Automatic Event Detection in Basketball

Suraj Keshri, Min-hwan Oh, Sheng Zhang, Garud Iyengar

September 22, 2017

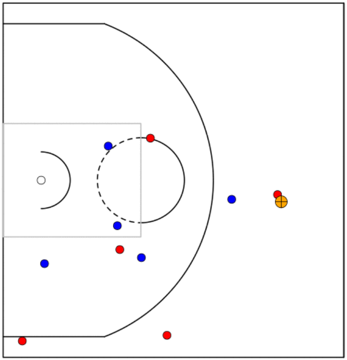
IEOR Department, Columbia University

Introduction

Motivation

- Recognizing player match-ups and game events, e.g. ball screen, drive, post-up, etc., crucial for gaining insights both on players and teams
- Manually labeling these events not scalable
- Goal: Detect events automatically using player tracking data!

Player Tracking Data



Installed in 2013. Tracks:

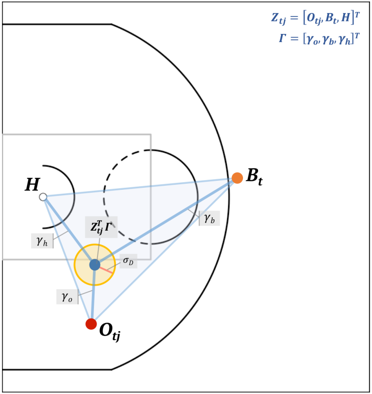
- (x, y) locations of all 10 players
- (x, y, z) locations of ball
- 25 observations per second
- Event annotations (shots, passes, fouls, etc.)

1230 games per season: $\approx$ 1 billion space-time points per season

Defense Assignment

Defense Attraction in Basketball

Basic Setting



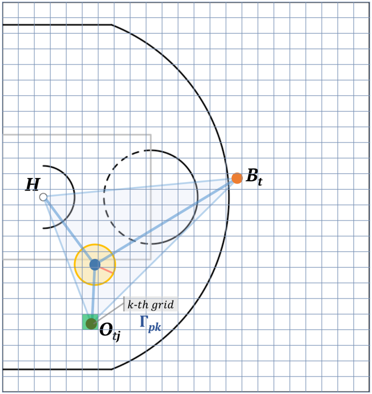
- D_{ti}: location of defender i at time t
- O_{tj}: location of offender j at time t
- $I_{tij} = 1$: i guards j at time t
- Stochastic model

$$D_{ti}|I_{tij} = 1 \sim \mathcal{N}\left(Z_{tj}^T\Gamma, \sigma_D^2\right)$$

[Franks et al. (2015)]

- Defender location is determined by offender characteristics

Basic Setting - player and location dependency



- Γ is player and location dependent:

$$\Gamma_{pk} = \left[\gamma^{o}_{pk}, \gamma^{b}_{pk}, \gamma^{h}_{pk}\right]$$
$$p = g(\cdot,\cdot) \text{ grid picker}$$

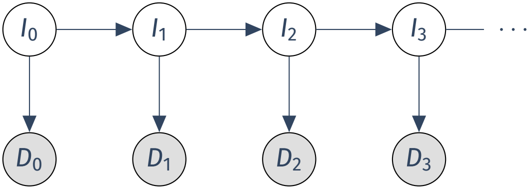
$$D_{ti}|I_{tij} = 1 \sim \mathcal{N}\left(Z_{tj}^{T}\Gamma_{g(t,j)}, \sigma_D^2\right)$$

- Prior on $\Gamma_p = [\Gamma_{p1}, \ldots, \Gamma_{pK}]$

$$\Gamma_p \sim \mathcal{N}\left(\mu_\Gamma, \mathcal{K}\right)$$

Hidden Markov Model

- Model the evolution of man-to-man defense using HMM
- Hidden states (I_t): defensive mapping



- How about transition probability?

Transition Probability

- Total of $5^5 (= 3125)$ matchings $\Rightarrow$ intractable to learn probabilities for all transitions
- Propose a bond energy based defensive assignment transition
  - Single defensive match-up: bond
  - Defensive switching: breaking and forming a new bond.
- 4 types of bonds: 1-on-1 on-ball (or off-ball) bond, extra on-ball (or off-ball) bond
- Transition probability proportional to energy difference

$$P(I_t \rightarrow I_{t+1}) \propto e^{-\Delta E_{t,t+1}}$$

Transition Probability: Example

- Double team match-ups have higher energy (more unstable) than 1-1 match-up. Hence, more likely to switch to 1-1 match-up

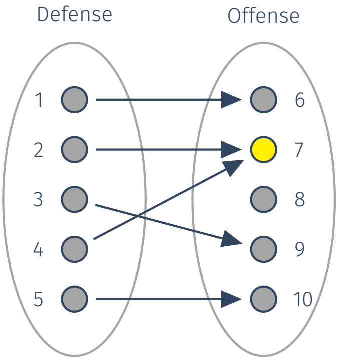


Figure 1: On-ball double team

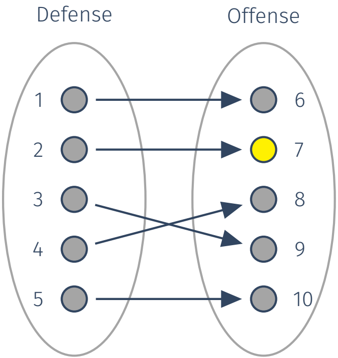
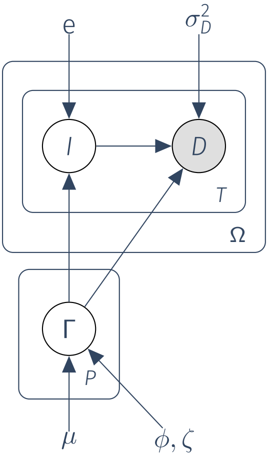


Figure 2: One-to-one match-up

Graphical Model: Defense Assignment



- For player p and location k: sample $\Gamma_{pk} \sim \mathcal{N}(\mu_\Gamma, \mathcal{K})$
- For all times t, sample defensive assignment I_t using energy based transition
- Sample each defender's location at t

$$D_{ti} | I_{tij} = 1 \sim \mathcal{N}\left(Z_{tj}^\top \Gamma_{g(t,j)}, \sigma_D^2\right)$$

- Iterate until convergence

Inference: Defense Assignment

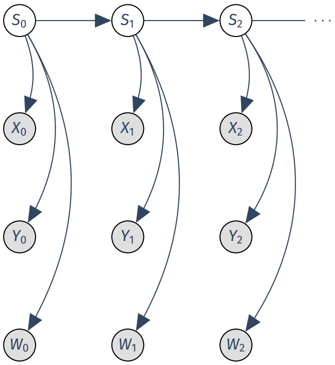
- Initialize all the fixed parameters for GP prior, bond energies $\mathbf{e}$, and σ_D^2. Let θ denote all the fixed parameters
- Until convergence
  - Sample from $P(I|\Gamma, D, \theta)$ using forward filtering backward sampling algorithm
  - Update energy parameters $\mathbf{e}$ given the sample of I
  - Sample $P(\Gamma|I, D, \theta)$
  - Update kernel parameters, and σ_D^2 given the sample of Γ

Event Detection

Event Detection

- Want to detect events without labeled data
- Model sequence of event progression using HMM
- Define the binary hidden state at each time point as an indicator of whether an event is taking place or not
- Specify the parametric form of the emission distributions which are characteristic to actions
- Using HMM, compute most likely sequence of hidden state

Ball Screen

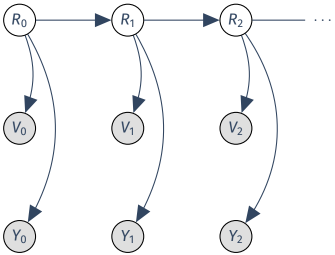


- S_t: indicator of ball screen event
- X_t: distance between on-ball defender and potential screener
- Y_t: distance between hoop and ball handler
- W_t: speed of potential screener

$$
\begin{aligned}
X_t | S_t = 1 &\sim \exp(\lambda_x) \\
Y_t | S_t = 1 &\sim \log N(\mu_y, \sigma_y^2) \\
W_t | S_t = 1 &\sim \exp(\lambda_w)
\end{aligned}
$$

$$
\begin{aligned}
X_t | S_t = 0 &\sim \log \mathcal{N}(\mu_x, \sigma_X^2) \\
Y_t | S_t = 0 &\sim \text{Unif}(0, \theta_y) \\
W_t | S_t = 0 &\sim \log \mathcal{N}(\mu_w, \sigma_w^2)
\end{aligned}
$$

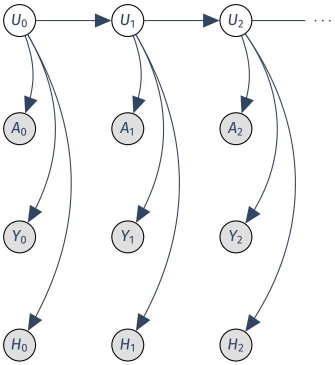
Drive



- R_t: indicator of drive event
- V_t: velocity of ball handler towards hoop
- Y_t: distance between hoop and ball handler

$$\begin{aligned}
\frac{1}{V_t^+} | R_t = 1 &\sim \exp(\lambda_v) \\
Y_t | R_t = 1 &\sim \exp(\lambda_y) \\
\\
V_t | R_t = 0 &\sim \mathcal{N}(\mu_v, \sigma_v^2) \\
Y_t | R_t = 0 &\sim \mathrm{Unif}(0, \theta_y)
\end{aligned}$$

Post-up



- U_t: indicator of post-up event
- A_t: distance between on-ball defender and ball handler
- Y_t: distance between hoop and ball handler
- H_t: speed of ball handler

$$
\begin{aligned}
A_t|U_t = 1 &\sim \exp(\lambda_a) \\
Y_t|U_t = 1 &\sim \log\mathcal{N}(\mu_y, \sigma_y^2) \\
H_t|U_t = 1 &\sim \exp(\lambda_h)
\end{aligned}
$$

$$
\begin{aligned}
A_t|U_t = 0 &\sim \log\mathcal{N}(\mu_a, \sigma_a^2) \\
Y_t|U_t = 0 &\sim \text{Unif}(0, \theta) \\
H_t|U_t = 0 &\sim \log\mathcal{N}(\mu_h, \sigma_h^2)
\end{aligned}
$$

Inference

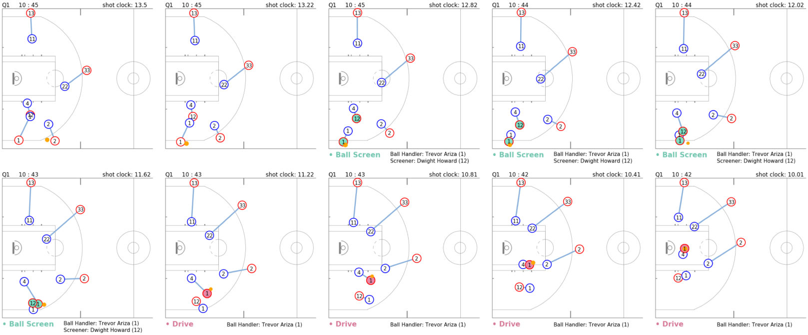
h = hidden state of event indicator (S_t, R_t, U_t)

$\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \ldots$ = sequences of observed states

- Initialize $\hat{P}(h_0), \hat{P}(x|h), \hat{P}(y|h), \ldots$, and $\hat{P}(h'|h)$ randomly
- Until convergence
  - **E Step**: For each sequence $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \ldots$, compute $\hat{P}(h_0|\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \ldots), \hat{P}(h_t, h_{t+1}|\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \ldots), \hat{P}(h'|h)$ using forward-backward algorithm
  - **M Step**: Update the model parameters $\hat{P}(h_0), \hat{P}(x|h), \hat{P}(y|h), \ldots$, and $\hat{P}(h'|h)$ using MLE
- Compute most likely sequence of hidden states, $\boldsymbol{h} = (h_0, \ldots, h_T)$ using Viterbi algorithm

Results

Estimated Defense Assignments and Events



- Lines represent estimated defense assignments
- Ball screen and drive actions are captured in the sequence

Accuracy

Table 1: Defense Assignment Accuracy Comparison

Model	Accuracy
Closest Defender	0.7597
Fixed Γ Model (Franks et al.)	0.9179
Player Attraction based Model	**0.9541**

Table 2: Event Detection Accuracy

Event	Accuracy
Ball Screen	0.868
Drive	0.953
Post-up	0.994

Table 3: Ball Screen Detection

	Prediction	
Actual	Positive	Negative
Positive	79	5
Negative	23	106

Table 4: Drive Detection

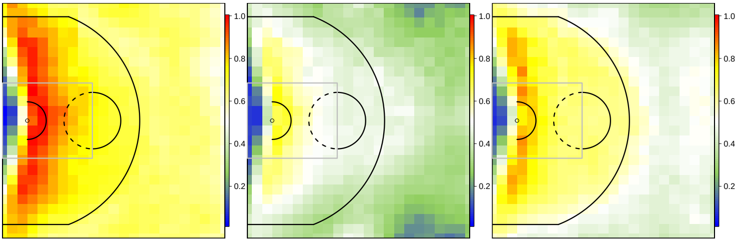
	Prediction	
Actual	Positive	Negative
Positive	65	2
Negative	4	58

Table 5: Post-up Detection

	Prediction	
Actual	Positive	Negative
Positive	8	0
Negative	2	334

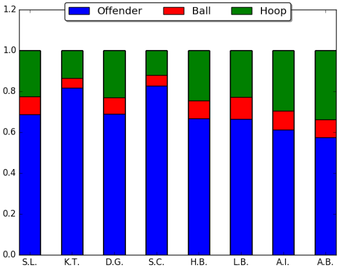
Γ Heatmap for Selected Players

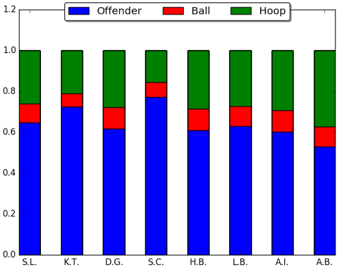
- Stephen Curry DeAndre Jordan LeBron James



Γ Results for Golden State Warriors

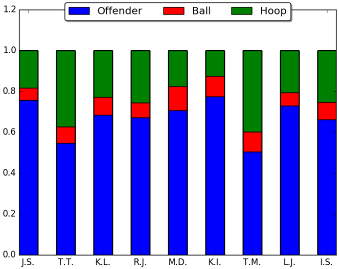
- High γ_o: (S.C.) Stephen Curry, (K.T.) Klay Thompson
- Low γ_o: (A.B.) Andrew Bogut, (A.I.) Andre Iguodala

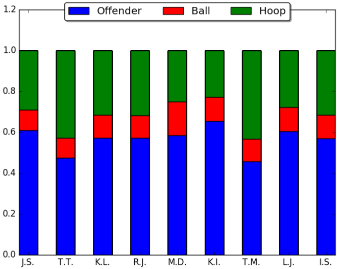




Γ Results for Cleveland Cavaliers

- High γ_o: (K.I.) Kyrie Irving, (J.S.) J. R. Smith, (L.J.) LeBron James
- Low γ_o: (T.M.) Timofey Mozgov, (T.T.) Tristan Thompson





Questions?