Player movement, Play evaluation, Defense

Jackson Wang, March 2017

Papers

- Generating Long-term Trajectories Using Deep Hierarchical Networks
- Defensive Metric
  - Counterpoints: Advanced Defensive Metrics for NBA Basketball
  - CHARACTERIZING THE SPATIAL STRUCTURE OF DEFENSIVE SKILL IN PROFESSIONAL BASKETBALL
- Expected Point Value
  - POINTWISE: Predicting Points and Valuing Decisions in Real Time with NBA Optical Tracking Data
  - A Multiresolution Stochastic Process Model for Predicting Basketball Possession Outcomes

Hierarchical Policy Network

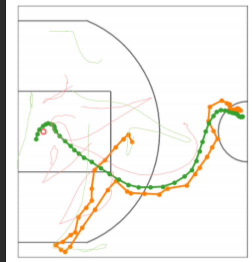


Figure 1: The player (green) has two macro-goals: 1) pass the ball (orange) and 2) move to the basket.

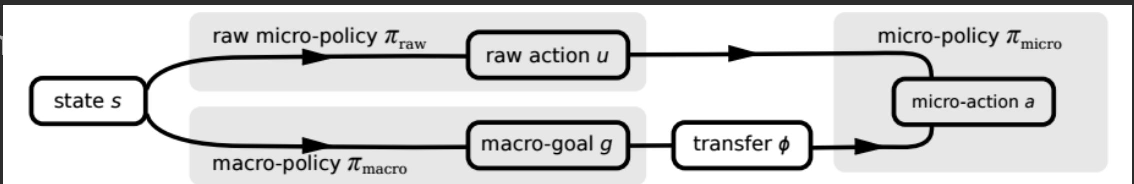
Goal

Given movement history, predict future movement

- At time t, an agent i is in state $s_t^i \in \mathcal{S}$ and takes action $a_t^i \in \mathcal{A}$. The full state and action are $s_t = \{s_t^i\}_{\text{players } i}$, $a_t = \{a_t^i\}_{\text{players } i}$. The history of events is $h_t = \{(s_u, a_u)\}_{0 \leq u < t}$.
- Macro policies also use a goal space $\mathcal{G}$, e.g. regions in the court that a player should reach.

Key insight

Player m



HPN - method

Given input state s, we want to decide on an action a

$$u = RNN_{micro}(s) \tag{1.4}$$

$$P(g|s) = g = RNN_{macro}(s) \tag{1.5}$$

$$P(a|s) = u \odot softmax(NN(g)) \tag{1.6}$$

$$L = L_{macro} + L_{micro} + R(\theta) \tag{1.7}$$

$$L_{micro} = -\log P(a = a_{truth}) \tag{1.8}$$

$$L_{macro} = -\log P(g = g_{weak}) \tag{1.9}$$

where $\odot$ is element-wise product and g_{weak} is pre-computed macro-level labels. They called the transformation $softmax(NN(\cdot))$ the attention mechanism ϕ_{ATT}. Training is done in 2 stages

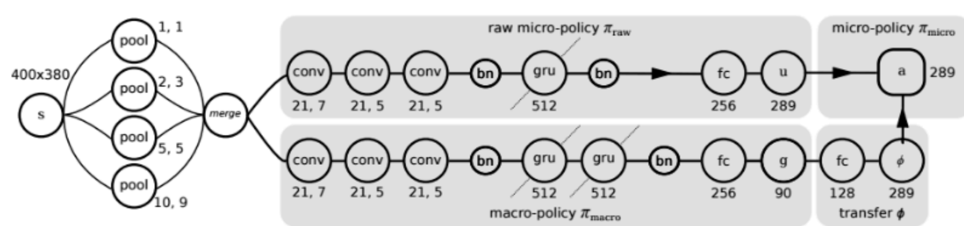
Algorithm 1 HPN Training

1: Pretrain RNN_{micro}, RNN_{macro}, ϕ_{ATT} using a_{truth}, and pre-computed g_{weak}, ϕ_{weak} independently
2: Train the whole HPN without L_{macro}

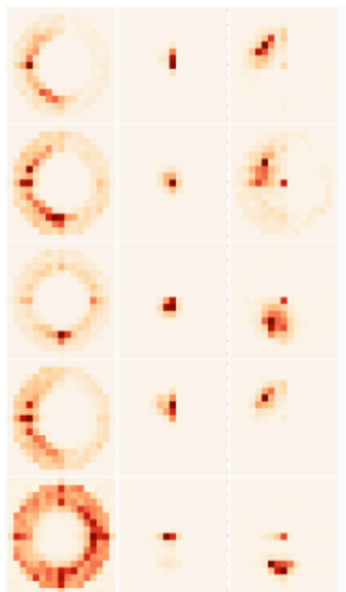
HPN-details

Table 1.1: HPN preproc ($N_{train} = 13k, \quad N_{heldout} = 1.3k$)

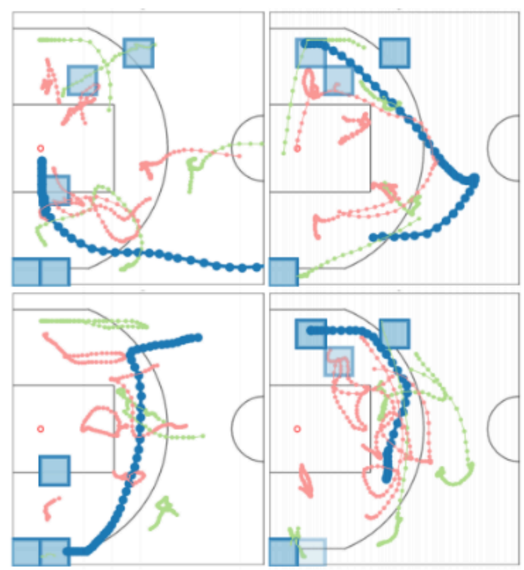
input state (image)
1. extract 200 frams from random starting point
2. downsample by 4
3. turn into $400 \times 380 \times 4$ images of player,ball,team,defense
micro label a (1-hot image)
1. 17×17 1-hot velocity ($1ft$ radius)
2. clipped if out of range (¡1% of the time)
weak label g (1-hot image)
1. identify location where player moved $< 1\frac{ft}{r}$ for 5 frames
2. 10×9 image of 1-hot occupancy
weak label ϕ (1-hot image)
1. 17×17 image mask (all zeros)
2. randomly one pixel to 1 in the direction of $\overrightarrow{g_t - s_t}$ with magnitude $\in [1, 7]$



HPN-visualization



(a) Predicted distributions for attention masks $\phi(g)$ (left column), velocity (micro-action) π_{micro} (middle) and weighted velocity $\phi(g) \odot \pi_{\text{micro}}$ (right) for basketball players. The center corresponds to 0 velocity.



(b) Rollout tracks and predicted macro-goals g_t (blue boxes). The HPN starts the rollout after 20 frames. Macro-goal box intensity corresponds to relative prediction frequency during the trajectory.

HPN-results (I)

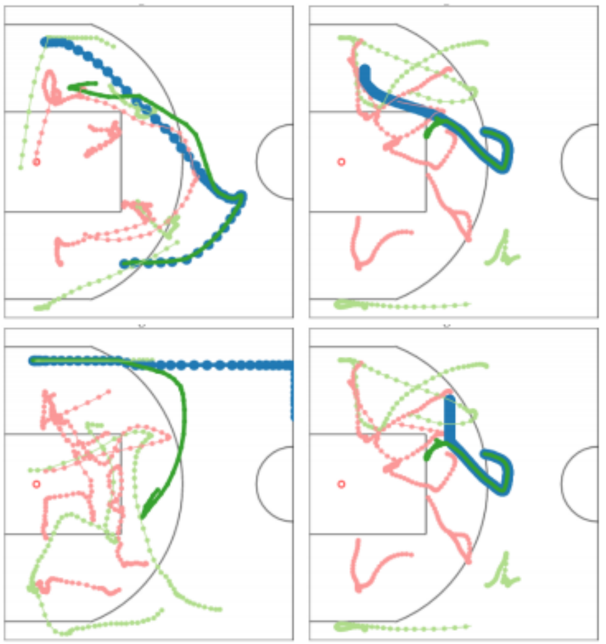
Model	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$	$\Delta = 3$	Macro-goals g	Attention ϕ
CNN	21.8%	21.5%	21.7%	21.5%	-	-
GRU-CNN	25.8%	25.0%	24.9%	24.4%	-	-
H-GRU-CNN-CC	31.5%	29.9%	29.5%	29.1%	10.1%	-
H-GRU-CNN-STACK	26.9%	25.7%	25.9%	24.9%	9.8%	-
H-GRU-CNN-ATT	**33.7%**	**31.6%**	**31.0%**	**30.5%**	10.5%	-
H-GRU-CNN-AUX	31.6%	30.7%	29.4%	28.0%	10.8%	19.2%

Table 2: Benchmark Evaluations. Δ-step look-ahead prediction accuracy for micro-actions $a^i_{t+\Delta} = \pi(s_t)$ on a holdout set, with $\Delta = 0, 1, 2, 3$. H-GRU-CNN-STACK *is an HPN where predictions are organized in a feed-forward stack, with $\pi(s_t)_t$ feeding into $\pi(s_t)_{t+1}$. Using attention (*H-GRU-CNN-ATT*) improves on all baselines in micro-action prediction. All hierarchical models are pre-trained, but not fine-tuned, on macro-goals $\hat{g}$. We report prediction accuracy on the weak labels $\hat{g}, \hat{\phi}$ for hierarchical models.*H-GRU-CNN-AUX *is an HPN that was trained using $\hat{\phi}$. As $\hat{\phi}$ optimizes for optimal long-term behavior, this lowers the micro-action accuracy.*

HPN-results (II)

Model comparison	Experts W/T/L	Avg Gain	Non-Experts W/T/L	Avg Gain	All W/T/L	Avg Gain
VS-CNN	21 / 0 / 4	0.68	15 / 9 / 1	0.56	21 / 0 / 4	0.68
VS-GRU-CNN	21 / 0 / 4	0.68	18 / 2 / 5	0.52	21 / 0 / 4	0.68
VS-H-GRU-CNN-CC	22 / 0 / 3	0.76	21 / 0 / 4	0.68	21 / 0 / 4	0.68
VS-GROUND TRUTH	11 / 0 / 14	-0.12	10 / 4 / 11	-0.04	11 / 0 / 14	-0.12

HPN-failure cases



(d) HPN (top) and failure case (bottom) (e) HPN (top), baseline (bottom)

HPN - comment

The notion of weak label

No player identity

...

Defense - Intro

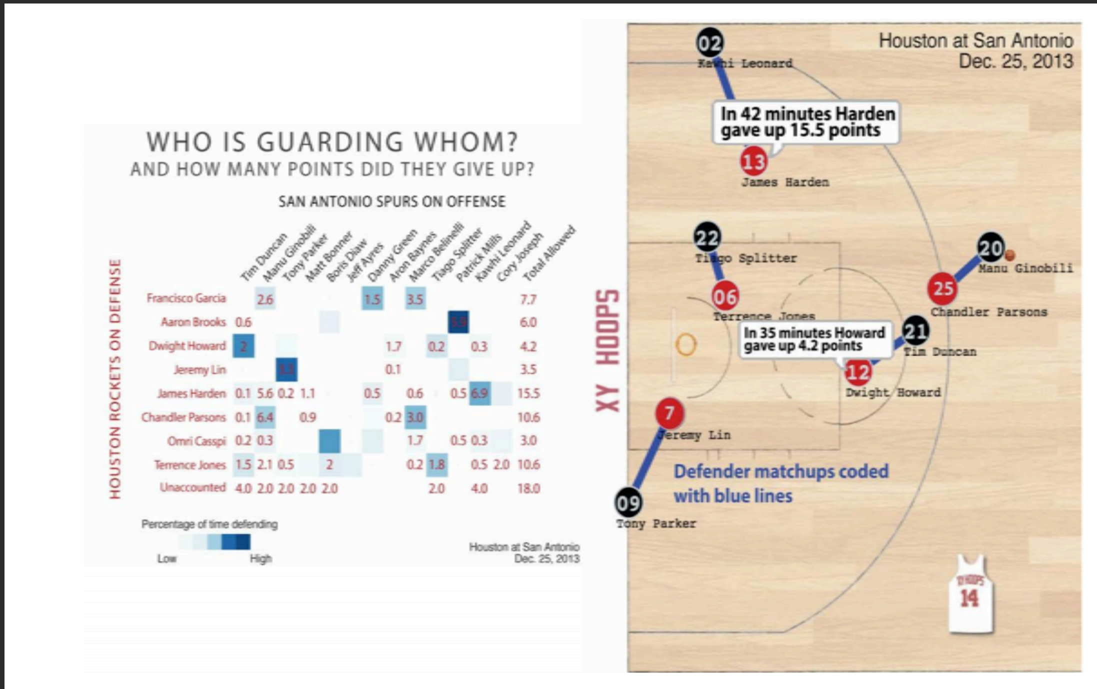


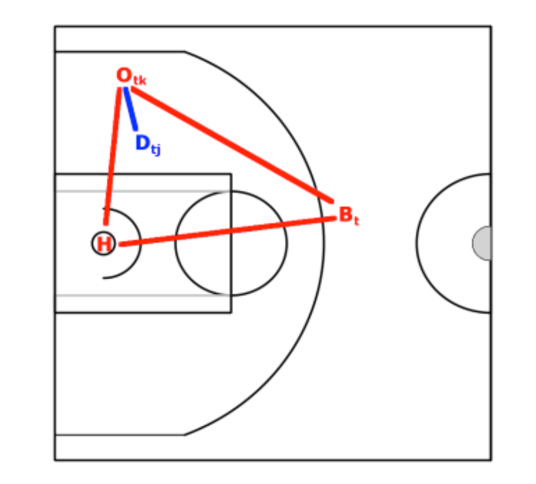
Figure 1. Matchup matrix for the Houston at San Antonio game on Dec 25, 2013. The matchup matrix has cells shaded according to the fraction of time spent guarding each offender. Counterpoints are assigned according to these fractions (see Methods). Points off of putbacks or fast breaks are not assigned to a defender ("unaccounted"). We visualize these responsibilities as a possession unfolds; the blue lines symbolize connections linking defenders to their offensive responsibilities (right side).

Defense - assignment

Defense (j), offense (k)

$$\mu_{tk} = \gamma_o O_{tk} + \gamma_b B_t + \gamma_h H,$$
$$\Gamma \mathbf{1} = 1$$

$$D_{tj} | I_{tjk} = 1 \sim N(\mu_{tk}, \sigma_D^2).$$



Defense - HMM

Parameters: 6 scalar weights, variance, 1 scalar “switch” probability

Hidden: “assignment”

$$L(\Gamma, \sigma_D^2) = P(\mathbf{D}, \mathbf{I} | \Gamma, \sigma_D^2) = \prod_{t,j,k} \left[P(D_{tj} | I_{tjk}, \Gamma, \sigma_D^2) P(I_{tjk} | I_{(t-1)j.}) \right]^{I_{tjk}},$$

$$P(I_{tjk} = 1 | I_{(t-1)jk} = 1) = \rho,$$

$$P(I_{tjk} = 1 | I_{(t-1)jk'} = 1) = \frac{1-\rho}{4}, \qquad k' \neq k$$

Defense - inference

Note: all defensive players are independent

2.1. *Inference.* We use the EM algorithm to estimate the relevant unknowns, I_{tjk}, σ_D^2, Γ and ρ. At each iteration, i, of the algorithm, we perform the E-step and M-step until convergence. In the E-step, we compute $E_{tjk}^{(i)} = E[I_{tjk} | D_{tj}, \hat{\Gamma}^{(i)}, \hat{\sigma}_D^{2(i)}, \hat{\rho}^{(i)}]$ and $A_{tjkk'}^{(i)} = [I_{tjk} I_{(t-1)jk'} | D_{tj}, \hat{\Gamma}^{(i)}, \hat{\sigma}_D^{2(i)}, \hat{\rho}^{(i)}]$ for all t, j, k and k'. These expectations can be computed using the forward–backward algorithm

Defense - inference E-step

z_{nk}. Using Bayes' theorem, we have

$$\gamma(\mathbf{z}_n) = p(\mathbf{z}_n|\mathbf{X}) = \frac{p(\mathbf{X}|\mathbf{z}_n)p(\mathbf{z}_n)}{p(\mathbf{X})}. \tag{13.32}$$

Note that the denominator $p(\mathbf{X})$ is implicitly conditioned on the parameters $\boldsymbol{\theta}^{\text{old}}$ of the HMM and hence represents the likelihood function. Using the conditional independence property (13.24), together with the product rule of probability, we obtain

$$\gamma(\mathbf{z}_n) = \frac{p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_n)p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{z}_n)}{p(\mathbf{X})} = \frac{\alpha(\mathbf{z}_n)\beta(\mathbf{z}_n)}{p(\mathbf{X})} \tag{13.33}$$

where we have defined

$$\alpha(\mathbf{z}_n) \equiv p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_n) \tag{13.34}$$

$$\beta(\mathbf{z}_n) \equiv p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{z}_n). \tag{13.35}$$

$$\alpha(\mathbf{z}_n) = p(\mathbf{x}_n|\mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} \alpha(\mathbf{z}_{n-1})p(\mathbf{z}_n|\mathbf{z}_{n-1}).$$

$$\beta(\mathbf{z}_n) = \sum_{\mathbf{z}_{n+1}} \beta(\mathbf{z}_{n+1})p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n).$$

Bishop, Christopher M. "Pattern recognition." Machine Learning 128 (2006): 1-58.

Defense - inference M step (emission)

Solved analytically using constrained generalized least square

In the ith iteration of the M-step we first update our estimates of Γ and σ_D^2,

$$(\hat{\Gamma}^{(i)}, \hat{\sigma}_D^{2(i)}) \leftarrow \arg\max_{\Gamma, \sigma_D^2} \sum_{t,j,k} \frac{E_{tjk}^{(i-1)}}{\sigma_D^2}(D_{tj} - \Gamma X_{tk})^2, \qquad \Gamma \mathbf{1} = 1.$$

Defense - inference M step (transition)

Next, we update our estimate of the transition parameter, ρ, in iteration i:

$$\hat{\rho}^{(i)} \leftarrow \arg\max_{\rho} \sum_{t,j,k} \sum_{k' \neq k} A_{tjkk'} \log\left(\frac{1-\rho}{4}\right) + \sum_{t,j,k} A_{tjkk} \log(\rho).$$

It is easy to show, under the proposed transition model, that the maximum likelihood estimate for the odds of staying in the same state, $Q = \frac{\rho}{1-\rho}$, is

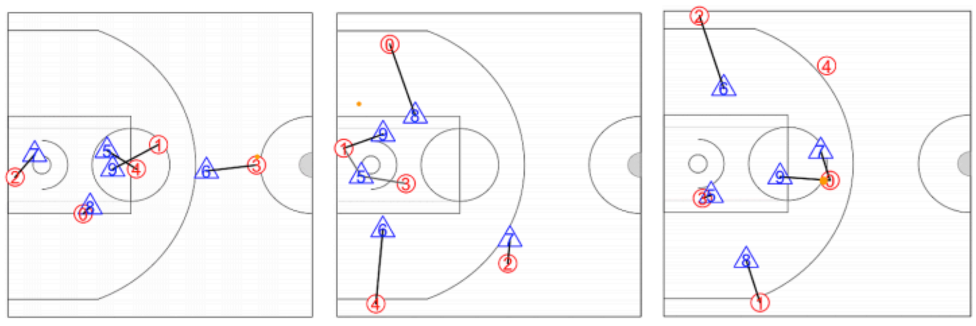
$$\hat{Q} = \frac{1}{4} \frac{\sum_{t,j,k} A_{tjkk}}{\sum_{t,j,k} \sum_{k' \neq k} A_{tjkk'}}$$

$$\hat{\rho} = \frac{\hat{Q}}{1+\hat{Q}}.$$

Defense - results

an offender. We use the EM algorithm to fit the HMM on 30 random possessions from the database. We find that a defender's canonical position can be described as $0.62 O_{tk} + 0.11 B_t + 0.27 H$ at any moment in time. That is, we infer that on aver-

Values of the transition parameter are more variable but have a smaller impact on inferred defensive matchups: values range from $\rho = 0.96$ to $\rho = 0.99$. Empirically,



(a) (b) (c)

FIG. 2. *Who's guarding whom. Players* 0–4 *(red circles) are the offenders and players* 5–9 *(blue triangles) are defenders. Line darkness represents degree of certainty. We illustrate a few properties*

Defense - App

TABLE 1

Average attention drawn, on and off ball. Using inference about who's guarding whom, we calculate the average attention each player receives as the total amount of time guarded by each defender divided by the total time playing (subset by time with and without the ball). At any moment in time, there are five defenders, and hence five units of "attention" to divide among the five offenders each possession. On ball, the players receiving the most attention are double teamed an average of 20% of their time possessing the ball. Off ball, the players that command the most attention consist largely of MVP caliber players

	On ball		Off ball	
Rank	**Player**	**Attention**	**Player**	**Attention**
1	DeMar DeRozan	1.213	Stephen Curry	1.064
2	Kevin Durant	1.209	Kevin Durant	1.063
3	Rudy Gay	1.201	Carmelo Anthony	1.048
4	Eric Gordon	1.187	Dwight Howard	1.044
5	Joe Johnson	1.181	Nikola Pekovic	1.036

Defense - App

TABLE 2

Team defensive entropy. A player's defensive entropy for a particular possession is defined as $\sum_{k=1}^{5} Z_n(j,k)\log(Z_n(j,k))$, where $Z_n(j,k)$ is the fraction of time the defender j spends guarding offender k during possession n. Team defensive entropy is defined as the average player entropy over all defensive possessions for that team. Induced entropy is the average player entropy over all defenders facing a particular offense

Rank	Team	Entropy	Rank	Team	Induced entropy
1	Mia	0.574	1	Mia	0.535
2	Phi	0.568	2	Dal	0.526
3	Mil	0.543	3	Was	0.526
4	Bkn	0.538	4	Chi	0.524
5	Tor	0.532	5	LAC	0.522
26	Cha	0.433	26	OKC	0.440
27	Chi	0.433	27	NY	0.440
28	Uta	0.426	28	Min	0.431
29	SA	0.398	29	Phi	0.428
30	Por	0.395	30	LAL	0.418

Defense - App

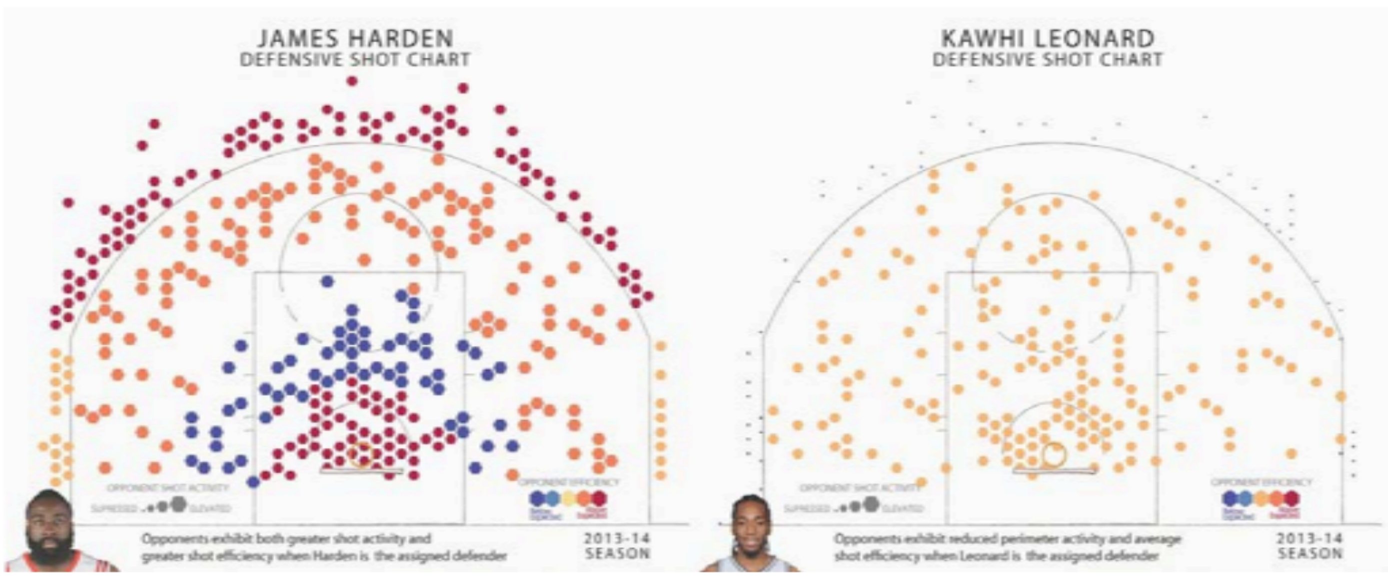


Figure 2a. Graphical depiction of a defender's volume (size) and disruption scores (color). Kawhi Leonard tends to suppress shots on the perimeter. More comparisons are provided in the Appendix.

Defense - App: counter-point

Points Against Comparison (Back Court Defenders)

Top Defenders				Bottom Defenders			
Player	Original	Shot	Fractional	Player	Original	Shot	Fractional
Chris Paul	14.4 (1)	17.7 (9)	10.8 (1)	Jrue Holiday	23.5 (61)	24.1 (50)	19.1 (63)
Norris Cole	15.0 (3)	17.0 (5)	11.1 (2)	Shaun Livingston	25.1 (63)	27.8 (62)	17.5 (62)
Nick Calathes	16.0 (5)	19.4 (18)	12.0 (3)	Jarrett Jack	21.1 (54)	22.3 (33)	17.5 (61)
C.J. Watson	18.8 (33)	19.3 (17)	12.0 (4)	Mo Williams	23.5 (62)	19.8 (19)	17.3 (60)
Greivis Vasquez	15.0 (2)	17.4 (7)	12.3 (5)	Patty Mills	23.1 (59)	23.1 (41)	17.1 (59)
Steph Curry	16.6 (7)	16.2 (2)	12.3 (6)	Kemba Walker	20.7 (51)	26.7 (60)	16.9 (58)

Table 2) Comparison of three points against metrics and their associated ranking for one defensive group (back court defenders). While highlighting slightly different aspects of defense, these metrics are largely consistent.

Defense - comment

- simple model
  - Not using player information, defense/offense strategy
  - Lack player dependent physics (e.g. speed, acceleration)
- “Average defense”, no notion of “good” defense

Halftime

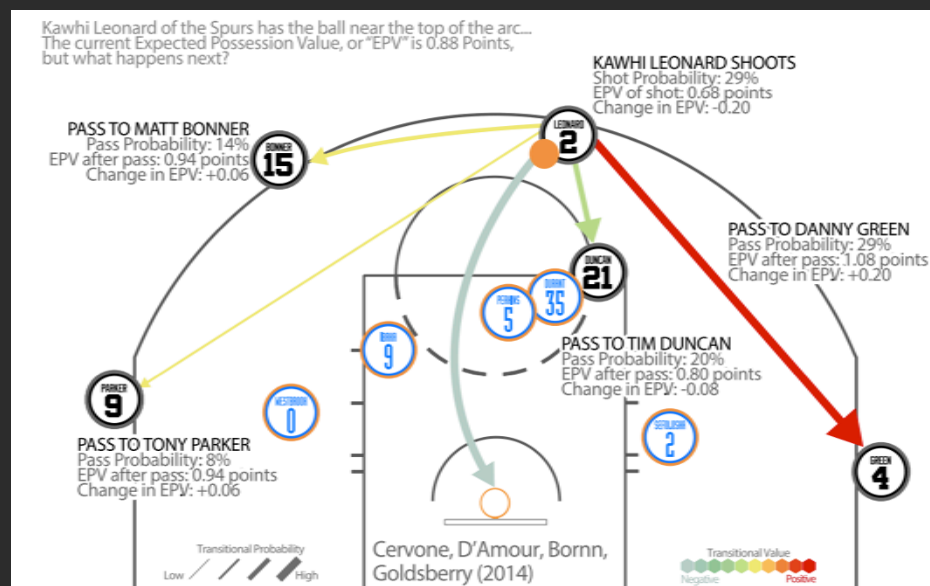
Sequence generation

Defense inference (maybe not generation)

Expected Point Value

What leads to a good shot is what’s really important (assist?)

Dynamic game



EPV app - stock ticker

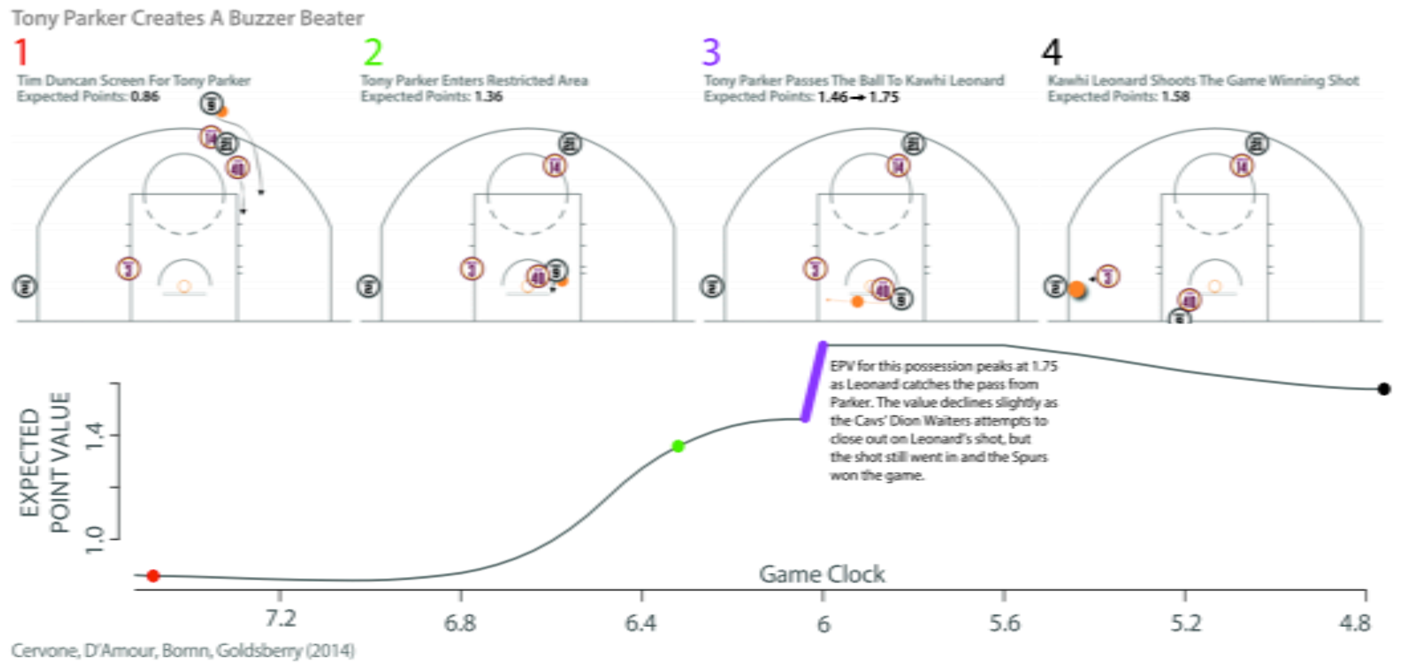
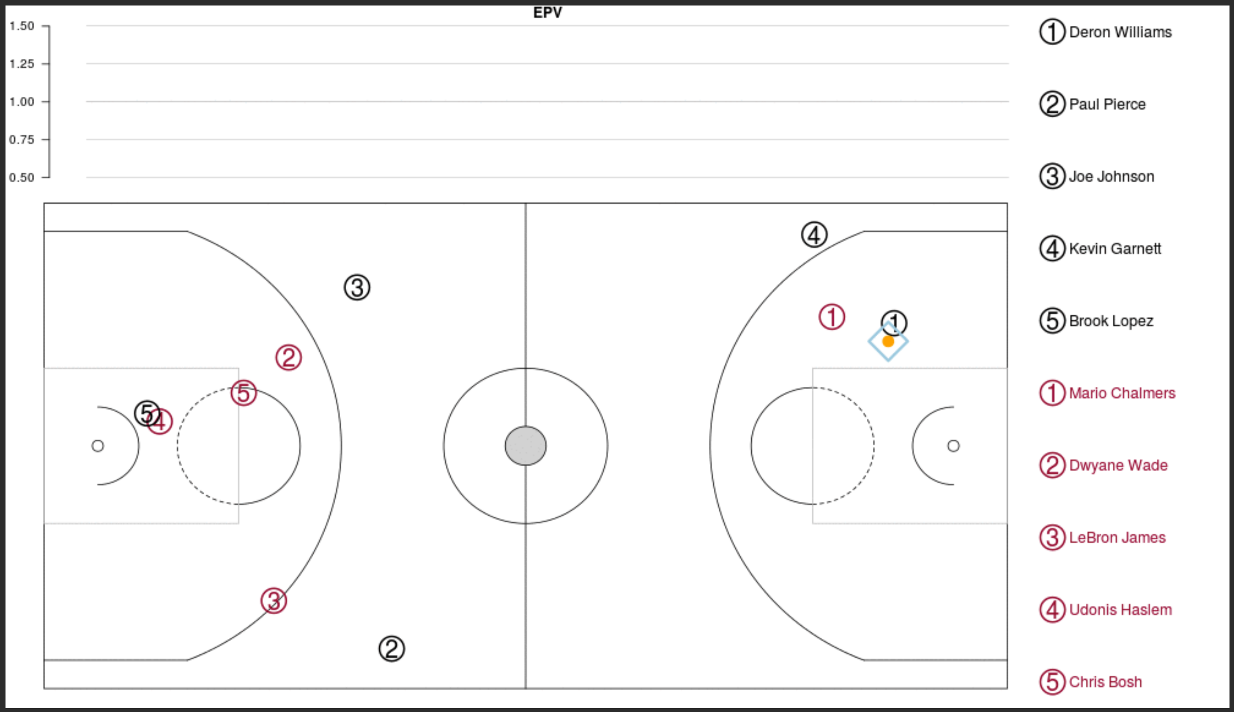


Figure 2. EPV throughout the Spurs' final possession, with annotations of major events.

EPV app - stock ticker



EPV-app: EPV-added

EPVA = EPV - EPV_r for all possessions

Player	EPVA
Chris Paul	3.48
Dirk Nowitzki	2.60
Deron Williams	2.52
Stephen Curry	2.50
Jamal Crawford	2.50
Greivis Vasquez	2.46
LaMarcus Aldridge	2.40
Steve Nash	2.09
Wesley Matthews	2.06
Damian Lillard	1.95

Player	EPVA
Ricky Rubio	-3.33
Kevin Love	-2.38
Russell Westbrook	-2.07
Evan Turner	-1.90
Austin Rivers	-1.84
Rudy Gay	-1.75
Jrue Holiday	-1.51
Paul George	-1.49
Chris Singleton	-1.48
Roy Hibbert	-1.44

EPV-app: Shot Satisfaction

$$\text{Shot Satisfaction} = \sum_{t \text{ for shot att.}} \text{EPV}(t) - E[\text{points } \mid \text{pass in } (t, t+\epsilon], \ d_t].$$

Table 2. Top 10 and bottom 10 players by average shot satisfaction in 2012-13 (per shot attempt, minimum 500 touches during season). The sampling bias concerns noted in Table 1 apply to these results as well.

Player	Shot Satisfaction
Lance Stephenson	0.362
Steve Nash	0.340
Pablo Prigioni	0.335
Chris Paul	0.334
Jamal Crawford	0.310
Jared Dudley	0.286
Martell Webster	0.283
Stephen Curry	0.258
Amir Johnson	0.256
Patrick Mills	0.255

Player	Shot Satisfaction
Alonzo Gee	-0.098
Daniel Gibson	-0.082
Ricky Rubio	-0.067
Patrick Beverley	-0.046
Michael Beasley	-0.033
Andre Miller	-0.005
Luc Richard Mbah a Moute	-0.005
George Hill	0.001
Evan Turner	0.001
Glen Davis	0.010

EPV-theoretical definition

$\omega \in \Omega$ possessions

$T(\omega)$ end time of possessions

$Z_t(\omega)$ SportVU snapshot

$X(\omega)$ outcome of possession

$F_t^{(z)} = \sigma(\{Z_s^{-1} : 0 \le s \le t\})$ history

$$EPV = \nu_t = \mathbb{E}[X|F_t^{(z)}]$$

$$\begin{aligned}\nu_t = \mathbb{E}[X|\mathcal{F}_t^{(Z)}] &= \int_\Omega X(\omega)\mathbb{P}(d\omega|\mathcal{F}_t^{(Z)}) \\ &= \int_t^\infty \int_{\mathcal{Z}} h(z)\mathbb{P}(Z_s = z|T = s, \mathcal{F}_t^{(Z)})\mathbb{P}(T = s|\mathcal{F}_t^{(Z)})dzds.\end{aligned}$$

EPV-"coarsened processes"

A coarsened description of the basketball world (i.e. someone has ball, passes, shoots, end)

Markov

Decoupling events

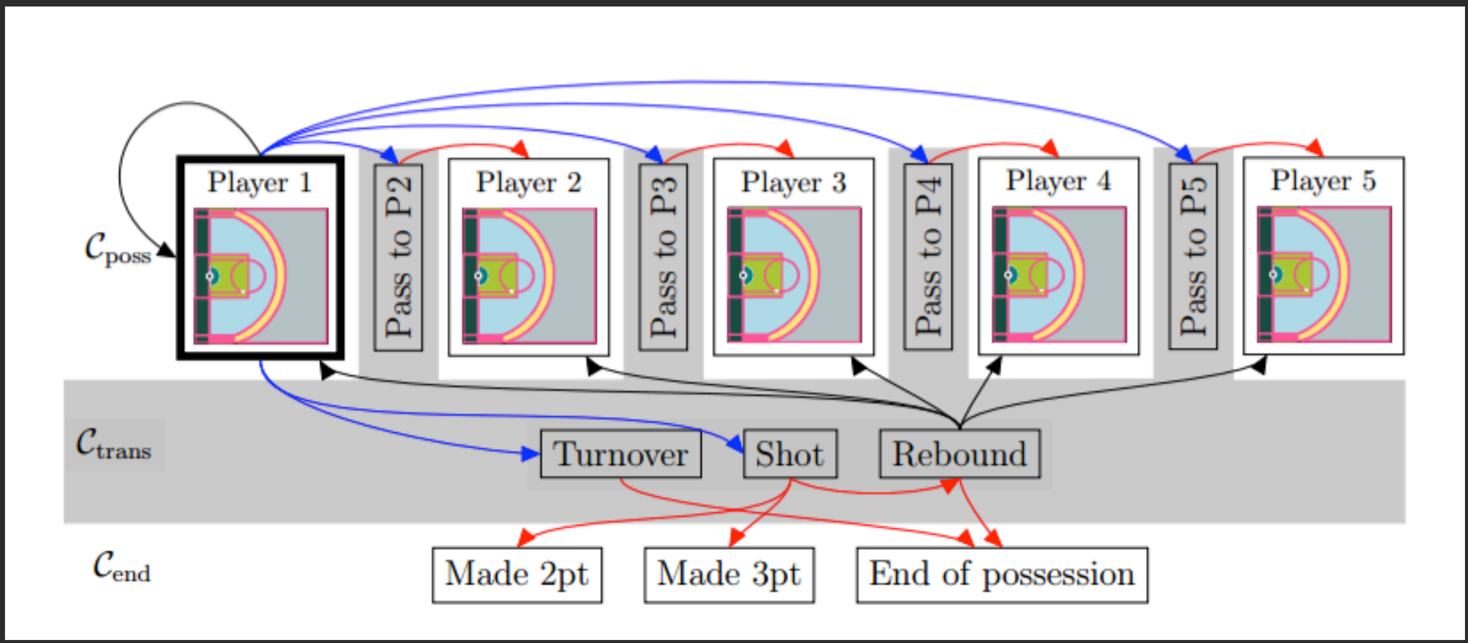
$$C_t = C(z_t) \in \{C_{poss}, C_{trans}, C_{end}\}$$

$$C_{poss} = \{\text{player ID}\} \times \{\text{region ID 1-of-7}\} \times \{\text{is guarded}\}$$

$$C_{trans} = \{\text{shot, pass, TO, rebound}\}$$

$$C_{end} = \{\text{2pt, 3pt, end}\}$$

EPV-schematic



EPV-combining C and Z

Assumption1: C is marginally semi-Markov

Assumption2:

For all $s > \delta_t$ and $c \in \mathcal{C}$, $\mathbb{P}(C_s = c | C_{\delta_t}, \mathcal{F}_t^{(Z)}) = \mathbb{P}(C_s = c | C_{\delta_t})$.

$$\tau_t = \begin{cases} \min\{s : s > t, C_s \in \mathcal{C}_{\text{trans}}\} & \text{if } C_t \in \mathcal{C}_{\text{poss}} \\ t & \text{if } C_t \notin \mathcal{C}_{\text{poss}} \end{cases}$$

$$\delta_t = \min\{s : s \geq \tau_t, C_s \notin \mathcal{C}_{\text{trans}}\}.$$

$$\nu_t = \sum_{c \in \mathcal{C}} \mathbb{E}[X | C_{\delta_t} = c] \mathbb{P}(C_{\delta_t} = c | \mathcal{F}_t^{(Z)}).$$

EPV-model

$$\nu_t = \sum_{c \in \mathcal{C}} \mathbb{E}[X | C_{\delta_t} = c] \mathbb{P}(C_{\delta_t} = c | \mathcal{F}_t^{(Z)}).$$

denote $M(t)$ as the occurence of 'decoupling' event before next frame

$\mathbb{P}(C_{\delta_t} | F_t^{\tilde{z}}) = \mathbb{P}(C_{\delta_t} | M(t), F_t^{\tilde{z}}) \mathbb{P}(M(t) | F_t^{\tilde{z}})$ (macrotransition)

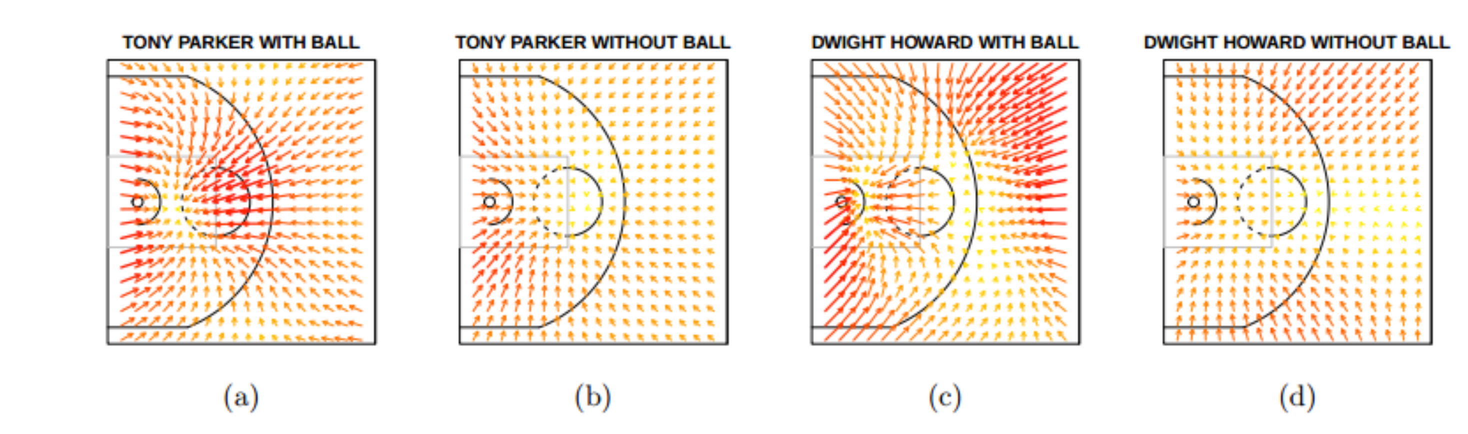
$\mathbb{P}(z_{t+} | M(t)^c, F_t^{\tilde{z}})$ (microtransition)

The *Markov transition probability matrix* $\mathbf{P}$, with $P_{qr} = \mathbb{P}(C^{(n+1)} = c_r | C^{(n)} = c_q)$.

EPV-microtransition model

Offense: players move based on “who”, “velocity”, “where”

$$x^\ell(t+\epsilon) = x^\ell(t) + \alpha_x^\ell[x^\ell(t) - x^\ell(t-\epsilon)] + \eta_x^\ell(t)$$



EPV-macro entry

$\mathbb{P}(M(t)|\mathcal{F}_t^{(Z)}) = \sum_{j=1}^{6} \mathbb{P}(M_j(t)|\mathcal{F}_t^{(Z)})$.

We parameterize the macrotransition entry models as competing risks (Prentice, Kalbfleisch, Peterson Jr, Flournoy, Farewell & Breslow 1978): assuming player ℓ possesses the ball at time $t > 0$ during a possession, denote

$$\lambda_j^{\ell}(t) = \lim_{\epsilon \to 0} \frac{\mathbb{P}(M_j(t)|\mathcal{F}_t^{(Z)})}{\epsilon} \tag{7}$$

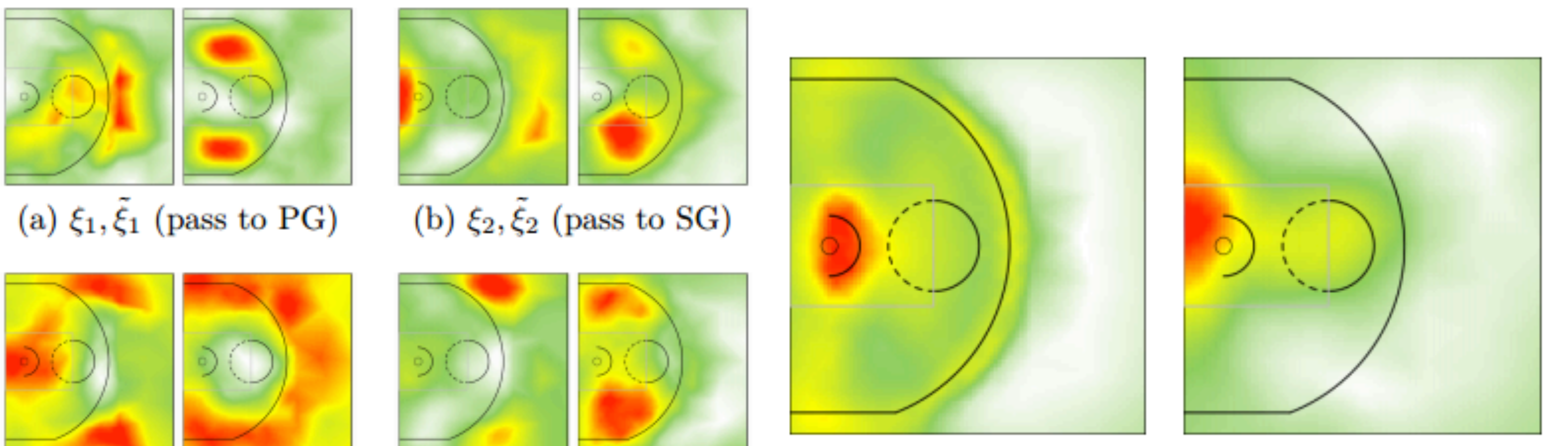
as the hazard for macrotransition j at time t. We assume these are log-linear,

$$\log(\lambda_j^{\ell}(t)) = [\mathbf{W}_j^{\ell}(t)]' \boldsymbol{\beta}_j^{\ell} + \xi_j^{\ell}\left(\mathbf{z}^{\ell}(t)\right) + \left(\tilde{\xi}_j^{\ell}\left(\mathbf{z}_j(t)\right) \mathbf{1}[j \leq 4]\right), \tag{8}$$

where $\mathbf{W}_j^{\ell}(t)$ is a $p_j \times 1$ vector of time-varying covariates, $\boldsymbol{\beta}_j^{\ell}$ a $p_j \times 1$ vector of coefficients, $\mathbf{z}^{\ell}(t)$ is the ballcarrier's 2D location on the court (denote the court space $\mathbb{S}$) at time t, and $\xi_j^{\ell} : \mathbb{S} \to \mathbb{R}$ is a mapping of the player's court location to an additive effect on the log-hazard,

EPV-macro entry

All model components—the time-varying covariates, their coefficients, and the spatial effects $\xi, \tilde{\xi}$ differ across macrotransition types j for the same ballcarrier ℓ, as well as across all $L = 461$ ballcarriers in the league during the 2013-14 season. This reflects the fact that



(a) $\xi_1, \tilde{\xi}_1$ (pass to PG)

(b) $\xi_2, \tilde{\xi}_2$ (pass to SG)

(c) ξ_5 (shot-taking)

(d) ξ_6 (turnover)

(e) $\xi_3, \tilde{\xi}_3$ (pass to PF)

(f) $\xi_4, \tilde{\xi}_4$ (pass to C)

EPV-macro exit

$$\mathbb{P}(C_{\delta_t}|M(t),\mathcal{F}_t^{(Z)}) = \sum_{j=1}^{6}\mathbb{P}(C_{\delta_t}|M_j(t),\mathcal{F}_t^{(Z)})\mathbb{P}(M_j(t)|M(t),\mathcal{F}_t^{(Z)})$$

$$= \sum_{j=1}^{6}\mathbb{P}(C_{\delta_t}|M_j(t),\mathcal{F}_t^{(Z)})\frac{\lambda_j^{\ell}(t)}{\sum_{k=1}^{6}\lambda_k^{\ell}(t)},$$

Turnover -> “end”

Pass -> continue w

$$\text{logit}(p^{\ell}(t)) = [\mathbf{W}_{\text{s}}^{\ell}(t)]'\boldsymbol{\beta}_{\text{s}}^{\ell} + \xi_{\text{s}}^{\ell}(\mathbf{z}^{\ell}(t))$$

Shot ->

EPV-learning

$$\prod_t \mathbb{P}(Z_{t+\epsilon}|\mathcal{F}_t^{(Z)}) = \Bigg(\overbrace{\prod_t \mathbb{P}(Z_{t+\epsilon}|M(t)^c, \mathcal{F}_t^{(Z)})^{\mathbf{1}[M(t)^c]}}^{L_{\text{mic}}}\Bigg)\Bigg(\overbrace{\prod_t \prod_{j=1}^{6} \mathbb{P}(Z_{t+\epsilon}|M_j(t), C_{\delta_t}, \mathcal{F}_t^{(Z)})^{\mathbf{1}[M_j(t)]}}^{L_{\text{rem}}}$$

$$\times \Bigg(\underbrace{\prod_t \mathbb{P}(M(t)^c|\mathcal{F}_t^{(Z)})^{\mathbf{1}[M(t)^c]} \prod_{j=1}^{6} \mathbb{P}(M_j(t)|\mathcal{F}_t^{(Z)})^{\mathbf{1}[M_j(t)]}}_{L_{\text{entry}}}\Bigg)\Bigg(\underbrace{\prod_t \prod_{j=1}^{6} \mathbb{P}(C_{\delta_t}|M_j(t), \mathcal{F}_t^{(Z)})^{\mathbf{1}[M_j(t)]}}_{L_{\text{exit}}}\Bigg) \tag{15}$$

The factorization used in (15) highlights data features that inform different parameter groups: L_{mic} is the likelihood term corresponding to the microtransition model (M1), L_{entry} the macrotransition entry model (M2), and L_{exit} the macrotransition exit model (M3). The remaining term L_{rem} is left unspecified, and ignored during inference. Thus, L_{mic}, L_{entry}, and

EPV-learning

inference. Thus, microtransition models are fit in parallel using each player's data separately; this requires $L = 461$ processors, each taking at most 18 hours at 2.50Ghz clock speed, using 32GB of RAM.

on macrotransition type. We perform this regression through the use of integrated nested Laplace approximations (INLA) (Rue, Martino & Chopin 2009). Each macrotransition type can be fit separately, and requires approximately 24 hours using a single 2.50GHz processor with 120GB of RAM.

EPV-result

Compare “transition probability” with simpler baselines

	Model Terms			
Macro. type	Player	Covariates	Covariates + Spatial	Full
Pass1	-29.4	-27.7	-27.2	-26.4
Pass2	-24.5	-23.7	-23.2	-22.2
Pass3	-26.3	-25.2	-25.3	-23.9
Pass4	-20.4	-20.4	-24.5	-18.9
Shot Attempt	-48.9	-46.4	-40.9	-40.7
Made Basket	-6.6	-6.6	-5.6	-5.2
Turnover	-9.3	-9.1	-9.0	-8.4

Table 1: Out of sample log-likelihood (in thousands) for macrotransition entry/exit models under various model specifications. “Player” assumes constant hazards for each player/event type combination. “Covariates” augments this model with situational covariates, $\mathbf{W}_j^\ell(t)$ as given in (8). “Covariates + Spatial” adds a spatial effect, yielding (8) in its entirety. Lastly, “Full” implements this model with the full hierchical model discussed in Section 4.

Conclusion

SportVU data opens up a new perspective on analytics (not just basketball)

Difference approaches (i.e. stats, ML, engineering, …)