .111, pp.585-599 , 2016.

[5] A. Miller, L. Bornn, R. Adams, and K. Goldsberry, "Factorized point process intensities: A spatial analysis of professional basketball," in International Conference on Machine Learning , vol 32, 2014.pp. 22-24.

[6] V. Guralnik and J. Srivastava, "Event detection from time series data," in *Proc: fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA:ACM Press, 1999, pp. 33–42.

[7] M. Salmenkivi and H. Mannila, "Using Markov Chain Monte Carlo and dynamic programming for event sequence data," *Knowledge and Information Systems*, vol. 7, no. 3, pp. 267–288, 2005.

[8] S. L. Scott and P. Smyth, "The Markov modulated Poisson process and Markov Poisson cascade with applications to web traffic data," *Bayesian Statistics*, vol. 7, 2003, pp. 671–680.

[9] A.Alahi, K.Goel, V.Ramanathan, A.Robicquet, L.Fei and S.Savarese, "Social lstm: Human trajectory prediction in crowded spaces, " in proc: IEEE conference on computer vision and pattern recognition, 2016,pp. 961–971.

[10] S.Haddad and S.Lam, " Self-growing spatial graph networks for pedestrian trajectory prediction" in Proc. The IEEE Winter Conference on Applications of Computer Vision, 2020, pp. 1151–1159.

[11] A. Graves, S. Fern´andez, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks,"in ICML, Garching, Munich:Germany ,2006.

[12] L.Lamas, J.Barrera, G.Otranto, and C.Ugrinowitsch, *invasion team sports: strategy and match modeling* , "*International Journal of Performance Analysis in Sport*, vol.14, no.1,pp.307–329, 2014.

[13] J.Chen, H.M Le, P.Carr, Y.Yue, and J.J.Little, " Learning online smooth predictors for realtime camera planning using recurrent decision trees ", in Proc:IEEE Conference on Computer Vision and Pattern Recognition, 2016,pp. 4688–4696.

[14] K.Kim, M.Grundmann, A.Shamir, I.Matthews, J.Hodgins,and I.Essa, " Motion fields to predict play evolution in dynamic sport scenes, " in proc. IEEE International Conference on Computer Society Conference on Computer Vision and Pattern Recognition, 2010,pp.840–847.

[15] K M Kitani, B.D Ziebart, J.A.Bagnell, and M.Hebert, "Activity forecasting, " In European Conference on Computer Vision," journal of the Springer, 2012, pp. 201–214.

[16] P .Felsen, P.Lucey, and S.Ganguly. "Where will they go? predicting fine-grained adversarial multiagent motion using conditional variational auto encoders, " in proc:The European Conference on Computer Vision ,2018.

[17] P .Vraćar, E .Strumbelj, and I .Kononenko ," Modeling basketball play-by-play data. Expert Systems with Applications, " vol. 44 , 2016 ,pp.58-66.

[18] T Lan, Y Wang, W Yang, and G Mori, " Beyond actions: Discriminative models for contextual group activities, " in proc: Conference Advances in neural information processing systems, pp.1216–1224, 2010.

[19] D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, vol. 401, no. 6755, pp. 788–791, 1999.

[20].J. P. Brunet, P.Tamayo, T.R .Golub and J.P .Mesirov , Metagenes and molecular pattern discovery using matrix factorization, proceedings of the National Academy of Sciences of the USA, vol.101,no. 12,PP.4164-9, 2004.

[21] J. Yang and J. Leskovec, "Overlapping community detection at scale: A nonnegative matrix factorization approach," in proc: ACM Conference on Web Search and Data Mining (WSDM), 2013.

[22] S. Scott, *Detecting network intrusion using a Markov modulated nonhomogeneous Poisson process, Submitted to the Journal of the American Statistical Association* , vol. 31 ,pp 80 , 2000.

[23]. A .Khan, B.Lazzerini. G.Calabrese, L.Serafini. " Soccer event detection, " in proc: 4thInternational Conference on Image Processing and Pattern Recognition , Copenhagen, Denmark ,2018; pp. 119–129.

[24]. M.A .Russo, L .Kurnianggoro, K .Jo, " Classification of sports videos with combination of deep learning models and transfer learning, " in proc. International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox's Bazar, Bangladesh,2019;pp. 1–5.

[25] J .Gudmundsson and M.Horton, *Spatio-temporal analysis of team sports*, journal of ACM Computing Surveys (CSUR),vol50, no 22, 2017.pp. 1–34.

[26]. A.Ihler , J.Hutchins, and P.Smyth, " Adaptive event detection with time–varying Poisson processes " in proc :twelfth international conference on Knowledge discovery and data mining. ACM Press, New York, NY, USA, 2006,pp.207–216.

[27] H.Le, P.Carr, Y.Yue, and P Lucey, " Data-driven ghosting using deep imitation learning". 2017.
[28]. K .Richly, F .Moritz. C .Schwarz, Utilizing artificial neural networks to detect compound events in spatio-temporal soccer data, SIGKDD Workshop MiLeTS, Halifax, NS, Canada, August 2017.
[29] J. Palmgren,"The Fisher information matrix for log linear models arguing conditionally on observed explanatory variables, journal of the *Biometrical* ,vol. 63, no. 19, pp. 1-25,1981.
[30] S. G. Baker, The multinomial-Poisson transformation, *journal of the Royal Statistical Society. Series D*, vol. 43, no. 4, pp. 495-504, 1994.
[31] M. L. Puterman , Markov Decision Processes: Discrete Stochastic Dynamic Programming , 2014. [E-book]
[32]. W. Li and M. K .Ng, " On the limiting probability distribution of a transition probability tensor. Linear and Multilinear Algebra," ,2014 , pp. 362-385.[e-book]
[33].R. P. Adams, George E. Dahl, and I.Murray, " Incorporating side information into probabilistic matrix factorization using Gaussian processes, " in proc: the 26th Conference on Uncertainty in Artificial Intelligence ,. 2010.
[34].K.Goldsberry, E.Weiss,"The Dwight effect:A new ensemble of interior defense analytics for the NBA ,"In proc: Sloan Sports Analytics Conf, 2012.
[35] B. Yu, J. Cunningham, G. Santhanam, S. Ryu, K. Shenoy, and M. Sahani," Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity." in Neural Information Processing Systems (NIPS),vol.102,no.1 2009, pp.614-35.
[36] I. Murray, R.P. Adams, and D.J.C. MacKay, Elliptical slice sampling, Journal of Machine Learning ,pp.541-548, 2010.
[37] D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in proc: the 13th International Conference on Neural Information Processing Systems(NIPS), pp.535–541, 2001.
[38] L. Xu, J. A. Duan, and A. Whinston, "Path to purchase: A mutually exciting point process model for online advertising and conversion," Management Science, vol. 60, no. 6, pp. 1392–1412, 2014.
[39] K.Yasuda, K .Tsuboi, k.Tanaka,T. Miyazaki,Estimation of aerodynamic coefficients for a ball by using characteristics of trajectory, Transactions of the Japan Society of Mechanical Engineers, vol.80, no.814, 2014, PP.1-10.
[40] A.Gupta, J.Johnson, L.Fei-Fei, S.Savarese, and A. Alahi, " Social gan: Socially acceptable trajectories with generative adversarial networks," in proc: IEEE Conference on Computer Vision and Pattern Recognition ,2018,pp.2255–2264.
[41] C.Rupprecht, I.Laina, R.DiPietro, and M.Baust. "Learning in an uncertain world: Representing ambiguity through multiple hypotheses," in proc: IEEE International Conference on Computer Vision , 2017.
[42] J. von Neumann, Various techniques used in connection with random digits, in *Monte Carlo Method*, Washington D.C., National Bureau of Standards Applied Mathematics Series, 1951, pp. 36-38.
[43] P. Diggle, A kernel method for smoothing point process data, *journal of the Royal Statistical Society: Series C,* vol. 34, no. 2, pp. 138-147, 1985
[44] D. M. Higdon, "Space and space-time modeling using process convolutions, " in proc: international of conference *Quantitative methods for current environmental issues*, pp. 37-54, 2002.
[45] F.Lindgren ,H. Rue, Bayesian spatial modeling with R-IN, *Journal of Statistical Software*,vol.63, no.19, pp.1-25, 2015.
[46] A. Datta, S. Banerjee, A. O. Finley and A. E. Gelfand, Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets, *Journal of the American Statistical Association*, vol. 111, no. 514, pp. 800-812, 2016.
[47] S. Banerjee, A. E. Gelfand, A. O. Finley and H. Sang, "Gaussian predictive process models for large spatial data sets, *Journal of the Royal Statistical Society*, vol.70, no. 4, pp. 825-848, 2008.
[48]. T.Misu, m.Naemura , w. Zheng , y.Izumi and k.Fukui K, " Robust tracking of soccer players based on data fusion. " in proc:16th international conference on pattern recognition, vol.1, 2002,pp 556–561.
[49]. D.Michel, C.Panagiotakis, A.A. Argyros, "Tracking the articulated motion of the human body with two RGBD cameras, ", 2015, pp.41–54.
[50] R.C .Eberhart ,Y.Shi, Y, " Comparing inertia weights and constriction factors in particle swarm optimization , *Proceedings of the 2000 Congress on Evolutionary Computation*, La Jolla, CA, USA, vol.1,2000, pp. 84–88.
[51] A.M .Tawab, M.Abdelhalim, S.D. Habib, Efficient multi-feature PSO for fast gray level object-tracking .Appl. Soft Comput. 2014 , pp.317–337.
[52] B .Kwolek, " Object tracking via multi-region covariance and particle swarm optimization, " in proc: IEEE International Conference on Advanced Video and Signal Based Surveillance ,, 2009, pp. 418–423.
[53].C. Guestrin, D. Koller, R. Parr, and S. Venkataraman, Efficient solution algorithms for factored MDPs, *Journal of Artificial Intelligence Research*, vol.19,pp.399–468, 2003
[54].T. Degris and O. Sigaud, Factored Markova Decision Processes, Markov Decision Processes in Artificial Intelligence, pp.99–126, 2010. [E-book]
[55].M. Araya-Lȯpez, O. Buffet, V. Thomas, and F. Charpillet , Active learning of MDP models, Workshop on Reinforcement Learning, pp. 42–53, 2011.
[56] N. Rong, Learning in the Presence of Unawareness, PhD Thesis, Cornell University, 2016.
[57] C. Innes and A. Lascarides, " Learning Structured Decision Problems with Unawareness, " in proc: international Conference on Machine Learning, 2019,pp. 2941–2950.
[58] H. Cui, V. Radosavljevic, F .Chou, T .Lin, T .Nguyen, T .Huang, J .Schneider, and N.Djuric,"Multi-modal trajectory predictions for autonomous driving using deep convolutional networks. " in proc: IEEE International Conference on Robotics and Automation (ICRA), 2019.
[59] J. Duchi, E. Hazan, and Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, Journal of Machine Learning Research (JMLR), vol. 12, pp. 2121–2159, 2011.
[60] T. A. Lasko, "Efficient inference of Gaussian process-modulated renewal processes with application to medical event data," in proc: Thirtieth Conference on Uncertainty in Artificial Intelligence, 2014 ,pp.469–476.

[61] R .Yurko, F.Matano, L. F.Richardson, N.Granered,T.Pospisil, K.Pelechrinis and S.L. Ventura, Going deep: models for continuous-time within-play valuation of game outcomes in American football with tracking data, *Journal of Quantitative Analysis in Sports* , pp.1559-0410,2019.
[62] N.G.Polson, J. G .Scott, On the half-Cauchy prior for a global scale parameter, journal of Bayesian Analysis, " vol.7,no.4. 2012 ,pp. 887-902.
[63].P. Diggle, Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, CRC Press, 2013, [E-book].
[64] Y. Ogata, "Space-time point-process models for earthquake occurrences," Annals of the Institute of Statistical Mathematics, vol.50, pp. 379–402, 1998.
[65] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in proc. International Conference on Machine Learning (ICML), 2001, pp.282-289.
.[66] C. Rasmussen and C. Williams, *Gaussian processes for machine learning MIT press Cambridge*, MA, 2006,
[67] M.Beetz, N. von Hoyningen-Huene, B. Kirchlechner, S. Gedikli, F. Siles,M. Durus, L.M, Aspogamo: Automated sports game analysis models,International Journal of Computer Science in Sport, vol. 8, no. 1, 2009.
[68] L .Morra, F.Manigrasso, G.Canto, C.Gianfrate,.E.Guarino, F.Lamberti, " Slicing and dicing soccer: Automatic detection of complex events from spatio-temporal data, " in proc: Conference on Image Analysis and Recognition , 2020, pp 107–121.
[69] M.L.Puterman,Markov Decision Processes:Discrete Stochastic Dynamic Programming, John Wiley&Sons,2014.[E-book]
[70] S. Hauri, N. Djuric, V. Radosavljevic and S. Vucetic, "Multi-Modal Trajectory Prediction of NBA Players," *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2021, pp. 1639-1648.

.