Classifying NBA Plays & Predicting Shots

Jackson Wang
Feb 27, 2017

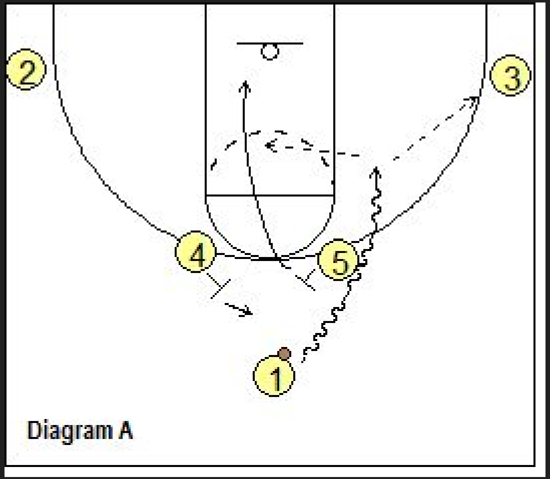
Today

- Some of my own work in the past
  - Introduction to SportVU data, some of the challenges/opportunities
  - What's important when we work in sports (my opinion)
  - An appetizer for a more general discussion about modern learning in sports

Wang, K. C., & Zemel, R. (2016). classifying NBA offensive plays using neural networks. MIT Sloan Sports Analytics Conference.

High Level Goal

"Classify the offensive play of a given sequence"



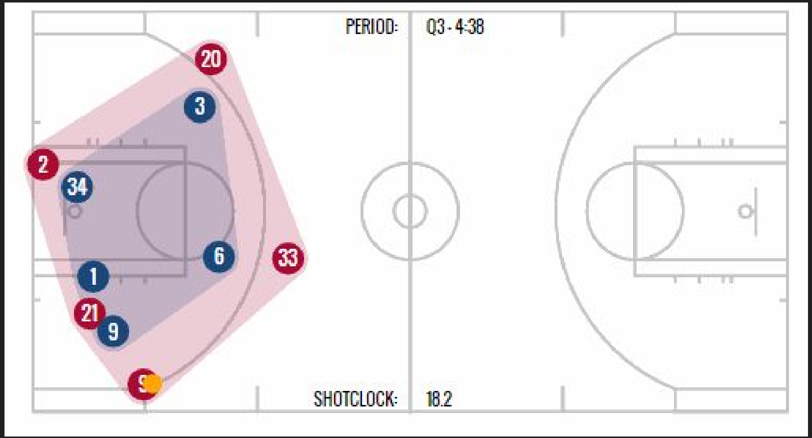
Some motivations

- Basketball Operations
  - Organize our team's history
    - Success rate of different plays
    - Play retrieval
  - Scouting the opponents
    - Knowledge for designing our defense
    - How is this done now?
- Other applications
  - Mass Media (entertainment)
    - Automatic high-level annotations
  - Gaming (entertainment)
    - More realistic game agents
- ....

* I use 'play', 'playcall', 'offensive strategy' interchangeably

Data

- SportVU data
  - (x,y,z) for ball, (x,y) for 10 players @ 25 Hz
  - Play-by-play annotations (much like what you see on nba.com)
  - Private?*
- Labels
  - Human labels provided by Raptors



https://github.com/linouk23/NBA-Player-Movements
https://github.com/neilmj/BasketballData

You're given this

… (x,y),(x,y),(x,y) …

… (x,y),(x,y),(x,y) …

… (x,y),(x,y),(x,y) …

… (x,y),(x,y),(x,y) …

… (x,y),(x,y),(x,y) …

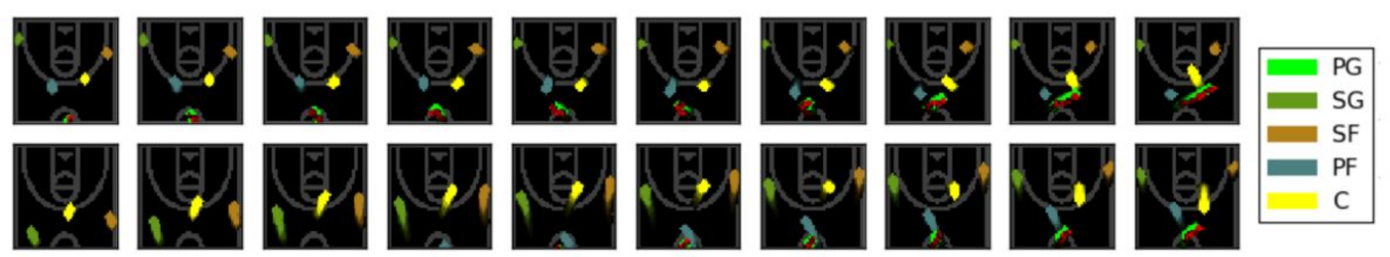
{..., 'pistol', 'fist', 'horns', 'horns X', 'horns fist', 'horns 53' ,…}

Data & Problem Definition

- Given SportVU sequences of plays, returns 1-of-K labels
  - >100 labels
    - Hierarchical in nature
    - Very unbalanced dataset
  - >7k sequences
    - When does it start/end?
- Pre-preprocessing
  - ???

Problem Definition

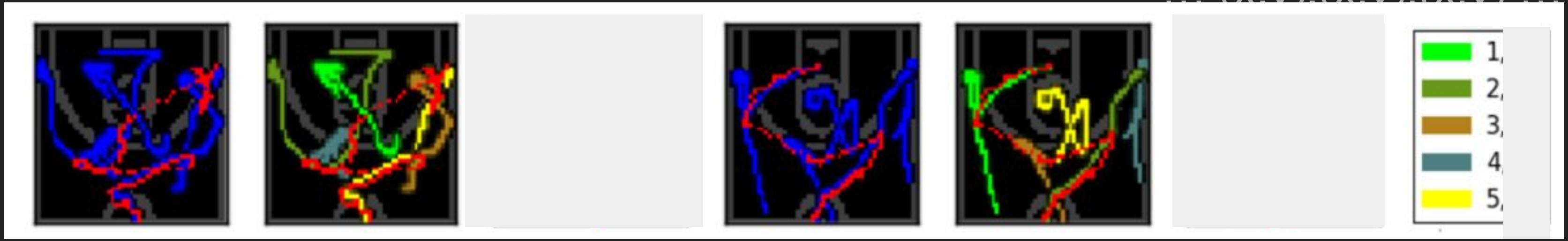
- Given SportVU sequences of a play, returns 1-of-K labels
  - 11 selected classes (from >100 labels)
  - 1435 sequences (from >7k sequences)
- Pre-preprocessing
  - **Data filtering**
    - Fairly open problem ...
  - **Temporal Segmentation**
    - A couple of seconds after the ball crosses the halfcourt
    - limitations?
  - Player Identifiability



Are they the same?

… (x,y),(x,y),(x,y) …

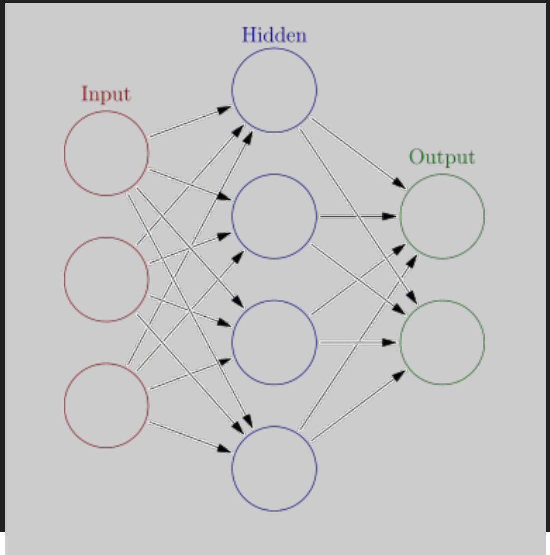
… (x,y),(x,y),(x,y) …

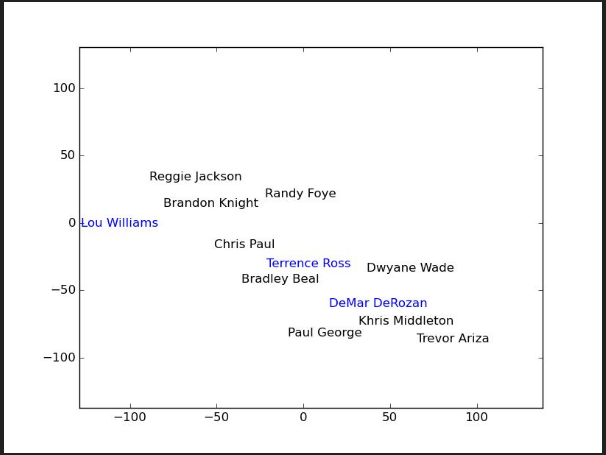


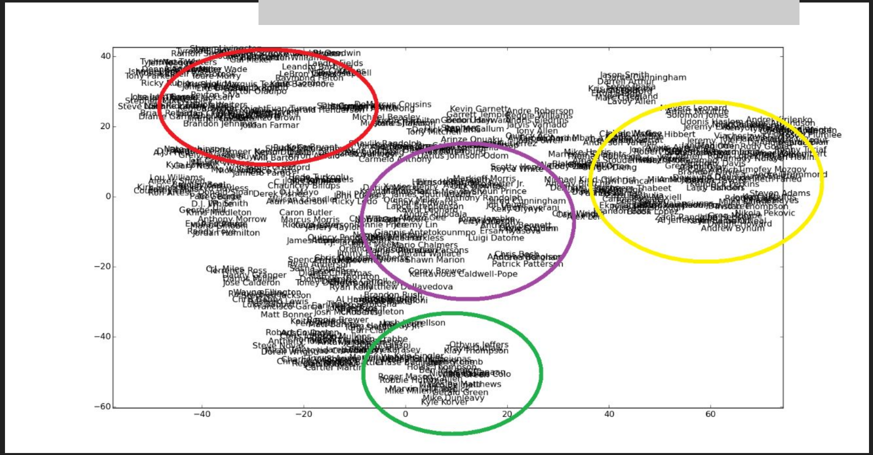
… (x,y),(x,y),(x,y) …

Player identifiability (I)

- Player position learned from simple NN classifier
  - Input = player id, output = sets of engineered features
  - “Embed players by their shooting tendencies”
  - Limitations?
  - Evaluation?







Player identifiability (end)

$$\text{argmin}_{p_{i,j}} \mathbb{I}[p_{i,j}]d(\text{emb}(p_j) - c_j)$$

$$i \in \{1, 2, 3, 4, 5\}, \quad j \in \{\text{pg}, \text{sg}, \text{sf}, \text{pf}, \text{c}\}$$

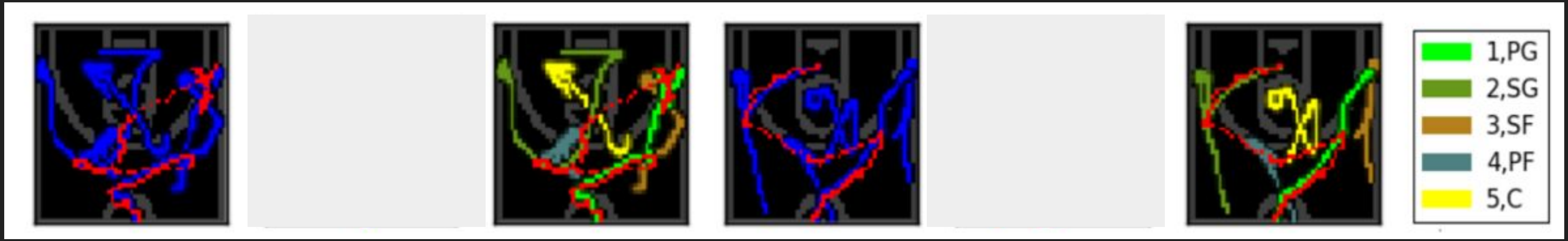
$\text{emb}(\cdot)$ is the learned embedding function

Table 3.1: Example of player position resolution. Here are 3 possible line-ups of the Toronto Raptors where Lou Williams, Terrence Ross, and DeMar DeRozan were assigned different positions depending on the line-up.

Kyle Lowry	Lou Williams	Terrence Ross	DeMar DeRozan	Jonas Valanciunas
Kyle Lowry	Lou Williams	DeMar DeRozan	Amir Johnson	Jonas Valanciunas
Lou Williams	Terrence Ross	DeMar DeRozan	Amir Johnson	Jonas Valanciunas

Are they the same?

YES



Now we have something like this

PG: T_0 -> … (x,y),(x,y),(x,y) … T_end

SG: T_0 -> … (x,y),(x,y),(x,y) … T_end

SF: T_0 -> … (x,y),(x,y),(x,y) … T_end

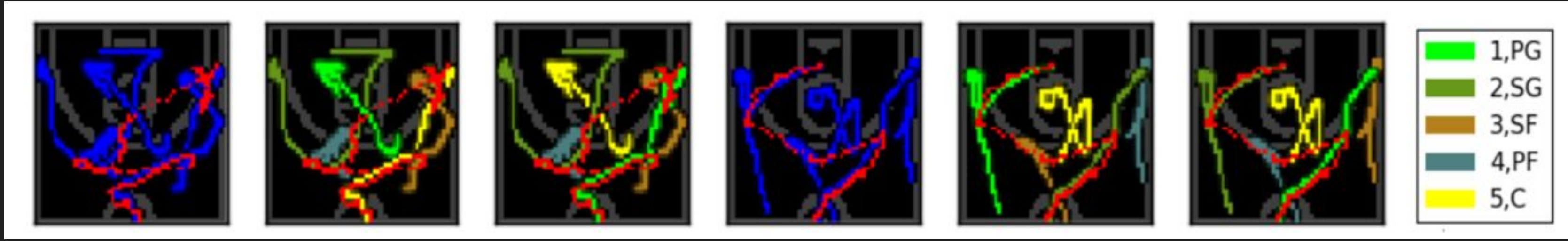
PF: T_0 -> … (x,y),(x,y),(x,y) … T_end

C: T_0 -> … (x,y),(x,y),(x,y) … T_end

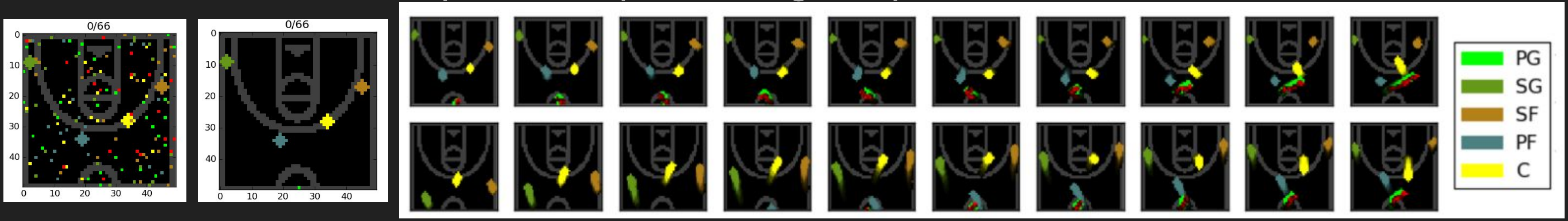
{‘pistol’, ‘fist’, ‘horns’} 1-of-K

Data Representation

- Data representation
  - (x,y) sequences? Low dimensional, but does not fit with modern NN tools
  - Images? Sparse signal, but fits with modern NN tools

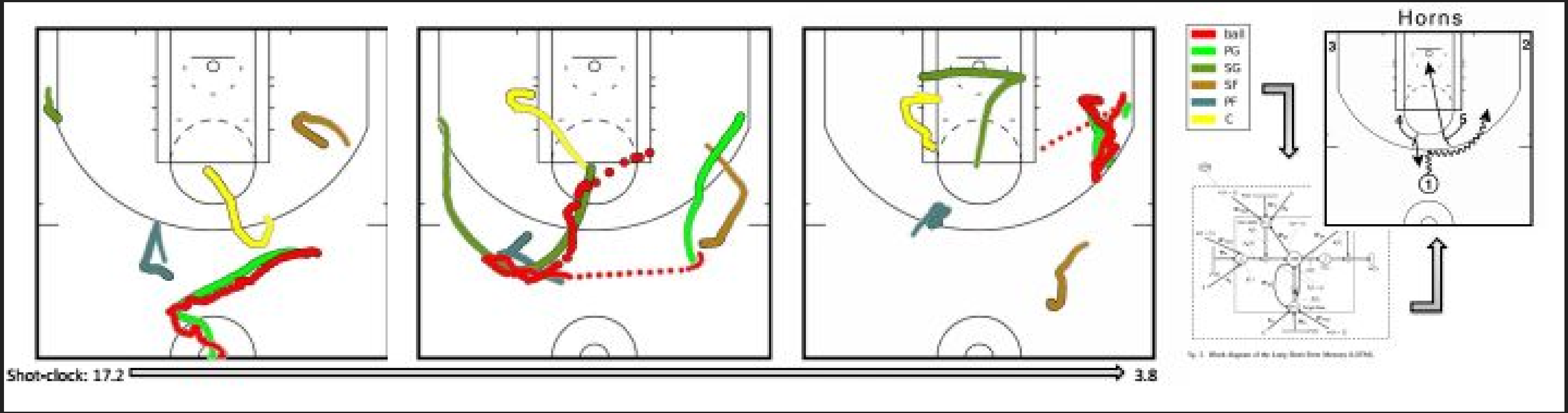


- Loses temporal information
- How does this compare with preserving temporal information?



Our classification task

1435 pairs -> 95 test, 1340 train+validation

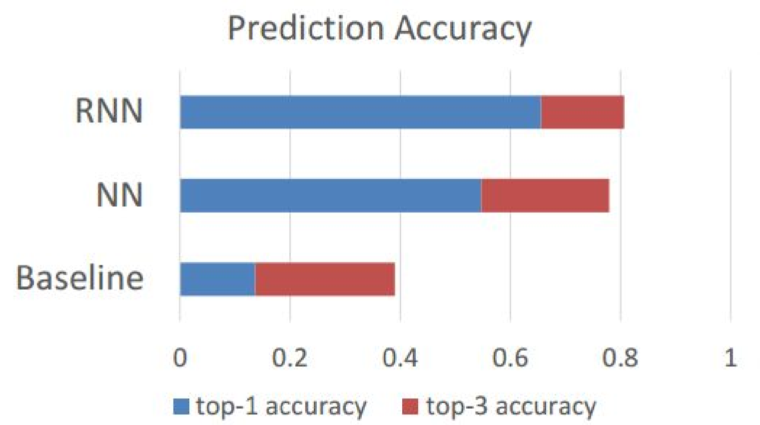


Results

Table 3.2: Classification Accuracy

Model	Top-1 accuracy	Top-3 accuracy		
base-rate	.137	.390	n/a	n/a
NN				
RNN				

Results



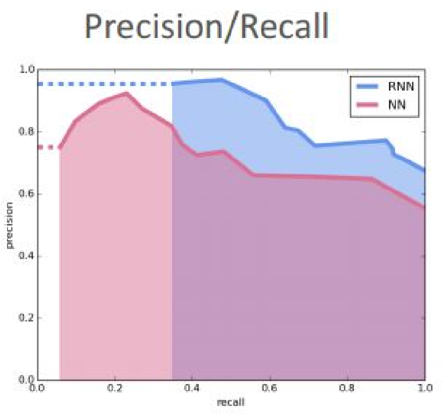
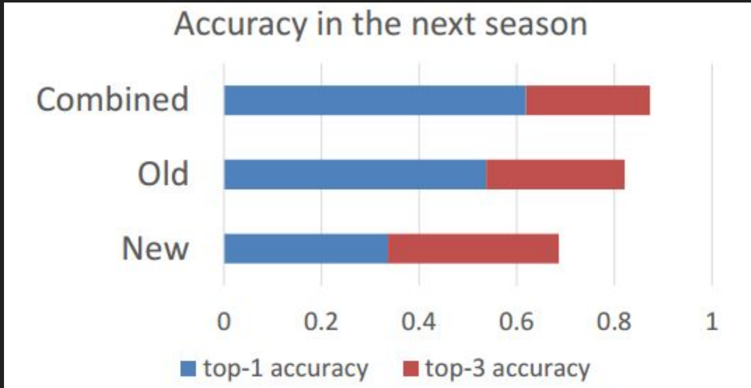


Table 3.2: Classification Accuracy

Model	Top-1 accuracy	Top-3 accuracy	Precision/Recall at T=.4	Precision/Recall at T=.7
base-rate	.137	.390	n/a	n/a
NN	.547	.779	.724/.412	.909/.196
RNN	.656	.806	.727/.918	.900/.590

Next season

Simulate early in season: given with 1st 3 months of data (N=327), train with 1/3



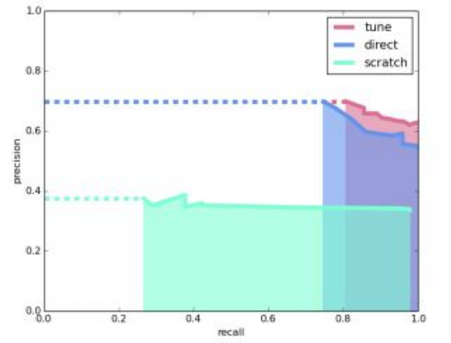


Table 3.3: Classification Acccuracy on the new season

Model	Top-1 accuracy	Top-3 accuracy	Precision/Recall at T=.4	Precision/Recall at T=.7
new	.336	.686	.336/.978	.347/.378
transfer	.537	.821	.541/1.0	.591/.958
fine-tune	.619	.873	.629/1.0	.639/.927

Failure mode/Limitations?

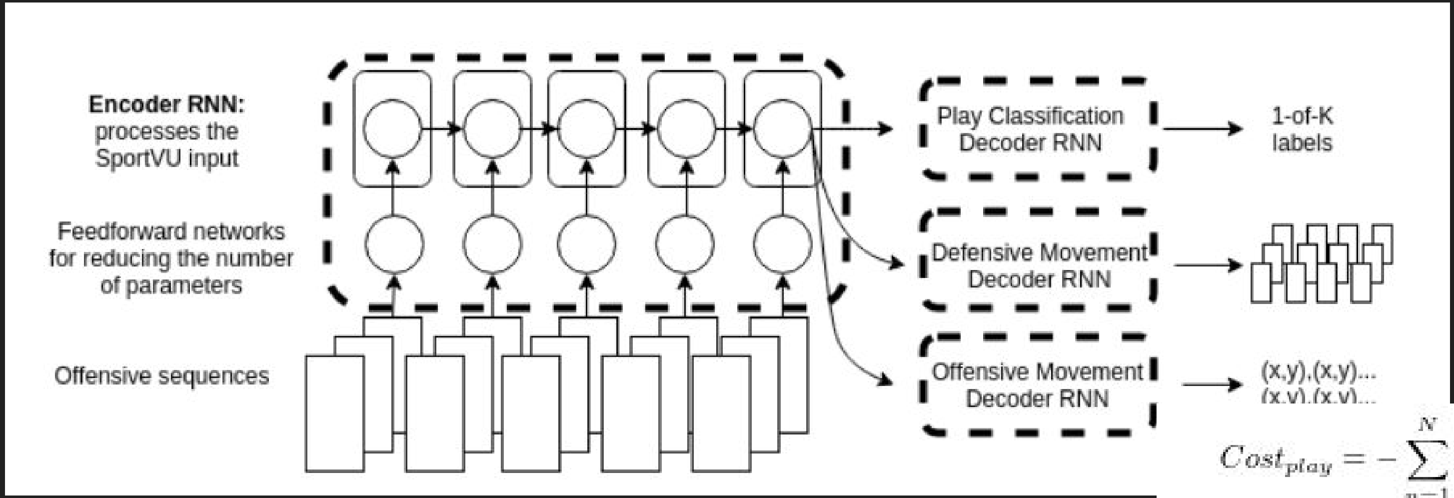
- Failure modes
  - Confuses between sibling classes
  - Some confounding factors to input (e.g. play interruption, defensive success)
  - Very long sequences (some plays last for only 4 seconds, but some >14)
- Limitations given the scope of our task definition?
- **Problems with our task definition?**
  - **Proper structure to the labels**
  - **Play-by-play annotations**

Auxiliary Tasks

- Supervised method is probably not the best way to go
  - Labels are hard to get
  - Not a lot of signal from label
  - ….
- Other types of learning signals

Auxiliary Tasks

Supervised method is probably not the best way to go



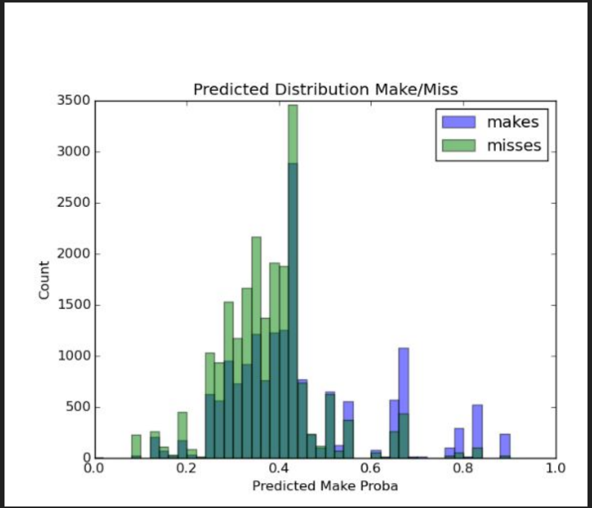
$$Cost_{play} = -\sum_{n=1}^{N} \log P(\hat{y}_n = y_n | \mathbf{X}_n)$$

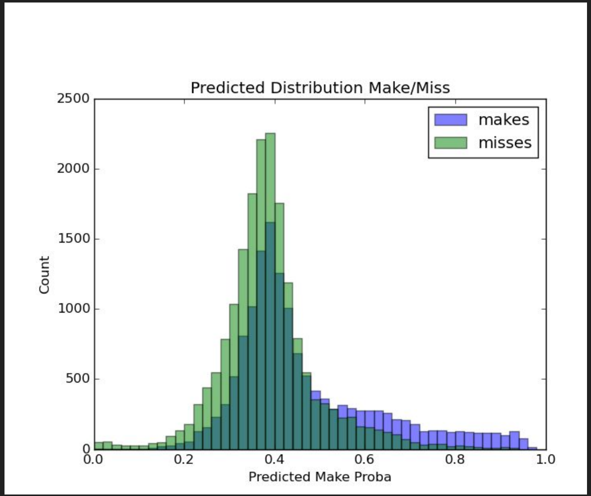
$$Cost_{seq} = \sum_{n=1}^{N} \sum_{i=1}^{5} \sum_{t} \|x_{n,i,t} - \hat{x}_{n,i,t}\|_2$$

$$Cost_{def} = \sum_{n=1}^{N} \sum_{p=1}^{50\text{x}50} \sum_{t} (x_{n,p,t} - \hat{x}_{n,p,t})^2$$

Shot Prediction

“What’s the expected outcome of a given shot?”



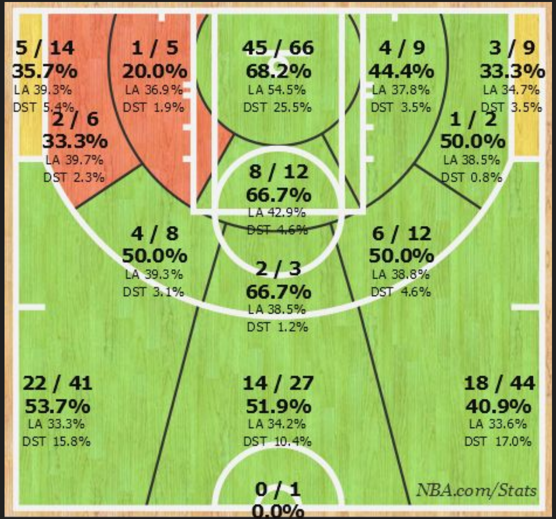


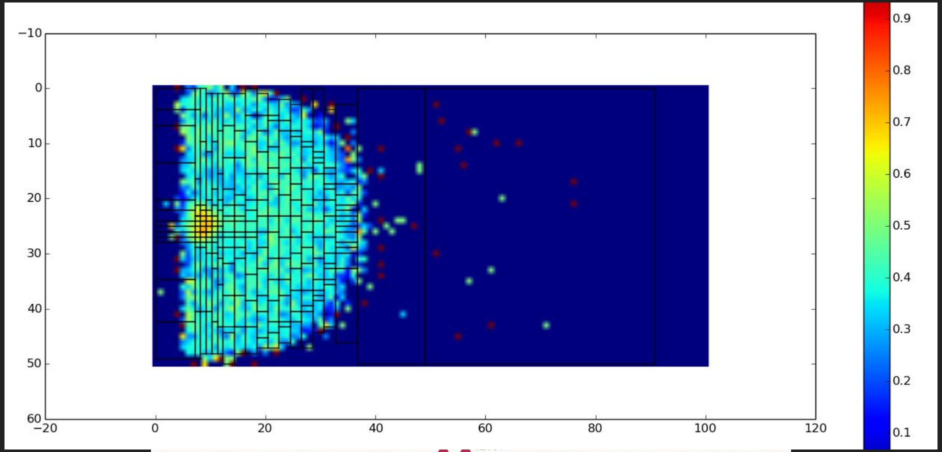
What’s important

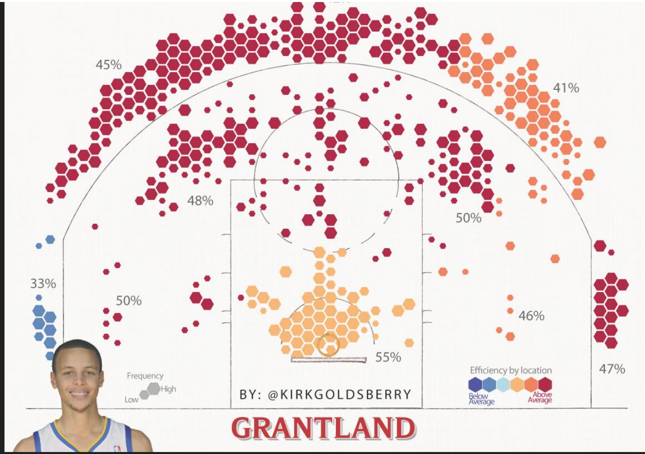
- Important factors to a shot
  - Contest-level, distance from basket, player’s movement history ….
  - **Where he is**
  - **Who is shooting**
- Evaluating our model

Spatial representation

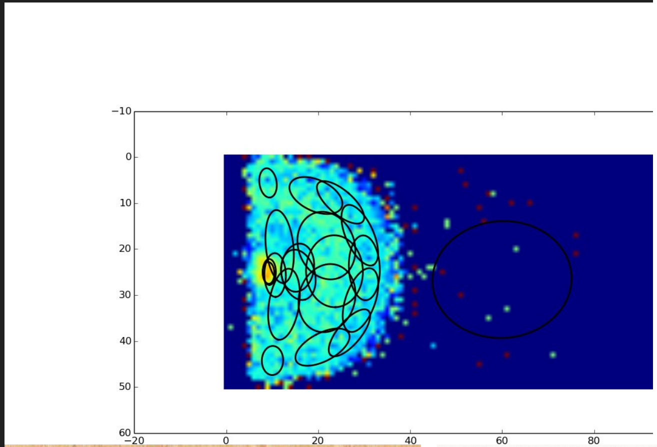
(x,y)?

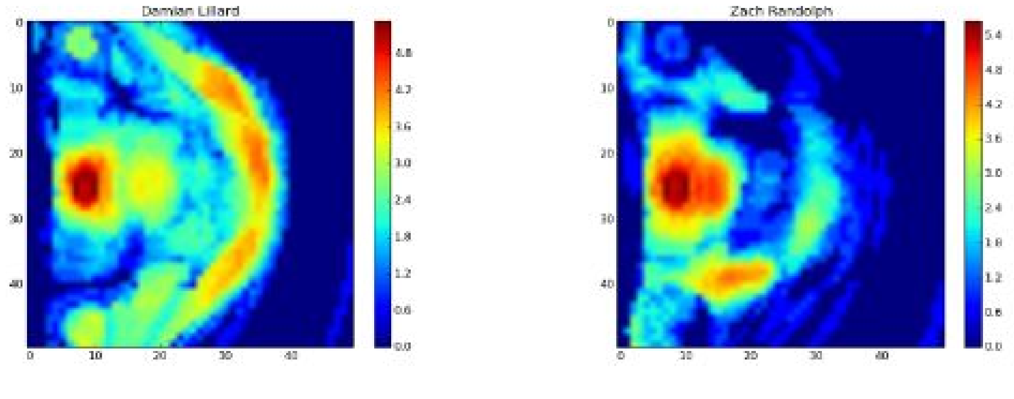




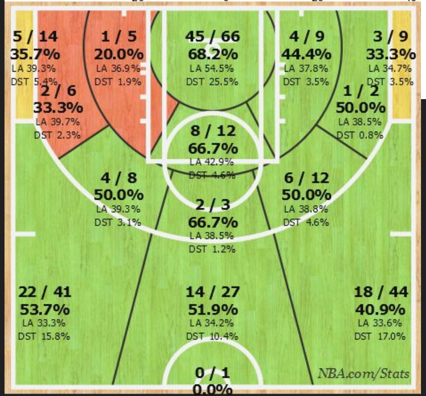


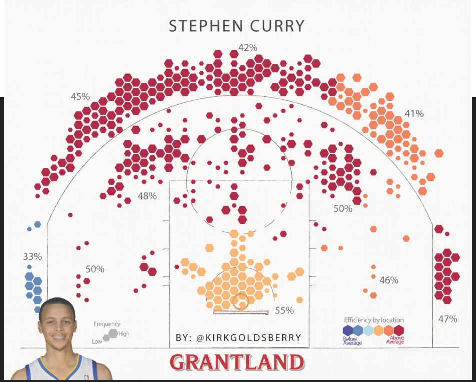
Spatial representation

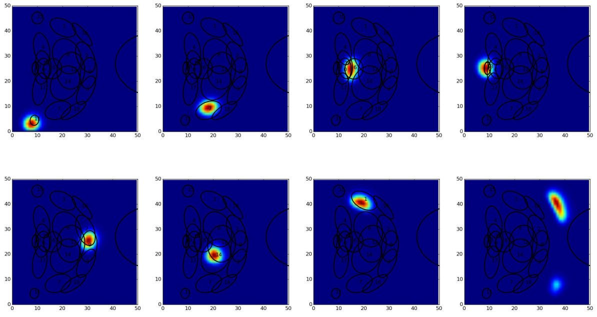




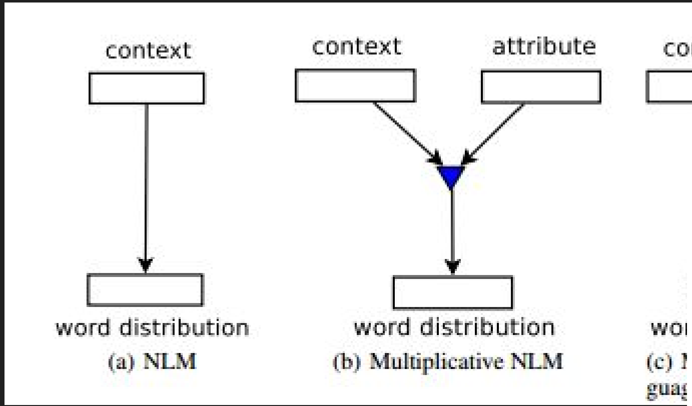
(a) Damian Lillard (b) Zach Randolph







Modelling player's effect



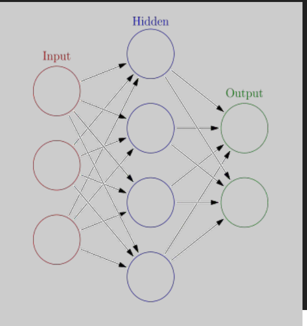


Table 2.1: Classification Acccuracy and Data Likelihood for all methods

Model	NLL			ACC		
	train	valid	test	train	valid	test
Baseline	0.6614	0.6601	0.6643	0.6195	0.6223	0.6153
LR	-0.0057	-0.0052	-0.0077	-0.0036	-0.0075	-0.0033
FB	-0.2216	nan	nan	0.132	-0.0327	-0.0323
LR-b	-0.0121	-0.013	-0.0132	0.0066	0.0063	0.0052
NN	-0.0182	-0.0165	-0.017	0.0112	0.0045	0.0074
NN-b	-0.0063	-0.0061	-0.0048	0.008	0.0056	0.0047
NN-add	-0.0205	-0.0191	-0.0197	0.014	0.0089	0.0099
NN-tensor	-0.0209	-0.0195	-0.0198	0.0143	0.0062	0.0077

Kiros, R., Zemel, R., & Salakhutdinov, R. R. (2014). A multiplicative model for learning distributed text-based attribute representations. InAdvances in neural information processing systems (pp. 2348-2356).

Today

- Some of my own work in the past
  - Offensive play classification
  - Shot prediction: neural representation of players, basketball court in alternative space
- Introduction to SporVU data, some of the challenges/opportunities
  - Data representation, segmentation, identification/grouping, label collection
- What’s important when we work in sports (my opinion)
  - Evaluation!
- An appetizer for a more general discussion about modern learning in sports
  - Player metrics
  - Game/Performance analysis
  - Strategy development
  - ….

Wang, K. C., & Zemel, R. (2016). classifying NBA offensive plays using neural networks. MIT Sloan Sports Analytics Conference.

Thanks!

Stay tuned for next week …