The Adaptability Index: An Information Theoretic Framework for Measuring Player Adaptability

Basketball

20251453

1. Introduction

Since the "super team" era in the NBA started, understanding a basketball player's offensive adaptability has been crucial to team construction. This holds true of all positions, skill levels, and usage rates. The best, highest usage players are expected to team up with other high usage players and must adapt to each other's playstyles in a way that maximizes the output from all players. Power forwards and centers are not only expected to grab rebounds and score from in the paint, but also space the floor and score further away from the basket more than ever. The "3" or small forwards are increasingly taking advantage of all the space, running offenses and creating opportunities for others as well as creating their own shots from all areas of the court. Even point guards are getting bigger and stronger, not just shooting from the perimeter but also scoring in the paint when available. Without at least a couple players on the team that are adaptable, it seems that offenses can become stagnant and predictable and star players' potentials can be wasted.

Despite the importance of adaptability, there are very few metrics that measure offensive adaptability. Versatility[1], developed by John Hollinger and defined in equation (1) below, measures the ability for a player to score, rebound, and assist *all the time*, but not selectively choose when to score, rebound, and assist *at any time* given the personnel around them. Players can be versatile, while not adapting their playstyles.

$$[(PPG) \cdot (RPG) \cdot (APG)]^{1/3} \tag{1}$$

There exists another metric, portability, which is also crucial to team construction. Portability hasn't been rigorously defined by many however, craftedNBA[2] does have their own definition which is a weighted average of portable skills like low usage rates, high true shooting percentage, and high assisting. Portability is meant to measure the ability for a player to succeed regardless of who's around them. Another interpretation is that portability is a measure of how easy it is for other players to adapt to the new player. Certainly, a valuable metric but doesn't quite capture how adaptable the player is themselves.

This work aims to define offensive adaptability and build a metric that adequately quantifies that definition.

2. Methods

The existing metrics that are close to adaptability are missing the concept of *changing* playstyle given the lineup. To adequately define adaptability, the metric must include the concept of changing playstyle. Secondly, the playstyle must be changing *predictably* with lineup. A player cannot be adapting if the change in playstyle from lineup to lineup is random.

A start for this metric is some sort of correlation. Correlations measure the relation of a change in one variable to the change in another. For example, Pearson's correlation measures how linearly related one variable is to another, with a correlation of 1 representing that a change in one variable

perfectly predicts the change in another and a correlation of 0 representing a change in one variable having no predictive power for the change in another, assuming a linear relationship. So, adaptability could be defined as the correlation between lineup and playstyle where the higher the correlation, the more likely a change in lineup results in a change in playstyle and the lower the correlation, the less likely a change in lineup results in a change in playstyle.

The natural selection at first thought might be Pearson's correlation. However, using Pearson's correlation presents some challenges. Lineup is a discrete, non-ordinal variable. Pearson's correlation only properly relates linear relationships, which without even defining playstyle, one might think that there is not much "linear" about playstyle. Lastly, Pearson's correlation can only relate two scalar variables. Playstyle most likely can't get distilled down to a single scalar without losing too much information.

There is a concept, mutual information, from information theory that correlates random variables. Mutual information makes no assumptions about the distributions of the random variables, the relationship of the random variables, whether they are discrete or continuous, or the dimensionality of the variables – making it a strong candidate for correlating lineup and playstyle. Adaptability will be defined as the mutual information between lineups that a player participates in and the player's playstyle.

2.1. Information Theory Primer
The following section is a brief introduction to information theory which will build intuition on the utility of mutual information and will provide the context for a more nuanced interpretation of future results that include information theoretic concepts.

2.1.1. Entropy and Information Content
To get to the idea of shared information between lineup and playstyle, "information" must first be defined. The information theoretic definition of information called "entropy" is used. Entropy in information theory measures the spread of a probability mass or probability density function. In other words, how random the process is that follows the function. The definition is below:

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) \tag{2}$$

Where: X is a discrete random variable
$H(\cdot)$ is the entropy of a discrete random variable
$p(\cdot)$ is the probability mass function for a discrete random variable
$\mathcal{X}$ is the set of values that X can take on
x is a realization of X sampled from the set $\mathcal{X}$

To illustrate the utility of the equation, imagine a biased coin that has heads and tails. The probability of heads is 100% leaving the probability of tails as 0%. Substituting into equation (2), $H(X) = -(1 \cdot \log 1) - (0 \cdot \log 0)$. In this context, the second term is defined as 0 or $0 \cdot \log 0 \equiv 0$. The resulting entropy is 0. The function has no spread or randomness. Sometimes, entropy is described as the average information content of a distribution. $-\log p(x)$ is large when lower probabilities exist in the distribution and these low probability outcomes are considered "surprising" which can carry more information. Then entropy would be the average of all the events, which will be larger if there are more surprising (low probability) events. There is an analog for continuous distributions

where the summation is substituted with integrals and the probability mass function is substituted with a probability density function.

$$h(X) = - \int_X f(x) \log f(x) \, dx \tag{3}$$

Where: $h(\cdot)$ is the differential or "continuous" entropy of a continuous random variable
$f(\cdot)$ is the probability density function for a continuous random variable
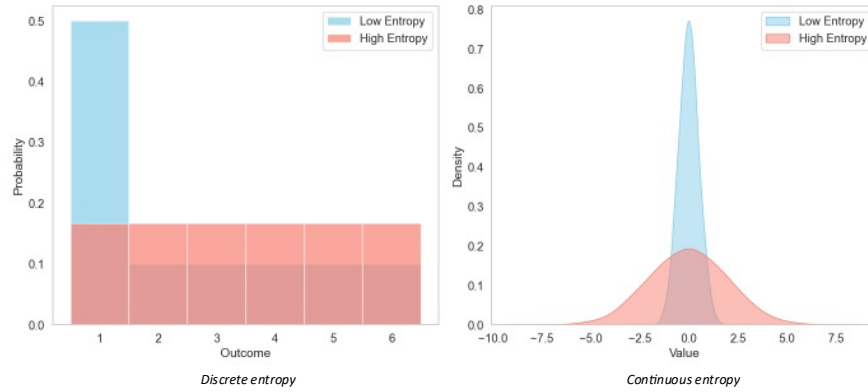

Figure 1: Low and high entropy for discrete and continuous distributions

2.1.2. Mutual Information

Now that there is a framework for getting the information content of a distribution, what does it mean for two distributions to share information? One could imagine that for two random variables, knowledge of the outcome of one variable might contain some information about the outcome of the other. This is analogous to Pearson's correlation r, where r^2 denotes how much variance in one variable can be explained by the variance in another. Another interpretation of two distributions sharing information is that knowledge about one variable reduces the average information content in another. This is defined mathematically in equation (4):

$$I(X;Y) = H(Y) - H(Y|X) \tag{4}$$

Where: $I(\cdot\,;\cdot)$ is the mutual information between two random variables
$H(\cdot\,|\,\cdot)$ is the conditional entropy – Entropy of first variable given knowledge of the second

This says that random process Y has some information content but knowledge about X reduces the information Y contains. The amount of that reduction is the mutual information. This is illustrated in Figure 2 where Y is a continuous random variable and X is a discrete random variable taking on two values. The marginal distribution, $f(Y)$, is flatter and contains some amount of information. Knowing X decreases the variance of Y on average. The information content of Y is reduced by the shared information between X and Y.
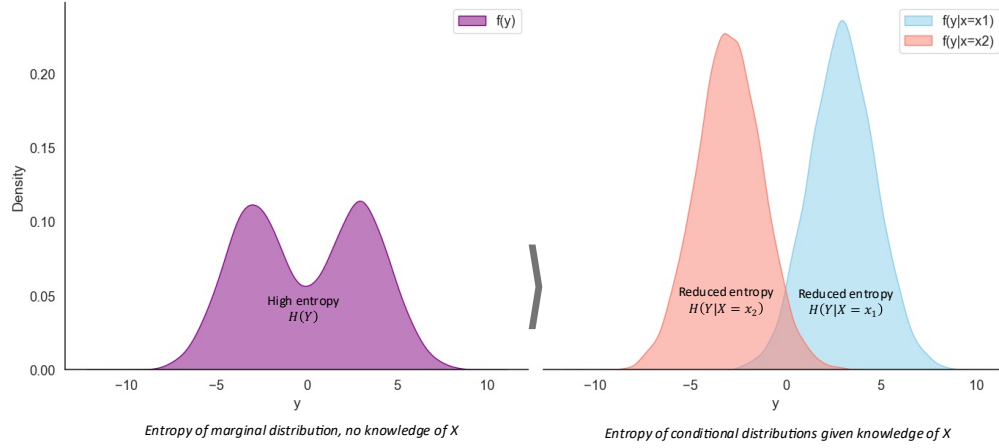
Figure 2: Illustration of reducing information content (entropy) through shared (mutual) information

Mutual information can be written in terms of probability mass functions.

$$I(X; Y) \equiv \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right) \tag{5}$$

This shows that if two random variables are independent, $p(x, y) = p(x)p(y)$, then the mutual information is 0. Mutual information can also be written in the continuous case with integrals

$$I(X; Y) \equiv \iint_{x \in \mathcal{X}, y \in \mathcal{Y}} f(x, y) \log\left(\frac{f(x, y)}{f(x)f(y)}\right) dy dx \tag{6}$$

2.2. Estimating Mutual Information

Although there is a clear definition of mutual information, computing its value still poses a challenge. For now, it's assumed that "playstyle" contains some continuous component, requiring the integral form of mutual information to be used. However, $f(x)$ is not known – only observations sampled from $f(x)$. $f(x)$ could be estimated using a kernel density estimator, however computing the integral is still a challenge. A Markov Chain Monte Carlo (MCMC) estimator could be used, however if "playstyle" is multi-dimensional, MCMC will also be challenging to compute.

It is possible to directly estimate the entropy of a random variable using a k-nearest neighbor approach like the Kozachenko-Leonenko estimator[3] which is what's used in this work. The derivation of the estimator is as follows:

It is observed from equation (3) that entropy is an expectation of $\log f(x)$. An estimate of the expectation would be to forego the integral and directly calculate

$$-\frac{1}{N} \sum_{i=1}^{N} \log f(x_i) \tag{7}$$

Where: N is the number of samples
$\quad\quad\ x_i$ is a single data point

That leaves estimating $\log f(x_i)$. This estimation begins with the approximation that the probability mass around a point is equal to a sufficiently small volume centered at that point multiplied by a constant probability density throughout the volume. The sufficiently small volume is defined as the hypersphere around the point with a radius equal to the distance to that point's kth nearest neighbor– more formally:

$$P_i(\epsilon) \approx c_d \epsilon^d f(x_i) \tag{8}$$

Where: ϵ is the distance from point i to its kth nearest neighbor
$P_i(\epsilon)$ is the probability mass of the volume around point i to its kth nearest neighbor
c_d is the volume of a hypersphere in d-dimensional space with radius 1

It then follows that:

$$\log P_i(\epsilon) \approx \log c_d + d \log \epsilon + \log f(x_i) \tag{9}$$

Focus is turned to estimating $\log P_i(\epsilon)$. Consider a probability density function $p_k(\epsilon)$ that is the probability density that ϵ is the distance to the kth nearest neighbor of point i. Then $p_k(\epsilon)d\epsilon$ is the probability mass that one point, point i's kth nearest neighbor, is contained in the range $[\epsilon, \epsilon + d\epsilon]$. Then exactly $k - 1$ points are at distances less than the kth nearest neighbor and $N - k - 1$ points, the rest, are at distances farther than the kth nearest neighbor. This is a trinomial distribution where the probability of success, exactly one point in the $d\epsilon$ region is the probability mass from the incremental probability density multiplied by the width $d\epsilon$, the probability of being at a distance less than ϵ is $P_i(\epsilon)$, and the probability of being at a distance greater than ϵ is the remaining probability mass. This takes the mathematical form of equation (10):
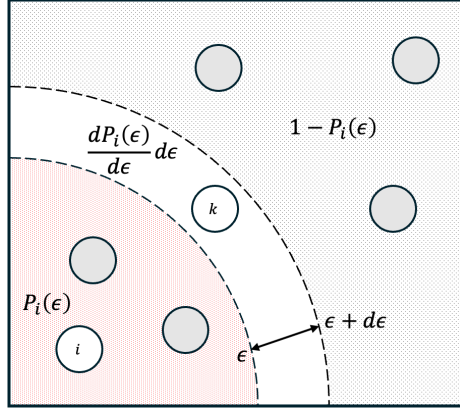

Figure 3: $p_k(\epsilon)d\epsilon$ trinomial distribution

$$p_k(\epsilon)d\epsilon = \frac{(N-1)!}{1!\,(k-1)!\,(N-k-1)!} \cdot \frac{dP_i(\epsilon)}{d\epsilon} d\epsilon \cdot P_i(\epsilon)^{k-1} \cdot \left(1 - P_i(\epsilon)\right)^{N-k-1}$$

$$p_k(\epsilon)d\epsilon = k \binom{N-1}{k} \cdot \frac{dP_i(\epsilon)}{d\epsilon} d\epsilon \cdot P_i(\epsilon)^{k-1} \cdot \left(1 - P_i(\epsilon)\right)^{N-k-1} \tag{10}$$

Then the expectation of the log of the probability mass around a point to its kth nearest neighbor

$$E[\log P_i(\epsilon)] = \int_0^\infty p_k(\epsilon) \log P_i(\epsilon)\, d\epsilon \tag{11}$$

Substituting equation (10) into equation (11):

$$E[\log P_i(\epsilon)] = k \binom{N-1}{k} \int_0^1 \log P \cdot P^{k-1} \cdot (1-P)^{N-k-1} dP = \psi(k) - \psi(N) \tag{12}$$

Where $\psi(\cdot)$ is the digamma function

The details of evaluating the integral in equation (12) are left to Appendix A. Then substituting the right-hand side of equation (12) into equation (9) and rearranging terms:

$$\log f(x_i) \approx \psi(k) - \psi(N) - \log c_d - d \log \epsilon_i \tag{13}$$

And finally averaging over all $\log f(x_i)$

$$\hat{h}(X) = \psi(N) - \psi(k) + \log c_d + \frac{d}{N} \sum_{i=1}^{N} \log \epsilon_i \tag{14}$$

Then to estimate mutual information between a continuous and discrete variable (this combination is chosen for reasons described in the next section)

$$\hat{I}(X;Y) = \hat{h}(Y) - \sum_{x \in \mathcal{X}} p(x) \cdot \hat{h}(Y|X = x) \tag{15}$$

To deal with compounding errors that can occur from the estimation of conditional entropies in the summation term, the kth nearest neighbor for every point in the estimation of the marginal entropy $\hat{h}(Y)$ is selected such that it is the mth nearest neighbor in the associated conditional entropy estimator $\hat{h}(Y|X = x)$. This approach is outlined by Ross[4]. Equation (15) with the Ross selection criterion for nearest neighbors in the marginal and conditional distributions is what is used to estimate mutual information.

2.3. Final Problem Formulation

With the framework for defining adaptability and the machinery to compute mutual information specified, all that's left is to formulate the variables for which mutual information will be computed between.

As described in the beginning of section 2, the mutual information between "playstyle" and lineup will be used as the proxy for adaptability. Defining lineup is trivial – it's a discrete random variable where each suitable lineup that a particular player was a part of is included in the set. Playstyle, for the purposes of this study, is simplified to the coordinates of the player on the court during a particular frame and a Boolean for whether the player has the ball during that frame. This will provide adequate information about how often the player has the ball and where they operate with the ball – a large portion of what defines offensive playstyle. Described later in section 4.3, it could be useful to update the ball possession variable to a discrete random variable sampled from the set of actions $\{None, Dribbling, Passing, Shooting, Rebounding\}$ however this simplified version of ball possession captures plenty of information and can lead to interesting insights on its own.

Mutual information is not well-defined for the mutual information between one variable and another joint random variable that is the joint of discrete and continuous variables, like in the case of the playstyle variable. The mutual information "chain-rule", proven in Appendix B, can separate the effects of mutual information between the components of the playstyle joint random variable. The chain rule is:

$$I(X;Y,Z) = I(X;Y) + I(X;Z \mid Y) \tag{16}$$

This says that the mutual information between X and a joint random variable Y, Z is sum the of the mutual information between X and Y and the incremental information gained about X from Z after already knowing Y.

Substituting for lineup and $playstyle = (Ball\ Possession, Player\ Coordinates)$

$$I(L; B, C) = I(L; B) + I(L; C \mid B) = I(L; B) + \sum_{b \in \{0,1\}} p(b) \cdot I(L; C \mid B = b)$$ (17)

Where: L is the lineup random variable

B is the player-ball possession Boolean random variable

C is the (x, y) player coordinates, random variable

This says that the mutual information between lineup and playstyle, defined as the joint of player coordinates and the ball possession Boolean, is the sum of the mutual information between the lineup and ball possession variables and the mutual information between the lineup and player coordinates, given knowledge of if the player has the ball or not. Given the lineup variable L and ball possession Boolean B are discrete, computing their mutual information $I(L; B)$ is trivial using equation (5). Finally, $I(L; C)$ is computed using equation (15) twice, once for the datapoints where the player has the ball and once for when they don't, with the result weight averaged together.

2.4. Data Preparation

Tracking data is required for the player coordinates and determining the ball possession Boolean. The 2016 SportVU dataset supplemented with play-by-play data is used. The merging of these two data sets is done by Donald Clayton and is publicly available on huggingface[5]. The SportVU dataset contains 631 games from the first half of the 2015-2016 season from the start of the season to January 23rd, 2016.

The structure of the dataset is illustrated in Figure 4. The graph represents a series of "events" in a play-by-play context. Events are point-in-time actions like a shot attempt, a foul, or a turnover. Each game consists of many events and each event has relevant information to describe the event like who the home and away teams are, the players on the teams, what happened during the event, and which player(s) were involved in the event. Finally, the event also contains "moments" which include coordinates for all players on the court and the ball as well as what the clock reads for that moment. Each moment is a frame captured at a sampling rate of 25 frames per second. Given the structure of the dataset, there are several data processing steps outline in sections 2.4.1 and 2.4.2.
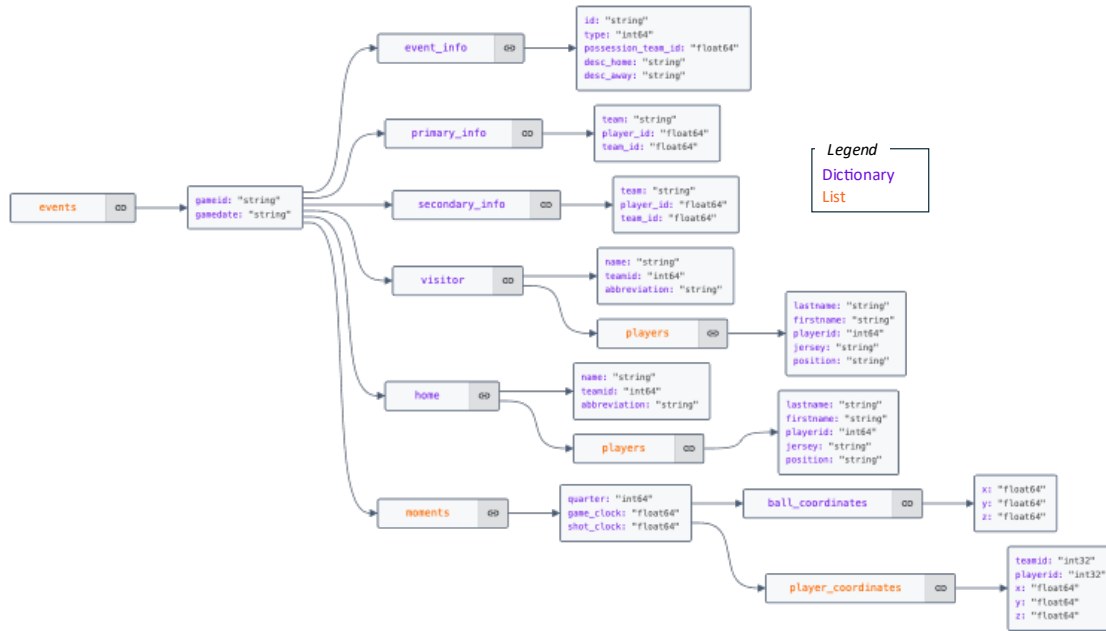
Figure 4: SportVU Dataset Schema

2.4.1. Deduplicating events

The moments, which capture frame-by-frame coordinates of the players and ball, are roughly grouped by possession, however the structure of the dataset maps groups of moments to events. Events are not an entire possession but rather a single moment in time. There may be multiple events in a possession. For example, a possession may consist of a shot attempt, an offensive rebound, and a foul. In this scenario, the shot attempt event, offensive rebound event, and the foul event will all contain the exact same set of moments from the possession that contains those events. Events need to be deduplicated such that each event in the dataset has a unique set of moments. To do so, events for a game are sorted in chronological order. Then starting with the first event, the following event's beginning and ending moments are checked to see if they have the same start and end times as the event that came before it. If so, the event is logged in a set for the possession. The process continues until the next event does not contain the same moments as the current event – it is not in the same possession. Then the set of events in the possession are checked for a made or missed shot event. If no events in the possession contain a shot, the offensive possession is not used for this analysis, and it's thrown out. If a shot event is contained in possession, the event labelled as the made or missed shot is retained to retain information about who had possession and the rest of the events in the possession are discarded. The algorithm is described in algorithm 1.

Algorithm 1 Deduplicate events

1: Initialize clean deduped events list c
2: Initialize Possession set $\mathcal{P}$
3: Sort uncleaned events list e in ascending chronological order
4: **for** each event e_i in $i = 1 \ to \ i = |e| - 1$
5:      **if** $|\mathcal{P}| = 0$
6:          Add e_i to $\mathcal{P}$
7:      **end if**

```
 8:     if start time of e_i = start time of e_{i+1} and end time of e_i = end time of e_{i+1}
 9:         Add e_{i+1} to P
10:     else
11:         if |P| > 1
12:           if shot made event in P
13:             Append shot made event to c
14:           else if shot missed event in P
15:             Append shot missed event to c
16:           end if
17:         else
18:           if e_i is a shot made or missed event
19:               Append e_i to c
20:           end if
21:         end if
22:         Clear Posession set → P = {}
23:     end if
24: end for
```

2.4.2. Removing event overlap

Even though events that contain the exact same set of moments have been deduped, it is still possible for a subset of moments from one event to exist in the next event. This is illustrated in Figure 5. Figure 5 tracks the ball location for two consecutive events during a Lakers vs. Celtics game. The two events are Julius Randle scores (on the right basket) followed by a David Lee missed jumper (on the left basket). In the top two visuals, the red oval highlights the same moments that exists in both events. Removing this overlap is trivial. Since only offensive possessions in the half-court are considered for this analysis, only frames where all 5 offensive players are on the offensive end of the court are retained. This fixes the overlap issue however some plays, like fast breaks, end up being completely removed given not all 5 offensive players are in the offensive halfcourt. This is okay as fastbreaks contain little information about change in offensive playstyle from lineup to lineup. The bottom two visuals in Figure 5 show which moments are removed using these rules.
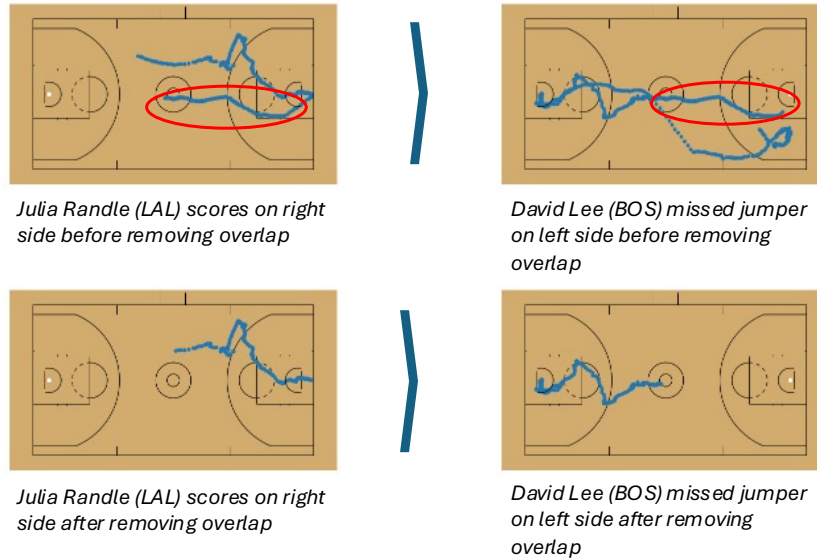
Julia Randle (LAL) scores on right side before removing overlap

David Lee (BOS) missed jumper on left side before removing overlap

Julia Randle (LAL) scores on right side after removing overlap

David Lee (BOS) missed jumper on left side after removing overlap

Figure 5: Location of the ball before and after removing overlap for two consecutive events

2.4.3. Ball Possession

With completely deduplicated, frame-by-frame player coordinate and lineup data, the last step is to compute the ball possession Boolean for the player of interest. A simple heuristic is used. For any frame, if the ball is within a 3-foot radius in the x-y plane and the ball is no higher than 3 feet above the player's z-coordinates, then the player is in possession of the ball. These rules were selected because the average NBA player height is around 6'6" with an average ape index of 1.063 giving an average wingspan of around 83 inches or nearly 7 feet. Players rarely fully extend the ball parallel to the court so 3 feet should capture when a player is in possession of the ball. The ball must also be no higher the three feet above the player's z-coordinate so that shots above a player's head do not count as possession.

3. Results

Equation (17) was computed for the top 8 players by minutes played for each of the 30 NBA teams. The samples used for computation came from the SportVU and play-by-play dataset which was cleaned using the process outlined in section 2.4. Only lineups that had sufficient data after cleaning were used in the adaptability computation, where sufficient is defined as 10,000 frames.

Figure 6 shows a plot of mutual information vs. lineup entropy. Both mutual information and lineup entropy have been shifted and scaled such that the minimum and maximum values are 0 and 100 respectively. The scaling is for increased interpretability. The first item to notice is the relationship between lineup entropy and mutual information; as lineup entropy increases, so does mutual information. This is an expected outcome that is standard to information theory. The more information that a variable has, the more information that can be shared with another variable. From equation (2), lineup distributions that are uniform and have more lineups available will produce higher lineup entropies, allowing for more information to be shared. Translating this phenomenon to a basketball context – players that have more diverse lineups have more opportunity to adapt. Figure 7 shows the mean lineup entropy by number of lineups available to the player.
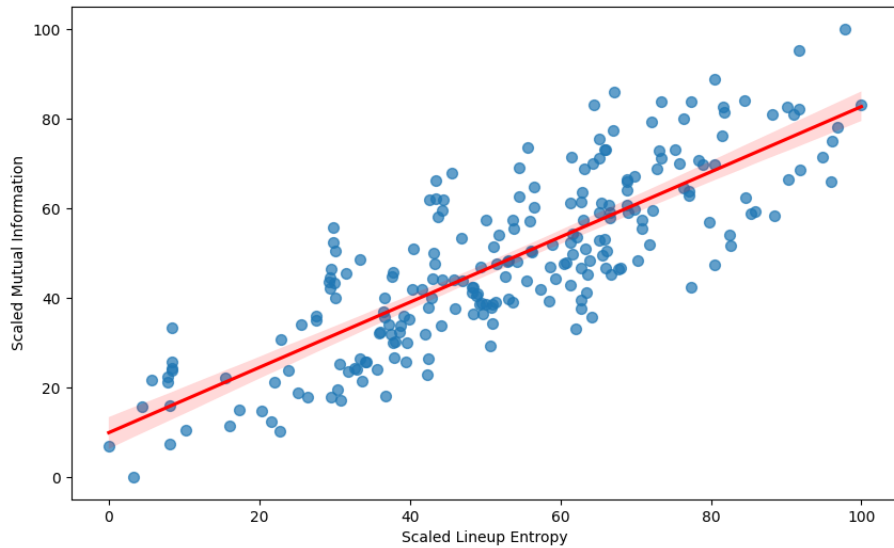
https://github.com/drmbeledogu/Adaptability-Index/

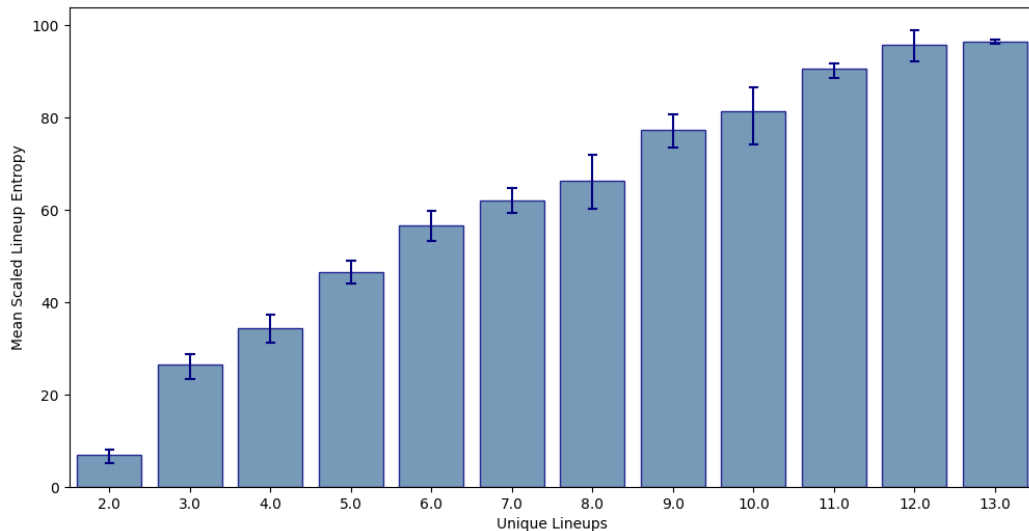*Figure 6: Mutual Information vs Lineup Entropy*

Figure 7: Mean Lineup Entropy by Unique Lineup Count

It seems that there should be some accounting for the number of lineups a player participates in. It might be unfair that a player participates in only 3 valid lineups while another participates in 10. It could be that the player with fewer lineups has a less imaginative coach not willing to try different lineups. It could be that this player is on a team that isn't as deep so there are fewer viable lineups. However, it also seems that players who have a higher mutual information regardless of number of lineups should not be penalized. It could be that these players are on so many lineups because they are so adaptable in the first place. On the low end of the lineup count scale, those players could be there because they are not very adaptable and only work in a couple lineups.

A weighted average of raw mutual information and mutual information above expected is used to capture the effects of being on a certain number of lineups because the player is adaptable and being on a certain number of lineups due to outside factors. Mutual information above expected is the difference between the measured mutual information and the predicted mutual information

given lineup entropy from the regression in Figure 6. Both raw mutual information and mutual information above expected are scaled to a value between 0 and 100. This scaling is done so that mutual information and mutual information above expected have similar magnitude and have equal effects when changing the weight. The final weighted average is scaled as well.

$$Adaptability = scaled[\alpha \cdot scaled[I(L;B,C)] + (1-\alpha) \cdot scaled[I(L;B,C) \ above \ expected]] \quad (18)$$

Where: α is the weighting factor

For this study, an alpha of 0.5 is used as a fair point in the middle although there is no rigorous testing for why it should be selected. It is up to the interpreter to select the weighting.

3.1. Individual Results
Given an alpha of 0.5, Table 1 and Table 2 show the most and least adaptable players in the set.

Table 1: Most Adaptable Players

Name	Team	Position	Adaptability
Dennis Schroeder	ATL	Guard	100.0
Rajon Rondo	SAC	Guard	98.6
Evan Turner	BOS	Guard	97.0
James Harden	HOU	Guard	95.4
LeBron James	CLE	Forward	93.2
Jameer Nelson	DEN	Guard	90.5
Darren Collison	SAC	Guard	87.6
Dion Waiters	OKC	Guard	87.0
Devin Booker	PHX	Guard	85.7
Jerryd Bayless	MIL	Guard	85.0

Table 2: Least Adaptable Players

Name	Team	Position	Adaptability
Ersan Ilyasova	DET	Forward	0.0
Steven Adams	OKC	Center	1.6
Andre Drummond	DET	Center	3.5
Tim Duncan	SAS	Center	6.3
Robin Lopez	NYK	Center	6.9
Hassan Whiteside	MIA	Center	6.9
Marcin Gortat	WAS	Center	8.3
Tony Snell	CHI	Guard	8.9
Ian Mahinmi	IND	Center	9.3
Andre Roberson	OKC	Guard	9.7

Familiar faces show up on the most adaptable players list. LeBron James is known as a player that is certainly versatile and can do it all, however he is also known for elevating the players around him, no matter who they are. Rajon Rondo and James Harden during this time are also known to be floor generals, most likely adapting to players around them. The list of least adaptable players is full of bigs like Steven Adams, Tim Duncan, and Andre Drummond.

The observation about bigs being less adaptable is supported by the entire dataset. Figure 8 shows the adaptability distributions by position. Centers are clumped near the lower end of adaptability

however, there seems to be some centers in the tail that are more adaptable than the bunch. These centers are Jahlil Okafor and Demarcus Cousins who were some of the premier centers during that time and who were asked to do a lot offensively. It now seems they were adapting to the players around them. Forwards and guards have a much more symmetric distribution while guards on the average are most adaptable. This makes sense as guards are typically the most skilled players and are asked to run their team's offense.
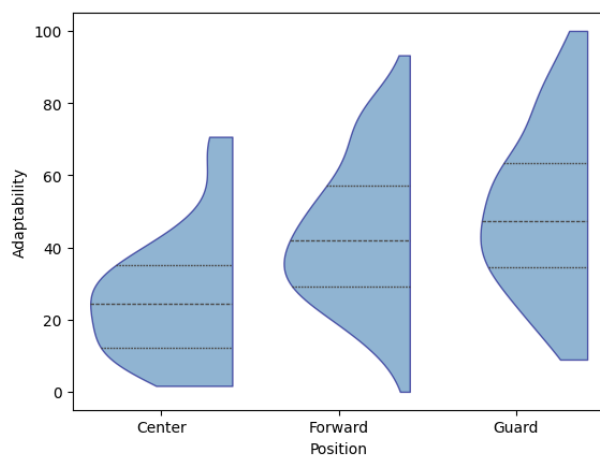

Figure 8: Adaptability Distribution by Position

Mutual information when defining playstyle as the joint of player coordinates and ball possession captures a complex relationship between player coordinates, ball possession, and lineup. Purely knowing the fraction of time that the player has the ball might give some information about which lineup the player is on. Knowledge of where the player is on the court regardless of if they have the ball or not may give some information about which lineup they're on. Then finally, knowledge of where the player is when they have the ball and when they don't have the ball will give some information about what lineup they're on. To illustrate a player being adaptable, I've chosen just where the player is on the court regardless of when they have the ball as a clear visualization.

Figure 9 shows a heatmap of LeBron James' coordinates for different lineups. Generally, LeBron is quite versatile and plays on many areas of the court with some warm spots along the perimeter and near the blocks. In lineup 2 of Figure 9, LeBron plays many areas of the court. Kyrie Irving was injured during part of the 2015-2016 season so the number of frames available for this lineup is not significantly larger than others which implies variation in location is not due to a larger sample size. A hypothesis is that Kyrie is an offensive play maker who can get his own points at any location on the floor. This gives LeBron more freedom to roam the court. It might even be a necessity that he roams the court to allow Kyrie space for when he decides to score. Conversely, lineup 1 contains three other bigs, Tristan Thompson, Timofey Mozgov, and Kevin Love. LeBron is playing a guard on this lineup and clears from the paint by operating along the perimeter so the other bigs can operate where they're comfortable. It's clear that LeBron is adapting to the lineup on the court.
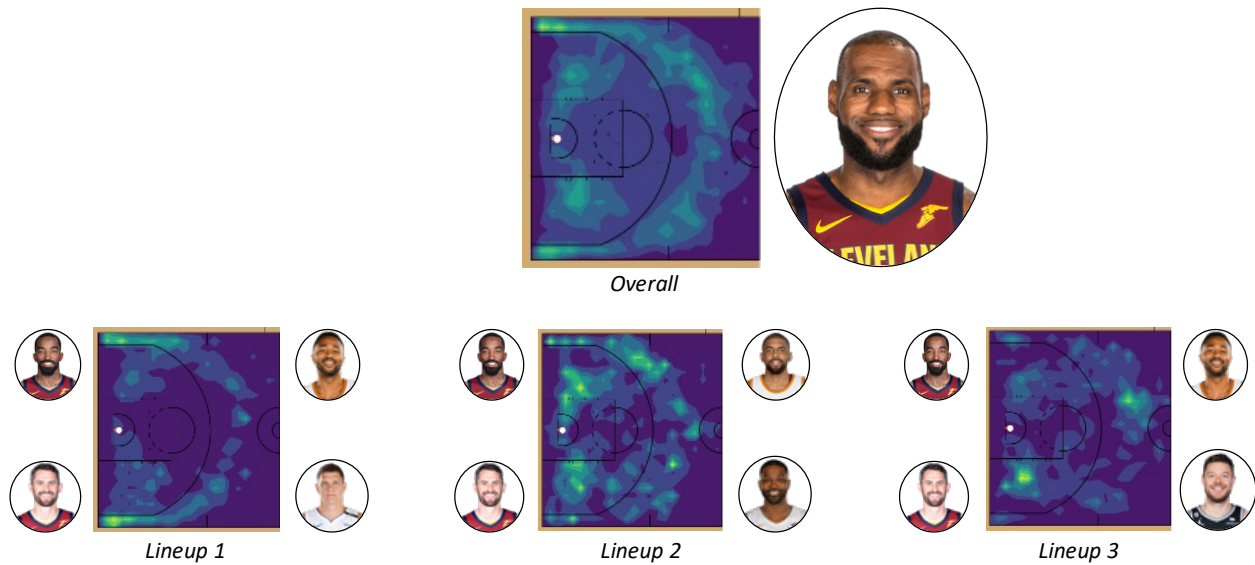
Overall

Lineup 1

Lineup 2

Lineup 3

Figure 9: LeBron James Position Heatmaps

Figure 10 shows a heatmap of Andre Drummonds' coordinates for different lineups. Generally, Andre Drummond operates in the paint. Across the three lineups shown in the figure, Andre operated almost exclusively in the paint with very little variation in position across lineups. Andre is not adapting to his lineup.
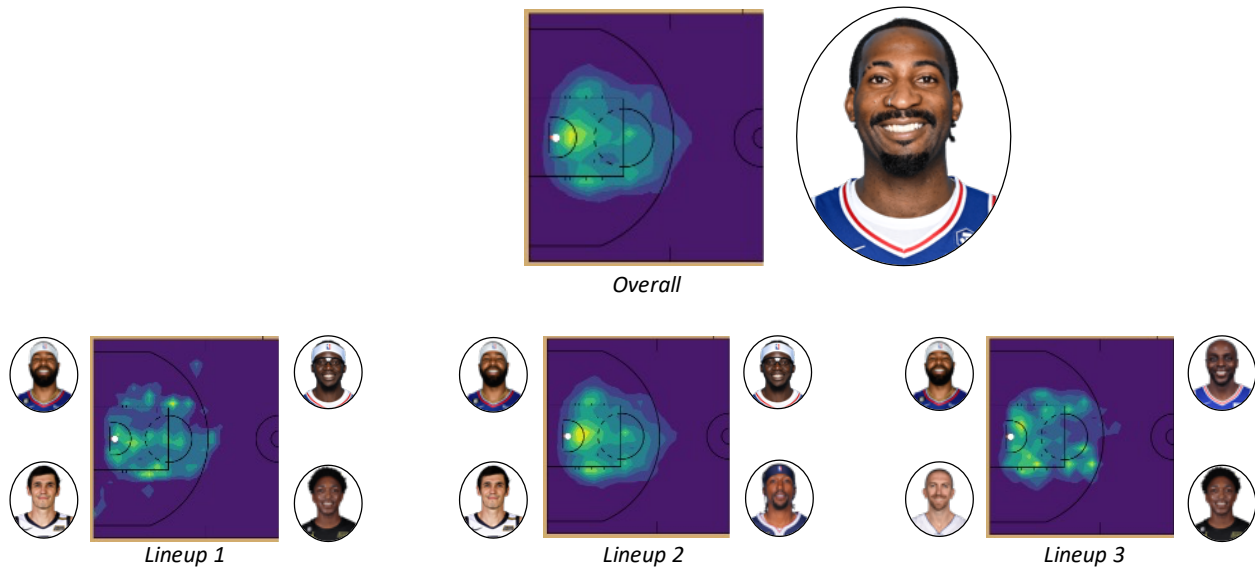

Overall

Lineup 1

Lineup 2

Lineup 3

Figure 10: Andre Drummond Position Heatmaps

3.2. Team Results

Adaptability is most important in the context of your team. The first analysis was to look at the mean adaptability index of the top 8 players by minutes on a team and see if that correlated with regular season wins. Figure 11 shows that there doesn't seem to be much correlation. In fact, the Spurs, one of the best teams in the league that year, on average did not have very adaptable players. This could be interpreted as each player has a role that they play regardless of the lineup that

they're on. The philosophy may be to know your role and excel at it. This is certainly easier on the players and the coach to execute, and it seems the Spurs do this extremely well. The Spurs are not without adaptable players though. Their most adaptable player is Manu Ginobili with an Adaptability Index slightly above average at 55.8. No other player on the team has an adaptability above 50. While most players on the Spurs during this time are tasked with executing their role, Manu is a "glue" person who comes off the bench and must play well with the second unit while also playing large minutes with the starting unit, where he must also perform as well.
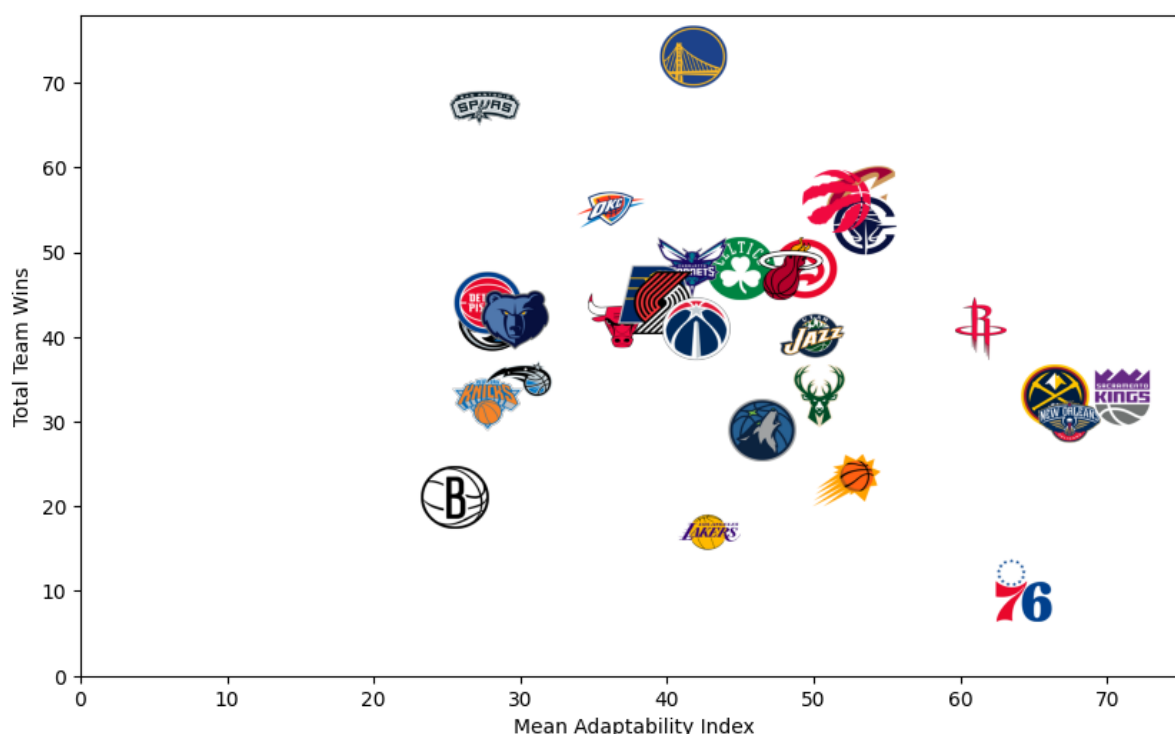

Figure 11: Regular Season Wins vs Mean Adaptability

It is interesting to investigate adaptability in the context of usage rate. Intuitively, it could be good for the highest usage player, the player most responsible for generating offense, to be adaptable for those who may not be as capable at generating offense. Figure 12 takes the usage rank for a particular player on their team and looks to see how high they rank in terms of adaptability on their team. For example, the first bar shows that for players who are the highest usage on their team tend to be the 4th most adaptable player on their team. Although, the error bars are wide (there are only 30 teams and thus 30 data points in each bar), the directional insight is interesting. Figure 12 says that the third highest usage player on the team, not the first, is often the most adaptable. This might be rationalized by the hypothesis that the highest usage player takes on the offensive load and must get points however they feel most comfortable, regardless of the other players on the court with them. However, the third option player must take the points as they come. They must adapt or risk being a lower scoring option on the team.
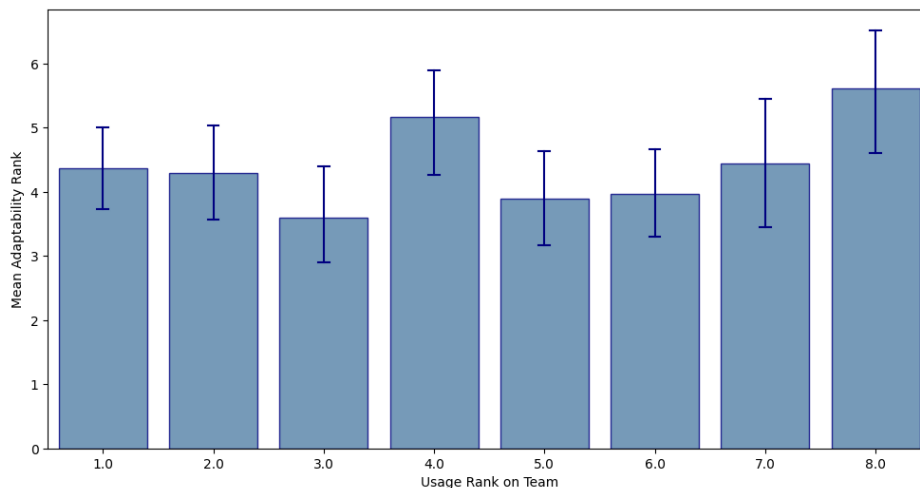
Figure 12: Mean Adaptability Rank by Usage Ranks

The third offensive option on the team might typically be the most adaptable, but is that the best strategy? Figure 13 takes the adaptability rank on the team for the highest usage player on that team and calculates the mean win rate for each bucket. For example, the first column is interpreted as teams for which their highest usage player is also the most adaptable player on the team, averaged 50 wins in 2015-2016. The second bar is read similarly, teams for which the highest usage player is the second most adaptable on the team averaged 40 wins for the season. The error bars are wide once again given only 30 datapoints total but directionally, it seems that it may be beneficial to have your highest usage player be your most adaptable.
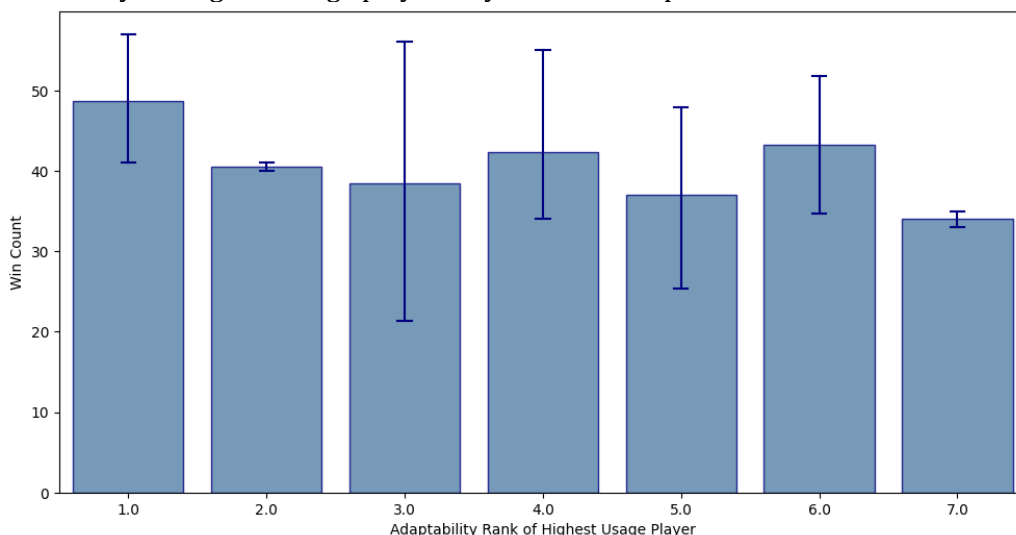

Figure 13: Mean Team Win Count by Adaptability Rank of Highest Usage Player

The teams in the first bar are the Cleveland Cavaliers, Houston Rockets, and Atlanta Hawks. Their highest usage and most adaptable players are LeBron James, James Harden, and Dennis Schroeder respectively. The strategy for these teams seems to be that their highest usage players are acting as floor generals fitting their teammates needs and making those around them better. Note that other high performing teams like the Golden State Warriors and the San Antonio Spurs are not in this list. Their most adaptable players are Andre Iguodala and Manu Ginobili who are the 7th most and 3rd

most used players on their team respectively. They fit the off the bench, glue-person role who provides value by fitting in where they can around stars that excel at their roles like Steph Curry and Klay Thompson for GSW and LaMarcus Aldridge and Kawhi Leonard for SAS. At first glance, it seems that if a team has a sufficient number of very strong offensive talent on the team, then lower usage players must be adaptable to provide value. Conversely, if there is not enough strong offensive talent on the team, the strongest offensive talent must be adaptable to lift their teammates up.

It's interesting, yet expected, that there is not a one size fits all winning solution to bringing in adaptable players. The decision must be made in the context of the existing team, each player's roles, and their excellence at their roles.

4. Discussion

4.1. Caveats

Estimating mutual information using the slightly altered Kozachenko-Leonenko estimator is a strong choice for this application however, it is still an estimate that can have error. Non-parametric estimators, like this one, can be prone to errors if the data are high dimensional ($d \geq 10$). These types of estimators can also have more error when the data are extremely correlated, given the assumption of constant density throughout the volume. For this analysis, the continuous vector was purposely kept to two dimensions to minimize this issue. 10,000 frames were also used to have a sufficient sample size to reduce noise from the measurement. Lastly, the number of neighbors used in the computation at the conditional level (for each lineup) was 3. A higher number will increase the bias in the error and a lower number will increase the variance. 3 is a standard sweet spot.

This metric measures how much a player has adapted in the past. It does not measure how much a player is able to adapt in the future. Possibly, in future studies, an analysis could be done to determine if the adaptability index has some autocorrelation with future values. The metric also does not measure how good a player is while adapting. A player can be adaptable but be poor at all the playstyles. This metric, like most metrics, must be used in totality with other metrics and the eye test to get the fullest insight.

4.2. Use cases for the metric

At the highest level, the main use case for this metric is to understand the effect of roster changes like bringing in a new player or the effect of bringing together several players who have not played with each other before. It could also be used to measure the effect of a player leaving – depending on the roster composition, players who are more adaptable may leave more of an offensive hole, all else being equal in terms of offensive capability.

At the lower level, there seems to be two main use cases for this metric. The first is roster design for team front offices. This metric would be used to select which available players would have the highest impact on a new team, both in terms of lifting the existing players' capabilities but also for the new player to be brought in. The second use case is for fantasy sports and sports betting where prediction of team/player performances after drafts, trades, and new signings is important. This metric could be used to help predict team/player performances before there is adequate data about how the team/player is performing in the new scenario. This early period is where sports bettors may find an edge and win money from the house.

https://github.com/drmbeledogu/Adaptability-Index/

4.3. Next Steps

The next steps are to improve the measurement of the metric and to continue to test for its usefulness in insight generation and future outcome prediction.

The definition of the playstyle vector has room for improvement. As defined in this paper, it accounts for where the player is on the court and where they have the ball. While where the player is on the court and whether they have the ball encodes information about what the player is doing, it is not a full replacement for explicit categorization of what the player is doing. In future iterations of the metric, playstyle would be defined as a joint random variable (C, A) where C is a two-dimensional vector containing the (x, y) coordinates of the player, like it is defined in this paper, and A is a discrete action vector extending the current B ball possession Boolean definition. The values of A would be $\{None, Dribbling, Passing, Shooting, Rebounding\}$. A machine learning model would be trained and implemented in the pipeline to determine if the player does not have the ball or if they are dribbling, in the motion of passing, in the motion of shooting, or in the motion of rebounding. This new joint random variable would encode information that would nearly fully encompass playstyle. The downstream computation would all be the same.

Unfortunately, there isn't much publicly available NBA tracking data. While the metric has some interesting insights with half a season's worth of data, gaining more confidence in those insights requires many seasons worth of data. With many seasons' worth of data, auto correlation or metric stability studies could be run. The effect on team performance could properly be studied with regression techniques that include proper controls. The effect of moving players from one team to another, by nature, requires several seasons worth of data. All these studies can hopefully be done in the future upon the public release of more tracking data.

5. Conclusion

In conclusion, mutual information is a powerful tool for measuring the adaptability of a player. While the Adaptability Index shows promise, the metric has room for improvement. Future iterations coupled with the public availability of more tracking data could generate even more interesting insights for roster development.

https://github.com/drmbeledogu/Adaptability-Index/

References

[1] "Versatility Index in NBA Explained," www.nbastuffer.com, May 08, 2017.
https://www.nbastuffer.com/analytics101/versatility-index/

[2] "NBA Stats & Analytics | CraftedNBA," craftednba.com. https://craftednba.com/glossary

[3] L.F. Kozachenko, N.N. Leonenko, "A statistical estimate for the entropy of a random
vector," Problemy Peredachi Informatsii 23 (1987), 9–16.

[4] B. C. Ross, "Mutual Information between Discrete and Continuous Data Sets," PLoS ONE,
vol. 9, no. 2, p. e87357, Feb. 2014, doi: https://doi.org/10.1371/journal.pone.0087357.

[5] "dcayton/nba_tracking_data_15_16 · Datasets at Hugging Face," huggingface.co.
https://huggingface.co/datasets/dcayton/nba_tracking_data_15_16

Appendix A: Kozachenko-Leonenko Integral Evaluation

The goal is to simplify the integral in equation (12) also listed below:

$$k \binom{N-1}{k} \int_0^1 \log P \cdot P^{k-1} \cdot (1-P)^{N-k-1} dP$$

Simplifying this integral requires knowledge of the Beta Function which is defined as:

$$B(m,n) = \int_0^1 t^{m-1} \cdot (1-t)^{n-1} dt$$

And also equals

$$B(m,n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}$$

Where $\Gamma(\cdot)$ is the gamma function. Then taking the derivative of the Beta function

$$\frac{\partial B(m,n)}{\partial m} = \Gamma(n) \frac{\partial}{\partial m}\left(\frac{\Gamma(m)}{\Gamma(m+n)}\right) = \Gamma(n)\left(\frac{\Gamma(m+n)\frac{\partial \Gamma(m)}{\partial m} - \Gamma(m)\frac{\partial}{\partial m}(\Gamma(m+n))}{\Gamma(m+n)^2}\right)$$

or

$$\frac{\partial B(m,n)}{\partial m} = \int_0^1 t^{m-1} \cdot (1-t)^{n-1} \log t \, dt$$

Then using the definition of the digamma function:

$$\psi(x) = \frac{d}{dx}\ln\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$$

And rearranging

$$\Gamma'(x) = \psi(x)\Gamma(x)$$

Then

$$\frac{\partial}{\partial m}\Gamma(m+n) = \psi(m+n)\Gamma(m+n)$$

Substituting into the partial derivative of the Beta function

$$\frac{\partial B(m,n)}{\partial m} = \Gamma(n)\left(\frac{\Gamma(m+n)\Gamma(m)\psi(m) - \Gamma(m)\Gamma(m+n)\psi(m+n)}{\Gamma(m+n)^2}\right)$$

https://github.com/drmbeledogu/Adaptability-Index/

And simplifying

$$\frac{\partial B(m,n)}{\partial m} = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}\left(\psi(m) - \psi(m+n)\right)$$

Then using the two definitions for $\frac{\partial B(m,n)}{\partial m}$ and equation one another

$$\int_0^1 t^{m-1} \cdot (1-t)^{n-1} \log t \, dt = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}\left(\psi(m) - \psi(m+n)\right)$$

Plugging P for t, k for m, and $N - k$ for n and simplifying

$$k\binom{N-1}{k} \int_0^1 \log P \cdot P^{k-1} \cdot (1-P)^{N-k-1} dP = \psi(k) - \psi(N)$$

Appendix B: Mutual Information Chain Rule

The chain rule stated in equation (16) is below:

$$I(X;Y,Z) = I(X;Y) + I(X;Z \mid Y)$$

To derive the mutual information chain rule, mutual information is converted to its definition of reduction in the entropy of the joint variable Y, Z after knowledge of X

$$I(X;Y,Z) = H(Y,Z) - H(Y,Z \mid X)$$

The next step is to rewrite the entropy of the joint variable Y, Z as the combination of singular and conditional entropies. To do this, the chain rule for entropy is used. The entropy chain rule is:

$$H(X,Y) = H(X) + H(Y|X)$$

This chain rule for entropy reads that the entropy of two variables is the entropy of one variable plus the entropy of the second variable given knowledge of the first. For completeness, the entropy chain rule will be derived. The proof for the entropy chain rule starts with converting entropy to the probability mass definition.

$$H(X,Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log(p(x,y))$$

Then the probability mass function for the joint random variable is split in the logarithm.

$$H(X,Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log(p(y|x)p(x))$$

Then using logarithm rules, the logarithm of a multiplication is split into the sum of logarithms.

$$H(X,Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y)[\log p(y|x) + \log p(x)]$$

Then the summations are split into two terms. In the first term, the conditional probability mass function is replaced by the definition where the joint distribution is divided by the marginal. In the right-hand term, the probability mass for the joint is split.

$$H(X,Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log\left(\frac{p(x,y)}{p(x)}\right) - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(y|x)p(x) \log p(x)$$

There is one last step for the term on the right-hand side. The terms that have no dependence on y is moved to the outer most summation. Then what's left on the inner summation summed over $\mathcal{Y}$ is just 1.

$$H(X,Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log(p(y|x)) - \sum_{x \in \mathcal{X}} p(x) \log p(x) \sum_{y \in \mathcal{Y}} p(y|x)$$

Then the term in the left-hand side is the definition for $H(Y|X)$ and the term on the right-hand side is $H(X)$.

$$H(X,Y) = H(Y|X) + H(X)$$

Using the entropy chain rule to reduce the entropy definition of mutual information to include only single variables, the following equation is achieved.

$$I(X;Y,Z) = H(Y) + H(Z|Y) - [H(Y|X) + H(Z \mid X,Y)]$$

Then rearranging terms.

$$I(X;Y,Z) = [H(Y) - H(Y|X)] + [H(Z|Y) - (Z \mid X,Y)]$$

And finally, the chain rule for mutual information.

$$I(X;Y,Z) = I(X;Y) + I(X;Z \mid Y)$$