Eindhoven University of Technology

MASTER

Visual analytics for basketball shot quality analysis

Wang, Shixiong

Award date:
2021

Department of Mathematics and Computer Science
Visualization Research Group

Visual analytics for basketball shot quality analysis

Shixiong Wang

under supervision of professors:
dr.Vilanova Bartroli
dr.Huub van de Wetering
dr.Kees Huizing

Eindhoven, June 2021

Abstract

Data analytic is widely used in modern competitive sports. Visualization techniques can be one of the intuitive ways to make sense of the data to sports participants. In this thesis, we design a visual analytic system to analyze basketball game information, especially for coaches. The users can explore the data and dig for the information helping to understand and teach the game better or prepare the game against their opponents. A primary focus of our visual analytic system is the shot quality. Shot quality has always been the core of offense in team-played ball games like basketball and football. However, what is shot quality is not well defined. We develop a method to evaluate the shot quality of teams, lineups and players. Coaches and players are provided an interactively visual analytic framework to explore the NBA basketball games. This framework can help the users learn the context of each shot and estimate the shot percentage under a specific context. Allows the users to evaluate the most essential, thus poorly defined factor in several competitive sports: shot quality.

Contents

Chapter 1

Introduction

1.1 Introduction

Basketball is one of the most popular team sports across the world. Numbers of people from Europe, Asia, and America, enjoy basketball in their daily lives. Data statistics and analysis are vital from the very beginning of this sport. Many professional scouts and analysts are paid to work with the numbers in the statistics of every game to evaluate the players' performance, adjust the teams based on the conclusions. However, it is challenging to present numbers in rows and columns to coaches or professional players then convince them with the conclusions, even more difficult if not impossible when the conclusions are against their basketball knowledge.

The most influential league in the basketball world is the NBA (National basketball association). Not only do they have the most talented players in the world, but they also have top technique devices in the area of capturing sports data. In this paper, we utilize NBA data and build a visualization tool for the users who want to dig into the basketball games. We use the data captured and provided by SportsVU, including every position of players and the ball in a game.

One of the meaningful questions one would like to find out from basketball data is: Is your team a winning team? To answer this question, we should look at the score table and look at your players shooting the shot that most likely will go in. Are the decisions of taking those shots good for helping your team winning? Or, it is a better idea to hold the ball for a while and not take the shot. This is what a basketball coach calls"shot quality". The higher the shot quality (or simply, the better the shot) is, the shot gives more chance to win the basketball game. Due to the complexity of a basketball shot, it is hard to maintain a 50% score at long range, even for the best shooters in the world. The performance of a player or a team should not be strict around the score itself. What we want is to reduce the luck factor and express the essential part of the game. A term commonly called "shot quality" in the basketball world, while every coach has a different definition of this term. This shows the difficulty in quantifying this term. However, suppose we have a way to evaluate the quality of shots. In that case, we can obtain information such as "this is a shot that will probably succeed", or a certain five players that are currently playing in the game what is so-called a "lineup", knowing whether a lineup performs better, e.g. because they are able to get better shots. In this work, we propose a solution for evaluating the shot quality, which depends on the distance to the basket and the defenders.

The shot quality is such a critical aspect of depicting the game of basketball. It is one of the most critical factors when we evaluate a team's offence. However, shot quality remains subjective and cannot be automated. Furthermore, shot quality is not an easy subject to communicate with others. Coaches usually use the filmed matches to address the shot quality problems with players; it takes time to filter the film and locate the moment they want to talk about. More often than not, especially in a limited time situation, during the halftime break, for example, they would trust their memories and speak to players something like: ”You should not shoot the ball in the middle of the first quarter, that was some bad shots.” This kind of communication is not efficient. We seek an alternative that explicitly demonstrates the players’ shot, why is it harmful and how bad it is. This leads to a crucial requirement of a visual analytic tool that allows coaches and players to filter and recognize shots, find problems, address them, and give the players direct insight into their decisions.

This thesis presents a data-based shot quality measurement and a visual analytic system, namely the Shot quality visual analytic(SQVA) system, allowing the exploration of the information in different facets: by team, lineup and individual. And also allow the user to observe the shot quality in both offence and defence point of view.

The rest of this thesis contains the following sections. In section 2, we present a few studies in the competition sports area. In section 3, we explain the data set we utilized in this project. We take a season-long of NBA player tracking data captured by SportVU. We briefly discuss how this data is captured. In addition, we also show how to enhance this data by using other data set so that it can be easily analyzed. In section 4, we analyze our tasks. We listed the external factors that can be treated as input of shot quality. We formulate the questions that the users can use our tools to answer. From that, we derive the requirements of the SQVA system. In section 5, we give an overview of our SQVA system. We illustrate every component in the system and its roles. Followed by section 6, we explain the detail of mining the data and detect the shooter and defenders in the shots. In section 7, we discuss the visual design of the SQVA system. In section 8, we explain how this design is realized, the software and libraries we learn and use in this project. In section 9, last but not least, we present an evaluation plan and discuss the result that comes out of the execution of the plan.

Chapter 2

Related work

2.1 Related works

2.1.1 Basketball shot analysis

The shot selection problem, or in general winning a basketball game problem, has been formulated in different ways. Skinner[19] introduces three factors that influence whether a shot should be taken. Namely (a) The probability that this shot will go in. (b) The distribution of shot quality that the offense is likely to generate in the future. (c) the number of shot opportunities that the offense will have before it is forced to surrender the ball to the opposing team. Moreover, classify this problem as an "optimal stopping problem[13]," including the "secretary problem[14]." We focus on describing the shot quality Q as just the probability of scoring the shot. The most hot topics in shot quality is the hot hand effect[8][22][9]. The hot hand effect is when a player makes a few shots in a row(he is hot), then he is more likely to make the next one. This effect is popular in video games. Unfortunately, Thomas Gilovich et al. [8] Robert L.Wardrop et al. [22], and Jonathan J Kochler et al. [9] do not support this theory, both in a real game and three-point contest. There are no defenders in a three-point contest. Therefore no defense adjustment can be made. In this thesis, we make the assumptions accordingly. When we analyze a shot, we do not consider the events that happen before or after it. The allocation of the three and two points shots is covered in studies[6][21][11]. However, a quantification of the shot quality is not given.

Some of the external factors of shot quality are studied. In 1993 Satern shows the change of kinematic parameters influenced by the distance factor.[17] Okubo, on the other hand, gave a complete study and efficient visualizations of shot defense in 2012.[12]. D. Daly-Grafstein and L. Bornn shows a way of predicting the shot percentage in their work. [5] They are using the players' past shot percentage to predict the future. They treated every shot as a Bernoulli random variable and did not take the external factors into account.

2.1.2 Basketball shot visualization

Basketball shots are often visualized and analyzed in a shot chart like in Figure 2.1. Reich, Brian J et al. develops hierarchical spatial models for shot-chart data [16]. Nowadays, data science plays a significant role again in all fields. Competitive sports are no exception. Similar basketball visual tools like SQVA are developed.[4][10] However, they focused on combining multiple features in a general way.

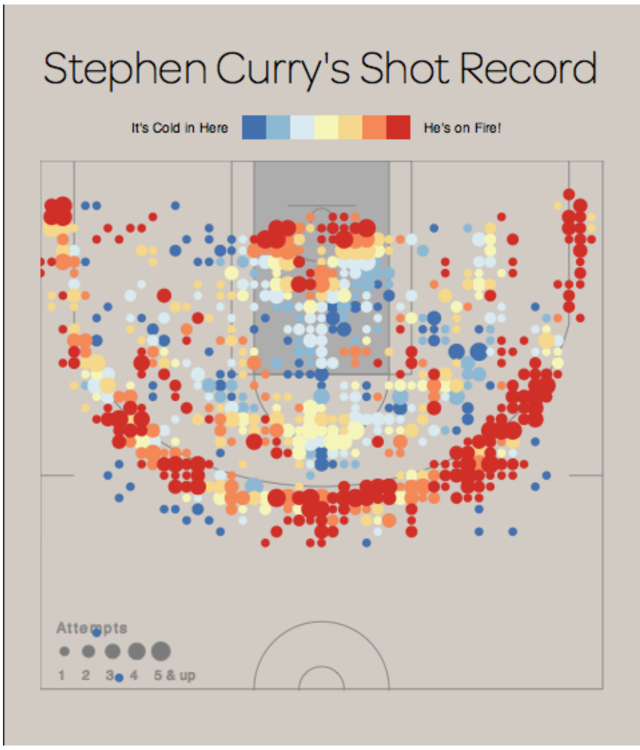


Figure 2.1: A shot chart of Stephen Curry, the size of the maker indicating the quantity of the shot from a particular position, and the color means the percentage of making those shots

Chapter 3

Data analysis

3.1 Data analysis

3.1.1 SportVU data

The data we are using in this project is provided by SportVU optical tracking system. This system is composed of six cameras installed in every 30 NBA teams' home court arenas, which can capture spatial-temporal data in real-time. In Figure 3.1 we showed such a camera in the Barclays center, the home court of Brooklyn Nets.

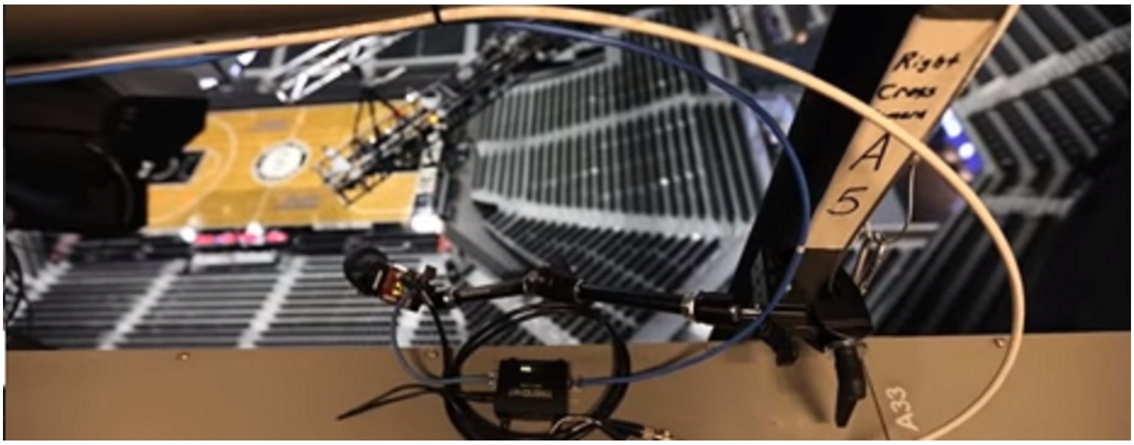


Figure 3.1: SportVU camera[2]

As Figure 3.2 illustrated, six pictures are taken at the same time. Thanks to the optical technologies, a frame of a basketball court with all players positions and ball position are calculated out of them.

This data can be accessed via APIs from the NBA static site [1] in the form of JSON. Every second 25 samples are taken, each containing coordinates of the ball and ten players, each of these samples is a moment. These moments are folded into multiple periods, and we call this period an episode. Every episode is a particular possession played by the players. There are 400-500 episodes captured per game. We have unique keys for games, episodes and moments.

In this project, we use data about 600 games in the 2015-2016 season. Additional information such as the players' names, shot clock, and game clock are also provided in this data. For more clarity, we can

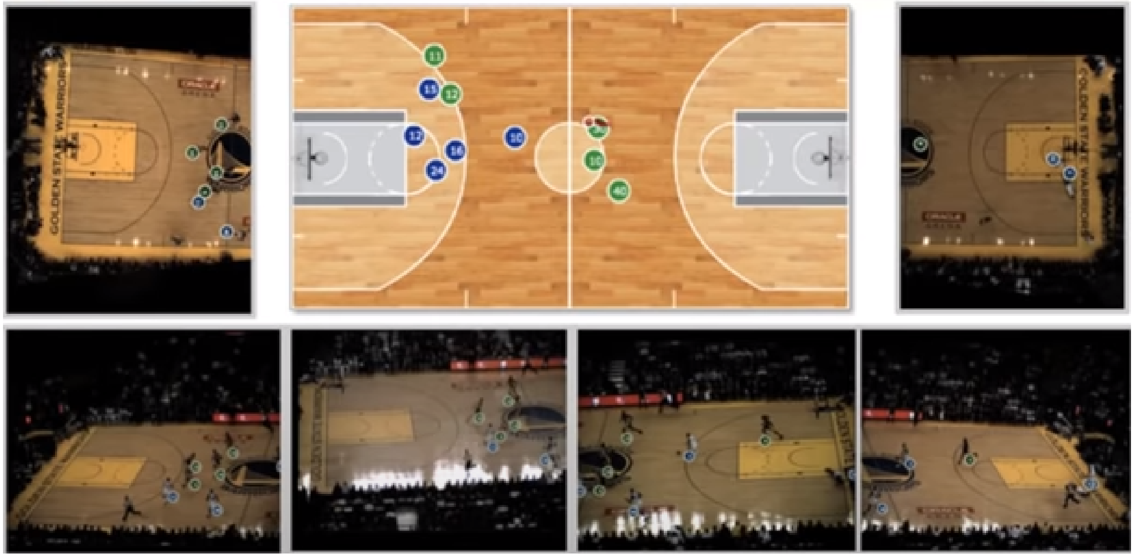


Figure 3.2: A SportVU sample, containing players' and ball's coodinate calculated from 6 pircures taken by the cameras

define the data set as a more mathematical format:
T is the set of all 30 registered NBA teams.
P are the set of all NBA players
$O = P \cup \{Basketball\}$ is the set of player union with the basketball
G is the set of all games played.
E is the set of the episode, which is a period of time from a special event till the other in a game.
$M = \{1, 2, 3, 4\} \times [0, 24] \times [0, 720] \times (O \times X \times Y \times Z)^{11}$ where $X \subset [0, 100]$,$Y \subset [0, 50]$,$Z \subset \mathbb{R}^{+}$ is the set of moment, which is associated with quarter number,shot clock, game clock and positions of all players and position of basketball on the court.
$PT \subset P \times T$ is a binary relation indicating the assignment of player to team.
$EG \subset E \times G$ is a binary relation indicating the episode belongs to which game.
$ME \subset M \times E$ is a binary relation indicating the moment belongs to which episode.
$MT \subset M \times T \times T$ is a relation indicating which two teams played the game.
$ET \subset E \times T$ is a relation indicating which team is the episode about.

Having the player tracking data alone is yet not sufficient for our analysis. We cannot tell if a player is shooting the basketball since the player data is 2-dimensional. All the z coordinates are recorded as 0 except for the ball. Moreover, we have only detected position data while a basketball move involves multiple movements of joints and muscles of a human body. We have to find extra information. Rajiv Shah's Projects Blog [18] suggests using NBA play by play data to provide more information to the player tracking data. We make use of the play by play data in our data analytic. We discuss this in detail in the following subsection.

3.1.2 Play-by-play data

To further analyze the player-tracking data, we can merge the play by play data with the player-tracking data. NBA provides play by play data since 1996-1997; each row contains 40 attributes, including the game ids and episode ids. In Figure 3.3 we showed a snapshot of the play by play data. By the fact that the play by play data share the game and episode ids with the player tracking data, we can merge these two datasets on the shared ids.

As the Figure illustrates, some of the attributes are duplicated since the player tracking data already has the players' information. The most relevant attribute in this data set is the type number. The type


Detail Compact Column 40 of 40 colum

URL	GameType	Location	Date	Time	WinningTe...	Quarter	SecLeft	AwayTeam	AwayPlay	AwayScore
/boxscores/2015 10270ATL.html	regular	Philips Arena Atlanta Georgia	October 27 2015	8:00 PM	DET	1	720	DET	Jump ball: A. Drummond vs. A. Horford (E. lyasova gains possession)	0
/boxscores/2015 10270ATL.html	regular	Philips Arena Atlanta Georgia	October 27 2015	8:00 PM	DET	1	701	DET	A. Drummond misses 2-pt layup from 1 ft (block by A. Horford)	0
/boxscores/2015 10270ATL.html	regular	Philips Arena Atlanta Georgia	October 27 2015	8:00 PM	DET	1	699	DET		0
/boxscores/2015 10270ATL.html	regular	Philips Arena Atlanta Georgia	October 27 2015	8:00 PM	DET	1	697	DET		0
/boxscores/2015 10270ATL.html	regular	Philips Arena Atlanta Georgia	October 27 2015	8:00 PM	DET	1	681	DET	M. Morris makes 2-pt jump shot from 13 ft (assist by A. Drummond)	2
/boxscores/2015 10270ATL.html	regular	Philips Arena Atlanta Georgia	October 27 2015	8:00 PM	DET	1	660	DET		2
/boxscores/2015	regular	Philips Arena	October 27 2015	8:00 PM	DET	1	644	DET	K. Caldwell-	4

Figure 3.3: A snap shot of play by play data

#1	Make
#2	Miss
#3	Free Throw
#4	Rebound
#5	out of bounds / Turnover / Steal
#6	Personal Foul
#7	Violation
#8	Substitution
#9	Timeout
#10	Jumpball

Table 3.1: Action type in play by play

number indicates which type of action happens in this particular episode. It can be a shot or turnover or anything else that happens in a basketball game. In Table 3.1 we list part of the types available in the data set. Adding this to the data model, we have an action type associated with the episode. We also have the involved player's ids in each episode used to calculate the line-ups. We merged the play by play data with our player tracking data, and we obtained an extended data model as illustrated in Figure 3.4.

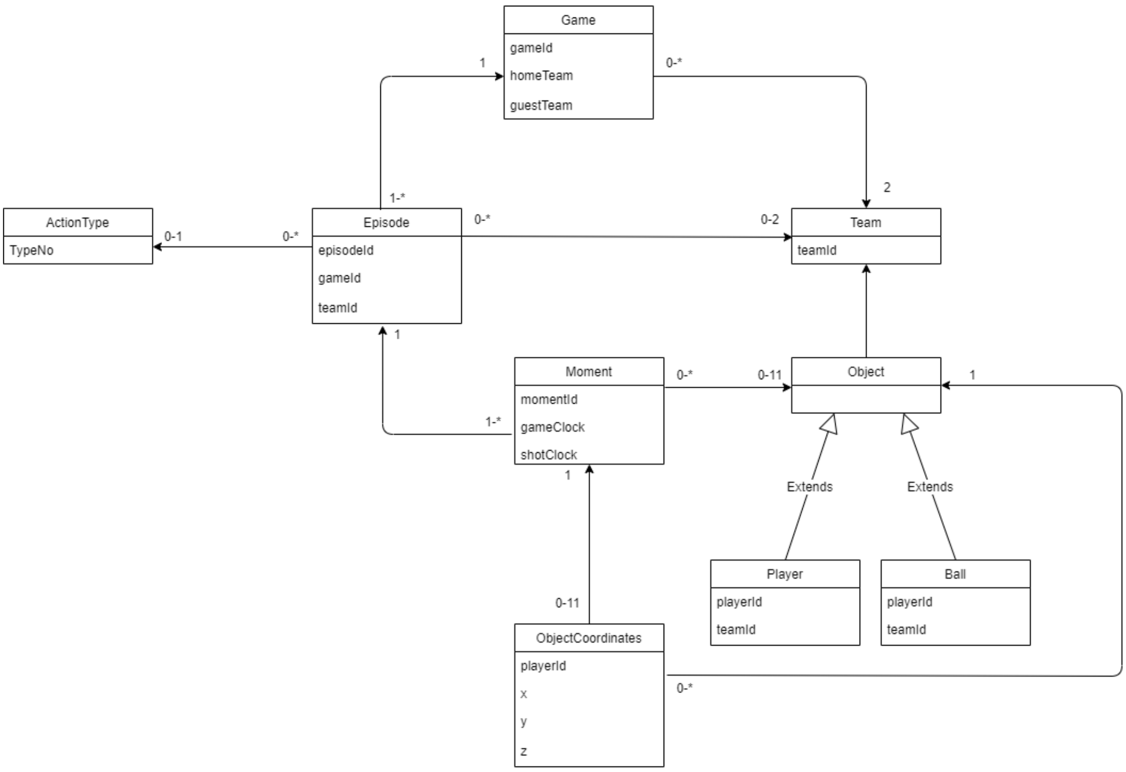


Figure 3.4: Data model diagram

Chapter 4

Tasks analysis

4.1 Tasks Analysis

Although shot quality is one of the most common terms used in basketball training and TV programs, one always struggles to explain what the shot quality really is. Most of the shot visualization that can be found are result-oriented. A scatter chart similar to Figure 2.1 can be found on a variety of NBA sites. It nicely presents both the shot preference and shot percentage of a player. Nevertheless, in this kind of visualization, we discover nothing about how the shot is being taken. How is the shooter being defended? Which is the chess game played by coaches behind the scenes. From the team's point of view, getting more good shots is equally essential as scoring more shots. Getting a good quality shot at the end of a play can be seen as a consequence of the successful execution of good strategies. It can be good execution of the offense strategies or some exceptional performance by creative players. Having more good shots than the opponent is often the aim of coaches offensively because once it comes down to the shooting itself, it is all about the players shooting mechanically. This is not an examination of the coach's strategic level and the team's cooperation level. On the other hand, if an individual player's shot quality is high, it might be the consequence of his careful choice of the shot. It could also be that his shooting ability is so high that every shot seems a high quality shot for him. Our task is to present this information in a different context to the user to find the patterns.

4.1.1 Domain tasks and user questions

Tasks

As we stated previously, shot quality remains subjective. Our task is to build a VA tool that helps users to evaluate the shot quality. Given any shot considering the shooter's position and defender's position. Our tool will provide a visual analytic based on the data, allowing the user to evaluate if the shot strategy implemented is what he wants. The users should get a clear impression of the context of each shot by interacting with the interface.

As an observer of the basketball season, the user wants to ask which team has a better shot quality overall. The user also would be interested in the connection between shot quality and the rank in the league.

A major part of the coaches' job is to find out how to make combinations out of his players, and distribute the playing time to those 5-player-combinations, what we called lineups, so that they can perform at their highest efficiency. Note that combining the players is not always to find the best five players. Some players are more fit playing with ones but not so much with others. Our tool should be able to help coaches with this task. Our system should be able to let the users compare the shot quality between the different lineups.

Basketball is played on two sides of the field. On the other side of the floor, the coaches consider exactly the opposite way. The user should be able to perform the tasks of evaluating the performance of teams/lineups/players on the defensive end. They should be able to see the shot quality estimation of the team's/lineup's/player's opponent. This will help the user construct a much-balanced team to dominate both ends of the field in their games.

To conclude, the tasks are listed as:

1. View defence and offence perspective.
2. Analyze a team, and a lineup average shot quality.
3. Comparisons between players and lineups.
4. Shot quality probability analysis.

User questions

To perform the tasks we mentioned, we emphasis some questions:

1. Team level
   (a) Which team has better shot quality generated during the season?
   (b) How much the game result depends on the shot quality of themselves and the shot quality of their opponents?
   (c) How the team shot quality relate to the players and lineups shot qualities?
2. Lineup level
   (a) Which lineups are played for each team, and how is the playtime distributed?
   (b) How good are the shots generated by each lineup?
   (c) How good are the shot qualities for the opponents when the lineup is playing. (are they giving easy shots or force difficult ones)
   (d) How are the coaches distributing the playing time within their players, and would there be a potential improvement?
3. Player level
   (a) Which player has more shots within a certain lineup?
   (b) What is the distribution of the players' shots and their success?
   (c) How player's shot distribution affect their team?
   (d) What shots are the known elite players specialized in?

Apart from these questions, users should be able to generate their own research with the SQVA system. The possible usages should not be limited to these tasks.

4.1.2 Requirements

Furthermore, in order to provide an effective experience to the users while exploring, there are some requirements we should meet

- Views in the system should be connected and accessible by interactions.
- The system should be interactive with no high latency.
- The system should be designed to be viewed in a web browser with 1920×1280 resolution.

Chapter 5

Shot quality visual analytics system

5.1 Shot Quality Visual Analytics System

5.1.1 Overview

Our SQVA system contains three main components. Including pre-processing, data model and the Views in the web server. The structure is illustrated in Figure 5.1.

The data processing component is responsible for processing the data set into such a form so we can use it as input of our algorithms. It reads both the play by play data and the player tracking data. Then two processes are run over the merged data set—namely, the episode analysis and the lineup analysis. The episode analysis component goes through all episodes, filters out the plays without ending up with shooting, and analyses the player tracking data to determine the necessary parameters for evaluating the shot quality. Records the shot data along with the player data as mediation data. Meanwhile, the lineup analysis process will add up the playing time of lineups for each team. Also, records the data for further usage.

The shot quality estimation unit is prepared for all views we create. The player VA lineup VA and Team VA acquire data once the user is interacting with it, render the visual analytic that we provide to the user.

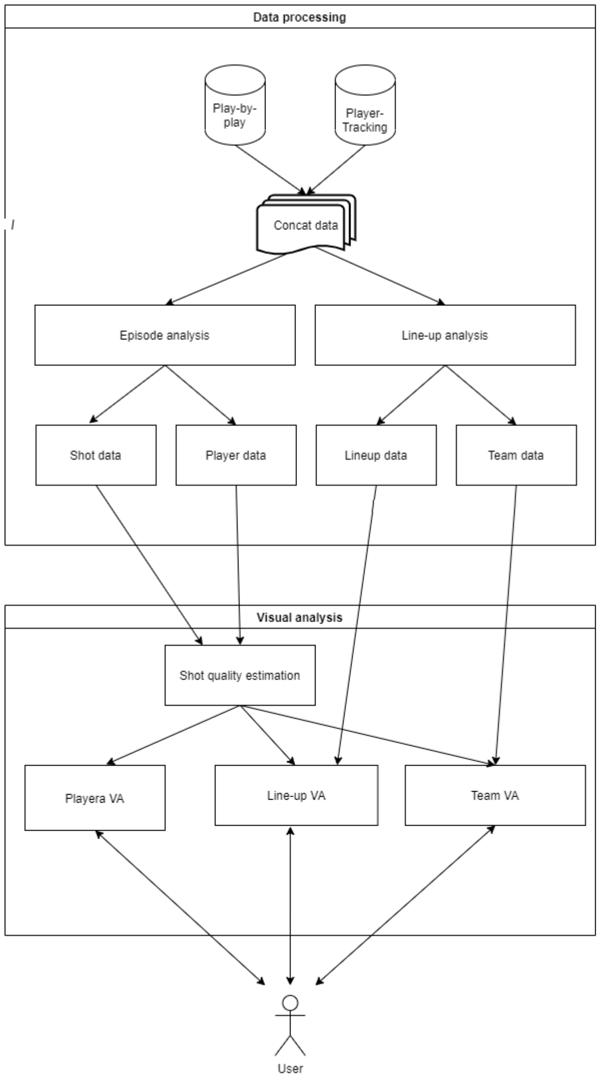


Figure 5.1: An overview of the SQVA system.The arrows denote the data flow.

Chapter 6

Shot quality estimation

6.1 Shot quality estimation

This section describes the approach we use to estimate the shot quality, which is central to our development. We implement this in the episode analysis process in Figure 5.1. To evaluate the quality of a shot, we know that many external factors are relevant. One of the greatest sports authors and basketball coaches, Alan Stein Jr[20] has listed the factors he thinks that can determine whether a player should shoot a basketball in a certain situation:

1. A good shot is one that is expected by your teammates.
2. A good shot is one that you are ready to shoot (on balance, square to the basket, etc.).
3. A good shot is one that you shoot a high percentage on in drills and practice (in your range)
4. A good shot is one that can be rebounded by at least two of your teammates.
5. A good shot is one that you can recover and play defense from if it is missed.
6. A good shot is one that is appropriate given the time and the score.
7. A good shot is one that is taken when you are not closely guarded (except for shots around the basket).

This list is a perfect guide for a basketball learner who wants to get an impression of when to shoot a basketball, although some of the points are debatable. For key no.4, it states that a good shot can be rebounded by the shooter's teammates. However, an open layup in a fast break is always a good shot, while you can only expect the player who is shooting it can rebound the basketball. For example, key no.1 is not fit for automation. It is extremely difficult to detect the expectation of a player. There are enough examples of shots that are not expected by anybody else but ended up in the basket. Some of them are remembered as great ones in the history of a basketball game. We would not believe this factor is a strong one to influence the quality of a shot. Key no.5 is also not that relevant, as we always tell players to have confidence at every shot. Therefore, nowadays, coaches are less strict on recovering from the misses.

Our goal is to estimate the probability of success on scoring the shot based on the historical data. Create a visual analytic system allowing the user to evaluate the shot quality. This is much more reproducible than examining the shot quality by using subjective methods. Coaches and players can evaluate the shot quality in a much more intuitive measure. Our assumption is:
A shot's quality is determined only by the probability of success on scoring the shot. Which depend on three factors

1. The skill of the shooter.
2. The distance to the defenders.

3. The distance of the shot.

These three factors are covered by Alan Stein Jr's [20] shot selection keys. Firstly, the skill of a shooter can conclude most of the keys no.2 and no.3. The choice of shooting technique is part of shooting skill. A good shooter can adjust the shooting style according to his situational awareness. In addition, if his teammates admit a player's shooting skill, his teammates would expect his shot if he is in a good position to shoot and vice versa. Therefore it matches the key no.1. Speaking of the good position to shoot, that involves the following two factors: The distance of the defender and the distance of the shot. We argue that these two factors are essential factors that influence the shot quality. The defender's job is always to either stop a shot or change one. The closer the defender to the shooter, the easier he can do so. It is obvious, the closer to the target, the easier you can hit. We assume shooting from left or right, i.e., the location of the shot. It Will not significantly change the shot quality since the mechanism is not changed in a different location but is changed at a different distance, as stated in Satern's work.[17]

6.1.1 Shot quality estimation

We can estimate the probability of success in scoring based on the historical data. We aim to build a model for each player. Our model is based on the relative distance to the defender D_d and the distance to the basket D_b as parameter space. The goal is to estimate the conditional probability of scoring a shot given D_d and D_b. Once the probability is calculated, we can then classify the shot quality as being a good or bad shot.

Configuration D_d and D_b

To provide the parameters D_d and D_b, we need to locate the exact moment the shooting action is happening. In the data set, we have the information of who shoots the ball, namely the player P. The position of this player is given as $\vec{X}_P$. We can use the information in Play-by-play data to derive which basket he is supposed to be shooting at and derive the position of basket $\vec{X}_B$. We also have the location of basketball as $\vec{X}_{ball}$. So the shooting moment M_{shot} is a moment i satisfied the condition that player P must hold the ball. That is:

$$d(\vec{X}_P, \vec{X}_{ball}) \leq \tau \tag{6.1}$$

where τ is a small threshold guarding the holding of the ball, and d denotes the 2-dimension Euclidean distance. Meanwhile, we can calculate the velocities of player P and basketball B in moment i by using :

$$\vec{V}_{P_i} = \vec{X}_{Pi+1} - \vec{X}_{Pi} \tag{6.2}$$

$$\vec{V}_{ball_i} = \vec{X}_{ball_{i+1}} - \vec{X}_{ball_i} \tag{6.3}$$

As a shot is supposed to go towards its target, we would like our moment to satisfy: The angle between the player to the rim and the velocity of the ball should be smaller than a threshold:

$$\arccos\left(\frac{\vec{V}_{ball_i} \cdot (\vec{X}_B - \vec{X}_{ball_i})}{||\vec{V}_{ball_i}||||\vec{X}_B - \vec{X}_{ball_i}||}\right) \leq \alpha \tag{6.4}$$

The α can be kept relatively small. If the shot goes far off the target, we would not call it a shot. Last but not least, at the shooting moment, the ball should be above a man's height and going even higher. This eliminates the possibility of the player is just dribbling to the rim with the ball, or some passing is just across the player P.

Once we detect the corresponding moment for all shots, the next step is to derive the parameters we need. In the following equation, we are in the context of the single moment of shooting. The distance to the basket is relatively simple to compute:

$$D_b = d(\vec{X}_P, \vec{X}_B) \tag{6.5}$$

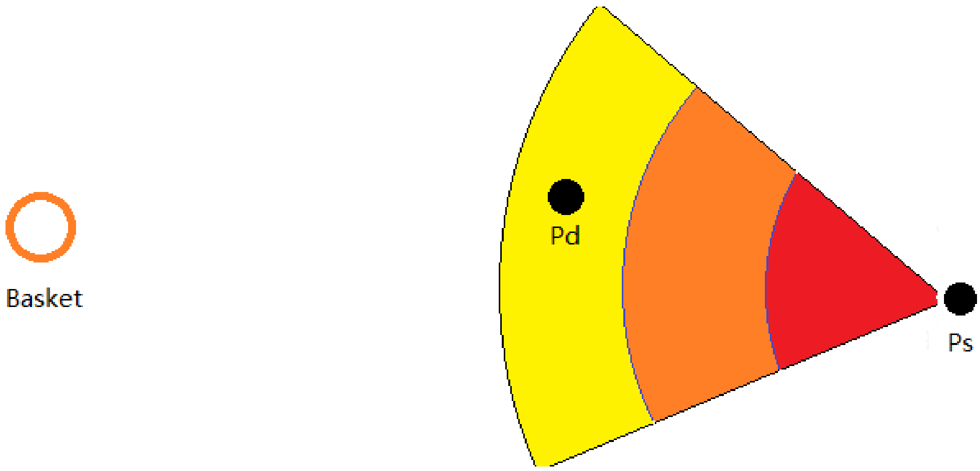


Figure 6.1: The effective defence area is modelled as a cone shape. The color scale denotes the interference that the defender gives to the shooter

To derive the defence factor D_b, first, we have to define who is the defender. As we know, the defender is trying to block the shot towards the basket. He is only effective when he is on the path of the shot projected position on the floor. The closer to the shooter, the more interference is given. We simulate this effect by using a cone shape illustrated in Figure 6.1. It is not the perfect model. We realize that people from right behind the shooter could influence the shooter as well. Besides, the variety of dominant hand should also have changed the shape of the area. However, for simplicity, we use a cone shape for now. This model can be replaced by a more accurate one in the future.

Usually, there is a single defender in this area. However, sometimes there could be multiple defenders. In this situation, we take the closest one in this area because other players further away do not create as much difficulty for the shooter P_s. To conclude, the defending player P_d is the player who found himself between the shooter and the rim, not in the same team of the shooter, and closer than all his teammates to the shooter in the shooting moment.

$$\arccos\left(\frac{(\vec{X}_B - \vec{X}_{Pd}) \cdot (\vec{X}_B - \vec{X}_{Ps})}{||\vec{X}_B - \vec{X}_{Pd})||||\vec{X}_B - \vec{X}_{Ps}||}\right) \leq \gamma \tag{6.6}$$

which γ is the angle threshold, it should be larger than α, we know that the pro players with long arms can reach in the shooting path from so far away. The angle α and γ is illustrated in Figure 6.2

We also have the conditions:

$$team(P_d) \neq team(P_s) \tag{6.7}$$

and

$$d(P_d, P_s) = min(d(P_i, P_s)) \tag{6.8}$$

where P_i is all the players on the court not in player P_s team. For each shot we have assigned a defending player guarding the shot, we can simply take his distance to the shooter to represent the relative position. That is:

$$D_d = d(P_d, P_s) \tag{6.9}$$

These are the input parameters that define our space.

Shot quality probability density function

There are patterns in the basketball shots. We observe that some shots similar to each other tend to have the same result. E.g., almost every player does not miss open layups very often. Every individual player also

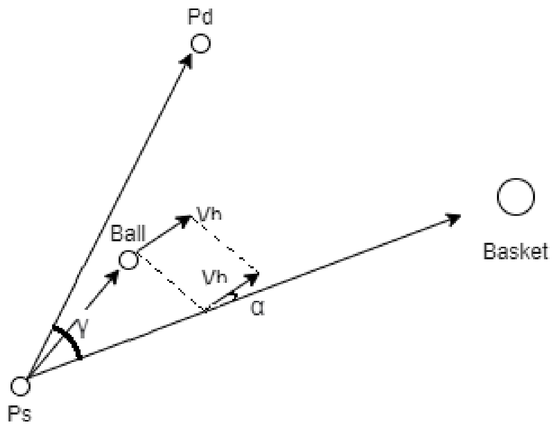


Figure 6.2: The 2D angle between the shooter to basket and the ball to basket/shooter to defender

has their own shot behaviors. E.g., DeAndre Jordan is not missing shots around the rim, and Stephen Curry has a high percentage in long-range. The core idea is to estimate the probability of a successful shot in the parameter space defined by D_b and D_d based on the data. A well-known classifier is KNN, namely the K-nearest neighbours algorithm. K-nearest neighbours are developed by Evelyn Fix and Joseph Hodges in 1951.[7] KNN can be used for both classification and regression. The concept is to find k nearest-neighbours of the point in the $\boldsymbol{X} \in D_b \times D_d$ space to be classified. Then we count the number of successful shots compares to the overall shots to obtain the probability of success of neighbours. The point can be labeled as the class with a larger number. KNN can also approximate the likelihood of making a shot by probability $P(x)$ density function.

$$p_s(\boldsymbol{x}, k) = \frac{k_s(\boldsymbol{x}, k)}{k} \tag{6.10}$$

where k is the number of neighbours we pick. We define a function to determine a label of the scored shot.

$$scored(\boldsymbol{x}_i) = \begin{cases} 0 & \text{if } \boldsymbol{x}_i \text{ data point scored} \\ 1 & \text{if } \boldsymbol{x}_i \text{ datapoint not scored} \end{cases} \tag{6.11}$$

Therefore $k_s(\boldsymbol{x}, k)$ is $\sum_{i=1}^{k} scored(\boldsymbol{x}_i)$, we can denote $p_s(\boldsymbol{x}, k)$ as:

$$p_s(\boldsymbol{x}, k) = \frac{\sum_{i=1}^{k} scored(\boldsymbol{x}_i)}{k} \tag{6.12}$$

where $\boldsymbol{x}_i \in N_k(\boldsymbol{x})$ and $N_k(\boldsymbol{x})$ are the k closest neighbours.

In addition, our historical data is sparse in the parameter space. We do not want the point we estimate influence by a point far away from it equally than one very close because of lack of data. We apply weight in our estimation to overcome this. One common weight function is the Gaussian kernel[15]. We derive weight by:

$$w(\boldsymbol{x}) = e^{(-\frac{|\boldsymbol{x}|^2}{\sigma^2})} \tag{6.13}$$

where $\boldsymbol{x}$ is the position of the point we want to give a weight, σ is a parameter we give. We can multiply the weight w into our equation (12) and get. Normalization is needed such it is a probability. The weighted probability is:

$$P_{ws}(\boldsymbol{x}) = \frac{\sum_1^k w(\boldsymbol{x} - \boldsymbol{x}_i) score(\boldsymbol{x}_i)}{\sum_1^k w(\boldsymbol{x} - \boldsymbol{x}_i)} \tag{6.14}$$

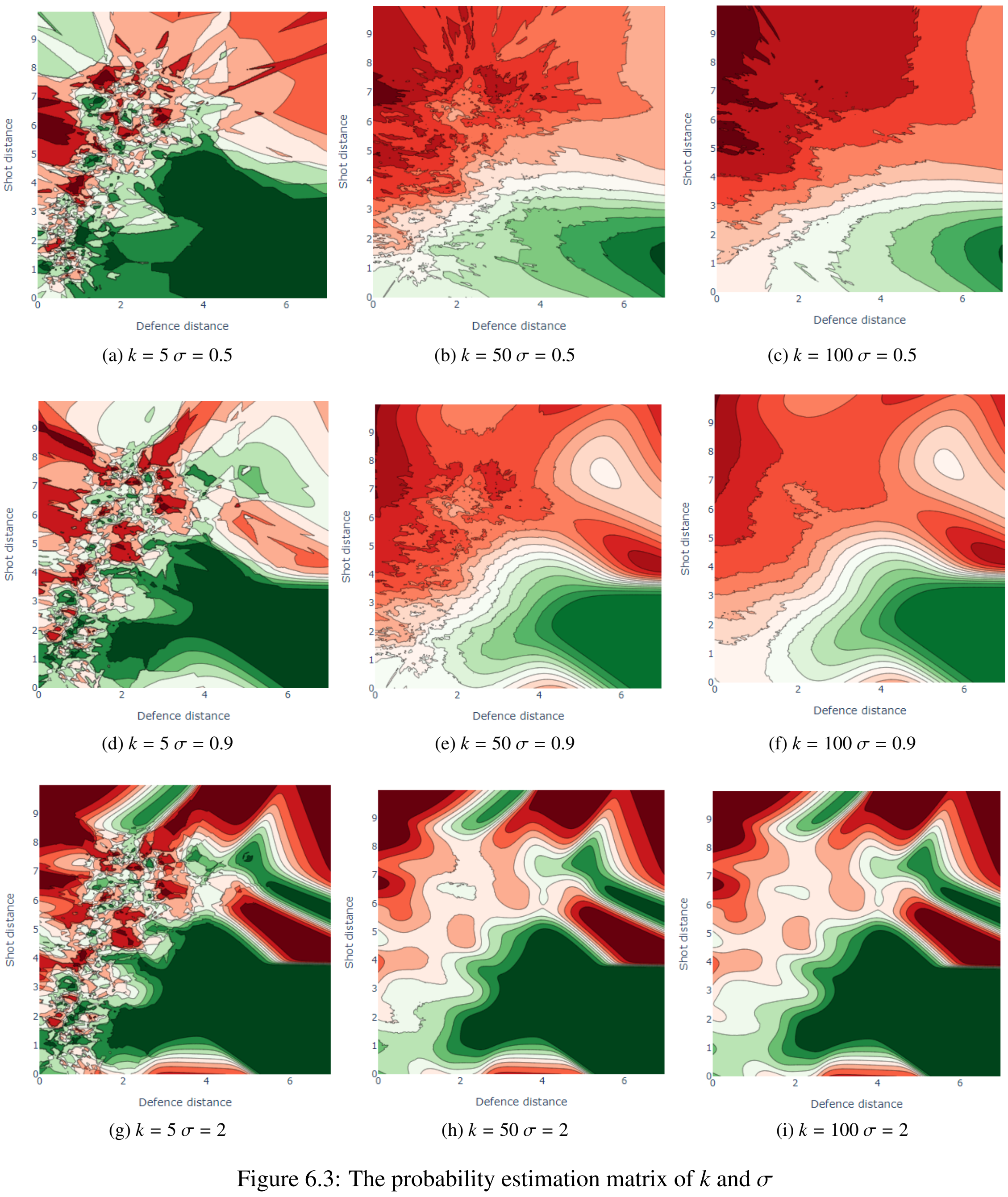
The opposite is the probability of missing the shot is:

$$O_s = 1 - P_{ws}(\boldsymbol{x}) \tag{6.15}$$

If $P_{ws} \geq O_s$ then we classify the shot is a good shot, and mark it by green colors.

We have to select the variables, namely k and σ, carefully. K is the number of neighbours we take into account when estimating a point. When k is too small, the estimation point will only sensitive to the few neighbours of it, the map getting overfit the user cannot evaluate anything from it, as shown in (a),(d),(g) of Figure 6.3. On the contrary, if the k is too large, we take almost every shot into account when we estimate a shot,(c),(f),(i) of Figure 6.3. This will result in almost no different colors in the parameter space. This k selection highly depends on the size of the data. For the size of one season's shots, we found when $k = 50$ shows us a good result when we estimate a 5-players lineup.

The σ is called the standard deviation, which determines the scale of the Gaussian function. The larger the σ we set, the estimation graph will become more smooth. We illustrate how k and σ combination affect the estimation graph in Figure 6.3.We found the $k = 50$ and $\sigma = 0.9$ setting is the best fit for our requirement to avoid loss of information, as well as to obtain a smooth estimation,



(a) $k = 5$ $\sigma = 0.5$ (b) $k = 50$ $\sigma = 0.5$ (c) $k = 100$ $\sigma = 0.5$

(d) $k = 5$ $\sigma = 0.9$ (e) $k = 50$ $\sigma = 0.9$ (f) $k = 100$ $\sigma = 0.9$

(g) $k = 5$ $\sigma = 2$ (h) $k = 50$ $\sigma = 2$ (i) $k = 100$ $\sigma = 2$

Figure 6.3: The probability estimation matrix of k and σ

Chapter 7

Visual design

7.1 Visual Design

Our aim of the SQVA system is to allow the user to analyze shot quality from various perspectives and provide comparisons between players, lineups, and teams. Figure 7.1 is a screenshot of our SQVA tool. We use capital letters to denote each element of the SQVA system. The A element is the Team shot quality VA. B is the lineup shot quality VA. The C, E, D and F together construct the shot quality probability estimation VA, including three sub-elements: C is the probability estimation contour. D is a basketball half court showing the exact position of shoot and defender for each shot in the historical data. E is the histogram of the historical shots distribution. F is an indicator showing the shot distance and defense distance of any point in the parameter space we want to estimate.

7.1.1 Team shot quality

To answer the team level questions we mentioned in section 4.1.2, i.e. 1(a) Which team has better shot quality generated during the season? 1(b) How much the game result depends on the shot quality of themselves' and the shot quality they gave to their opponents? One of the most widely used and famous metrics in this area is John Hollinger's offensive efficiency. John Hollinger is the former Vice President of Basketball Operations for the Memphis Grizzlies of the NBA and current Senior NBA columnist at The Athletic. He thought of some analytic features. One of the famous ones is called offensive efficiency. The concept is simple: how many points can a team produce per 100 possessions. This is related to our shot quality metric because, in our theory, if a team keeps getting the shot they are good at, they are more likely to score more points with each of them. The offensive efficiency can be approximately seen as the expectation score per shot times 100. It is not precisely the expectation score per shot because possession can be ended in other ways than a shot. e.g. A turnover. The rank of all 30 teams in the NBA is shown in Figure7.2.

Our approach is different. We choose to use a 2D-scatter diagram with 30 teams in NBA to visualize our shot quality estimations. The x-axis is the team's win rate, namely the number of winning games divided by the total played games. The y-axis is the average shot quality we calculated by adding up all shots' P_{ws} that are taken by a team in the whole season and divided by the total shots to get an average shot quality of a team.

$$average = \frac{\sum_{i=1}^{n} P_{ws}(\boldsymbol{x}_i)}{n} \tag{7.1}$$

where n is the total number of shots taken by the team in the season. $\boldsymbol{x}_i$ is the shot's parameters taken by the team during the season. If we change n to the total number of shots taken by a lineup/player, then we calculated the average shot quality of a lineup/player. Therefore this number is how likely the shot will go in when a certain team take a shot in average. As shown in Figure 7.3, we plot every team in NBA 2015-2016 season with their average shot quality and their game-winning percentage in that season. The more to the top denotes the better shot quality the team shoot with, and the more to the right means the

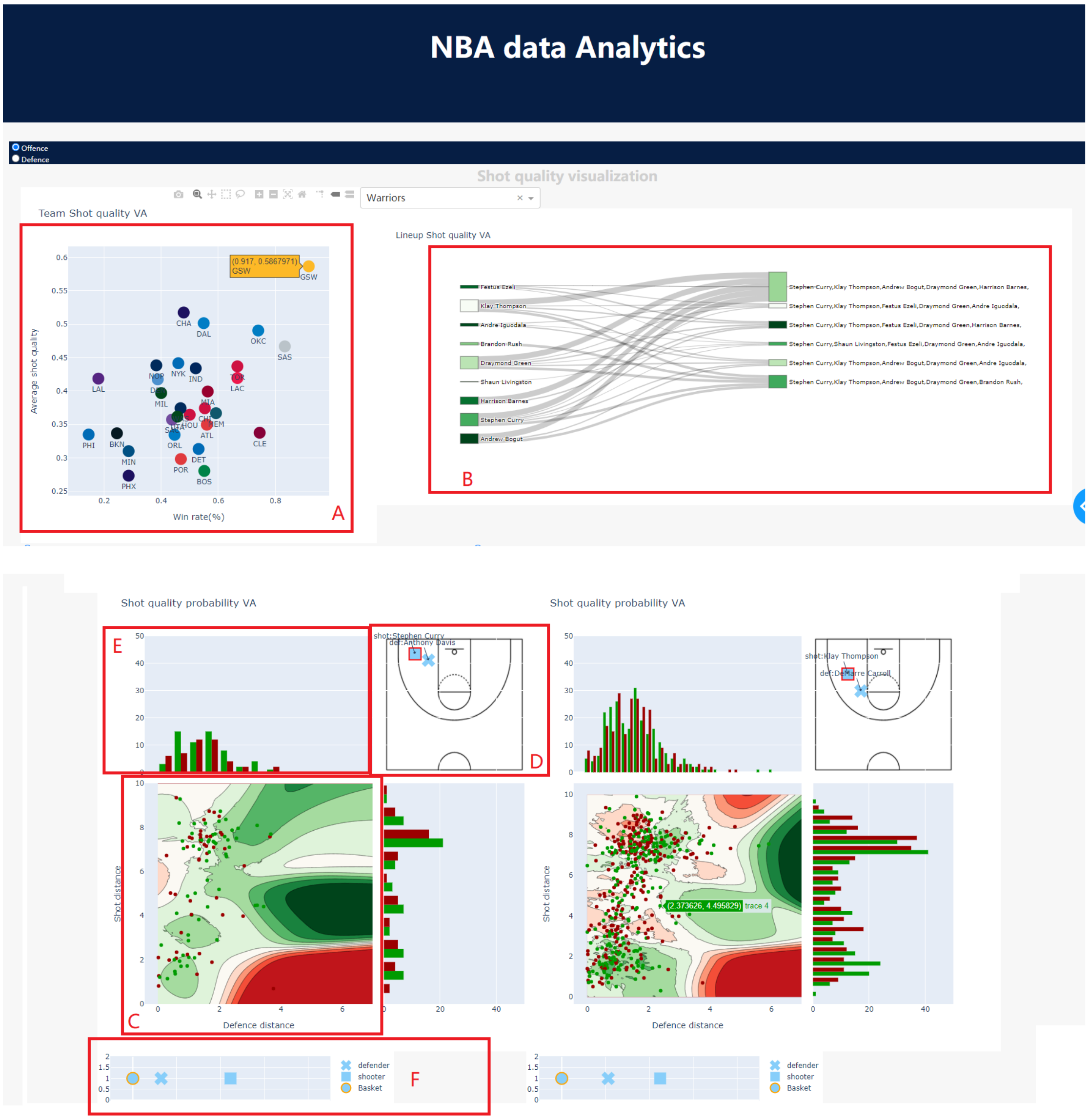


Figure 7.1: A screenshot of the SQVA.

Hollinger Stats - Offensive Efficiency

RK	TEAM	PACE	AST	TO	ORR	DRR	REBR	EFF FG%	TS%	OFF EFF	DEF EFF
1	Golden State	101.6	20.5	13.5	23.5	76.0	51.3	56.3	59.3	**112.5**	100.9
2	Oklahoma City	99.4	16.8	14.0	31.1	76.0	54.7	52.4	56.5	**109.9**	103.0
3	San Antonio	95.7	18.9	12.4	23.0	79.1	52.0	52.6	56.4	**108.4**	96.6
4	Cleveland	95.5	17.5	12.7	25.1	78.5	52.0	52.4	55.8	**108.1**	102.3
5	Toronto	95.3	15.0	12.3	24.6	77.7	51.6	50.4	55.2	**107.0**	102.7
6	LA Clippers	98.0	17.6	12.1	20.1	73.8	47.4	52.4	55.6	**106.5**	100.9
7	Portland	98.3	16.2	13.2	25.9	76.2	51.0	51.1	54.8	**106.1**	105.6
8	Houston	100.1	16.5	14.2	25.7	72.8	49.1	51.6	55.3	**105.5**	105.6
9	Charlotte	97.8	16.8	11.7	20.0	79.8	49.6	50.2	54.5	**105.1**	101.8
10	Dallas	96.4	17.2	12.0	20.6	76.2	48.5	50.2	54.4	**104.8**	104.3
11	Minnesota	97.6	17.8	13.9	24.3	74.7	49.8	49.8	54.9	**104.3**	107.1
12	Miami	95.7	16.4	13.3	23.8	77.8	51.7	50.8	54.5	**104.2**	101.6
13	Boston	101.1	17.6	12.1	25.1	74.6	49.4	48.8	53.1	**103.9**	100.9
14	Sacramento	102.2	17.7	14.2	23.9	74.9	49.6	51.0	54.6	**103.3**	106.3
	Detroit	97.4	14.9	12.2	27.0	79.3	52.1	49.1	52.2	**103.3**	103.4
16	New Orleans	98.9	16.9	12.3	21.2	78.8	49.0	49.8	53.7	**103.2**	107.3
17	Utah	93.3	15.2	14.2	25.9	77.7	51.9	50.1	54.0	**103.1**	101.6
18	Atlanta	99.4	19.1	13.8	19.1	74.6	47.5	51.6	55.2	**103.0**	98.8
19	Washington	100.6	18.2	13.1	20.6	77.7	48.6	51.1	54.4	**102.9**	103.6
20	Denver	98.2	17.0	13.2	25.8	77.3	51.1	48.9	53.1	**102.7**	106.4
21	Orlando	98.2	17.7	12.8	23.1	76.5	49.4	50.0	53.3	**102.6**	104.6
	Memphis	95.7	16.1	12.3	25.3	75.1	49.1	47.7	52.4	**102.6**	105.4
23	Indiana	99.0	16.2	13.5	23.4	76.0	49.8	49.7	53.6	**102.4**	100.2
24	Milwaukee	96.6	17.7	14.2	24.9	73.1	49.2	49.9	53.7	**102.2**	105.7
25	Chicago	98.2	17.1	12.6	24.5	74.9	50.2	48.7	52.6	**102.1**	103.9
26	New York	95.8	16.1	12.6	23.7	75.8	50.1	48.3	52.7	**102.0**	104.8
27	Brooklyn	97.4	17.1	13.6	24.1	75.7	49.4	49.2	52.7	**100.9**	108.5
28	Phoenix	100.9	15.5	15.2	25.4	77.1	50.6	48.7	52.6	**99.4**	107.0
29	LA Lakers	98.0	14.1	12.5	23.1	74.7	48.0	46.0	50.9	**98.6**	109.3
30	Philadelphia	100.2	16.3	14.9	20.6	74.0	46.4	48.7	51.9	**96.6**	106.7

Figure 7.2: The offensive efficiency ranks of 30 teams in NBA, the highlighted column is the value of offensive efficiency

more games they won during that season. We use the classical color of the team's theme and the short name to indicate different teams. Users can discover, for example, the correlation between the shot quality we analyze and the result of that season. As for all the graphs, if we change to the defence perspective, the horizontal axis becomes the average shot quality of shots taken by the team's opponents. This graph also acts as a team picker. User can select a team by clicking on the markers in this graph and expand the information within the particular team.

Let us compare John Hollinger's offensive efficiency Figure 7.2 with our team shot quality VA of SQVA in Figure 7.3. We observed that some of the teams are consistent in both rankings. For instance, the Golden state warriors are in a high place in both of the rankings. The Oklahoma Thunder and the San Antonio Spurs are also doing well. Phoenix Suns and Philadelphia 76ers are not doing so well in both the rankings. On the contrary, there are teams that rank much higher in one metric than the other. As we stated earlier in this thesis, shot quality is not the direct result of offensive possession. It reflects only the success of the offensive strategy and its execution. Some of the team works hard together to get the many shots they want successfully, but they are just not good enough to make them. Charlotte Hornets is a perfect example of this case. They rank the highest in the second legion, just behind the Golden state in terms of shot quality. However, drop to No.9 in the offensive efficiency ranking. They run better offence than any teams besides Golden state. However, their team lacks talent, as they do not have a single all-star player in that season. The Cleveland Cavaliers is on the exact opposite way. They have the best player in the world, Lebron James, and arguably the best offensive player Kyrie Irving who both well known as capable of making ridiculous difficult shots. Their team is not bothered to create better shots as they can make those difficult ones. As a result, the Cleveland Cavaliers' ranking only 23rd place in the shot quality metric while finding themselves a fourth place in the offensive efficiency.

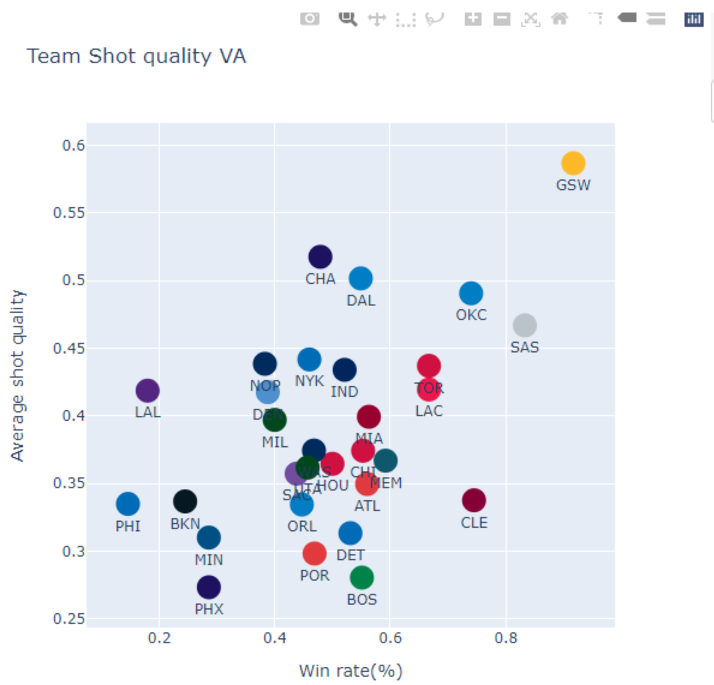


Figure 7.3: The scatter plot of teams shot quality and their win rates.

7.1.2 Lineup Sankey diagram

In section 4.1.2, we also have a question related to lineups. To answer questions like 2(a) Which lineups are played for each team, and how is the playtime distributed? 2(b) How good are the shots generated for each lineup? For a given team? We want to display these elements in a single visualization, emphasizing the playing time, average shot quality and the inclusive relation of players and lineups. We choose to use a colored Sankey diagram as shown in Figure **??**. The Sankey diagram is composed of nodes and lines with width. We can visualize the lineups effectively by utilizing the features. This diagram explicitly shows how the playing time is allocated and the shot quality comparisons among them. The size of the nodes, along with the width of the lines representing the playing time of a certain individual/lineup. We use the color to represent the average shot quality calculated as equation (15). Furthermore, the Sankey diagram enables us to show the relation between players and lineups clearly. When the user hovers on a lineup, the players that belong to this lineup is highlighted. On the other hand, when the user hovers on a player, all his involved lineup is highlighted. User may want to explore how these are related and find out if a player is working well with some specific teammates to answer question 2(d) to provide potential improvement of the lineup strategy.

If the user is on the defence view option, the Sankey diagram shows differently than the offence. The color scale turns red, and the color indicating the average shot quality of the lineup's opponents'. The closer to red color means the opponents' shooting is worse, representing the lineup doing a better job.

The team shot quality diagram and the Lineup Sankey diagram is closely connected. We put them in the same row, so users can always switch between teams and compare their lineups strategies easily. Users can switch to the defence perspective to examine the defence performance of these lineups. The nodes appear red, and the redder it is, meaning the opponents shoot bad quality shots, giving the opponents a hard time. If users click on the node representing players or lineups, we can see the shot quality predictions in a contour.

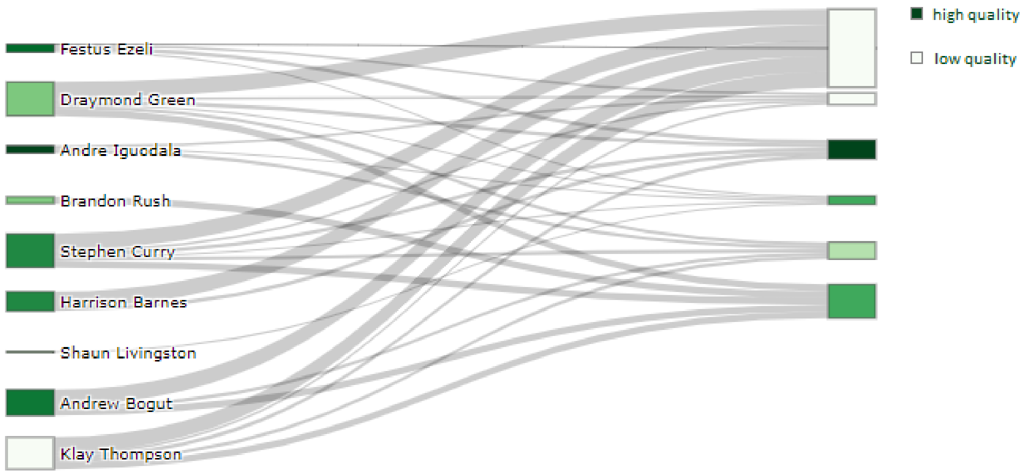


Figure 7.4: The Sankey diagram showing how players are allocated in lineups.Colors represent the average shot quality, Size of the nodes represent the playing time of the player/lineup.

7.1.3 Visualization of shot quality probability

To answer the players and lineup related questions in section 4.1.2, we introduce the visualization of shot quality probability. In this section, we explain each element the shot quality probability is composed of.

From the historical data of individual shots, we created a density map for both the successful shots and missed shots as illustrated in Figure 7.5. The users can evaluate the shots from a lineup/player and answering where are the most shots success and where are the most shots miss. However, most of the miss and success shot are from a similar area because most shots are taken from that area in the $D_d D_b$ space. Therefore, we want to show the P_{ws} we estimate in the view.

To enable the user to estimate every shot probability, we want to create a contour showing the probability by the scale of colors. As Figure 7.6 shows, the C part of the visualization is the prediction of shot quality according to our KNN algorithm. The horizontal axis represents the defender's distance D_d from left to right. The vertical axis represents the distance to the basket D_b from bottom to top. The color scale in this contour is the shot quality we predict by equation (14), from red 0% probability bad shot to green 100% probability good shot.

The scatter plot on top of the contour is the individual shots. From the individual shots, we calculate the P_{ws}, and build the contour base on it. Each dot represents a shot, the green dot indicating that shot is a success, and the red ones are those missing. If the user is distracted by the historical data scatter, he can switch it off and only keep the contour for analysis.

Suppose the user wants to see the historical data closely. In that case, he can see the exact shot position view by clicking on the dot, and the exact location in the basketball court is shown in the D part of the Figure 7.6 with the shooter's name and defender's name. With the shooter's location and defenders location, the user will get more insight into what is happening in the basketball court and even recall the shots in memories.

To answer the distribution-related questions in section 4.1.2, i.e., the distribution of the player's shots and their success. We choose to attach histograms to the top and right side of the contour as shown in the E part of 7.6. The histograms present the distribution of the shots. It shows the user the density of the shots

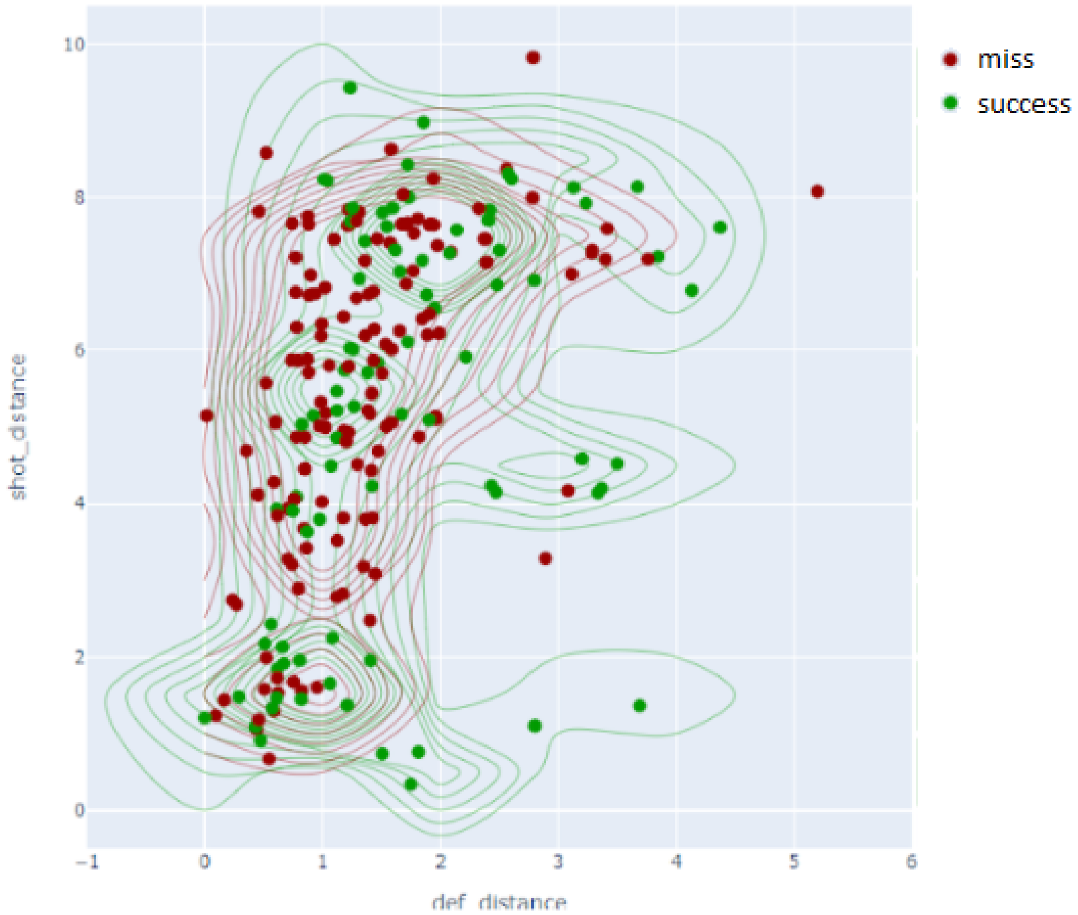


Figure 7.5: The density map view

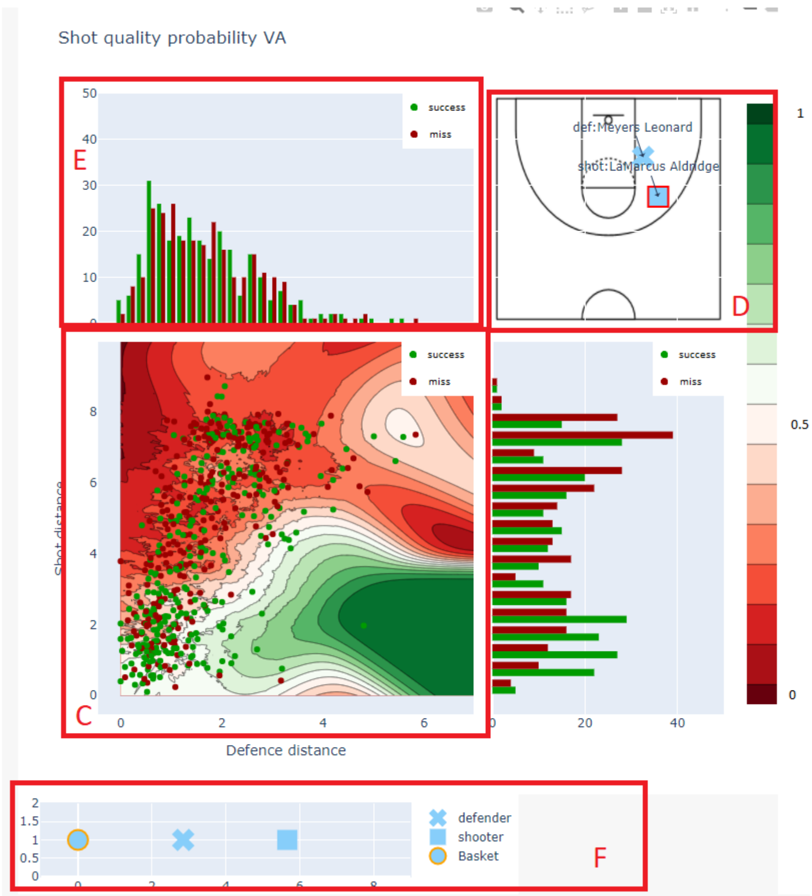


Figure 7.6: The shot quality probability view

in both dimensions, we use the green bar to represent the successful shots, and the red bar represent the miss shots. It provides a clear overview of what kind of shots in terms of D_b and D_d is produced more by the lineup and the amount of successes.

The widget at F in Figure 7.6 is the shot distance view. This visualized the D_b and D_d to give the user an idea of the relative distance of the basket, defender and shooter. When the user hovers on any point of the contour, the shot distance view dynamically shows a square, a cross and a round, where the square representing the shooter, the cross representing the defender and the round representing the basket. The users can see the distance changing while he is moving his mouse so that he always get the feeling of what shot he is estimating.

The SQVA system has all views connected to each other via interactions. In the A part of 7.1, if the user clicks on any team in the scatter plot, the corresponding lineup will show up in the B part of 7.1. Similarly, if the user clicks on any node in the Sankey diagram in the B part. The corresponding shot quality probability estimation view will appear in the C part. Furthermore, if the user click on any player or lineup once again, another shot quality probability estimation view appear next to the previous one, providing a comparison for the user, so he can evaluate the two lineups/ player together. Once the comparison is started, Part A and B of 7.1 will fade away from the screen so that the user can concentrate on the shot quality probability views. The user can always return to the team shot quality and lineup shot quality view by clicking on a button after entering the comparison view.

Chapter 8

Implementation

8.1 Implementation

We implement the SQVA in Python. Popular libraries like Pandas and NumPy fits the data processing job. We can manipulate the data set using very few lines of code.

We also use the scikit-learn machine learning tool, Skikit-learn is a Python library build for machine learning. It can perform various classification, regression and clustering algorithms, including the KNN we have used in our system. Skikit-learn is designed to interoperate with the Python numerical libraries NumPy. Thus make our implementation much easier.

We build our user interface based on the Dash plotly framework. Dash is a Python framework for building web visual analytic applications. It is built on top of several JavaScript libraries. As a result, we can build web applications by only coding in Python. We choose to use a web-based framework because we want to build a visual analytic system that every part connects closely to each other so that user can explore and interact with the data on one screen. Dash plotly suites our demands. The visual components are built in the web server and rendered in a web browser. It provides the experience we want the users to have.

Chapter 9

Evaluations and use cases

9.1 Evaluations and use cases

We put the SQVA system in the test of real cases to prove the system is helpful for NBA game analysis and find our system's weaknesses. We consult the Sports club in our university, the ESBV Tantalus. The ESBV Tantalus basketball club was founded in 1961. Men's first and ladies' first teams are competing in the national level first division and second division, respectively. They also organized an international tournament for forty years, inviting teams from all over the world to Eindhoven for the championship. [3]
We are fortunate enough to invite three participants from ESBV Tantalus to evaluate our SQVA system. As our domain expert, we invited the men's third team's head coach and one of his players.
First of all, we demo the SQVA to them. We explained the estimation of the probability. We show all the components to them as well as how to interact with them. Then we show an example of evaluating the offense of San Antonio Spurs and their lineups. After the demo, we asked them to use the SQVA system to perform the tasks we gave to them. The evaluation result is discussed in this section. We try to form tasks that conclude the questions of Lineup level and player level in section 4.1.2. The tasks we designed are:

1. Evaluate the shot quality from two favourite players in NBA. Answer the question: What kind of shot is the player good at? Is the player shooting the shots he is good at?

2. Using the SQVA tool to find out strategies for both teams in the Finals of the 2015-2016 season. How can they do better?

9.1.1 Evaluation results

The player chooses to evaluate his two favourite players Lebron James from Cleveland Cavaliers and Klay Thompson from Golden state Warriors as Figure 9.1 shows.
He sees the shot quality VA reflecting Lebron's shot quality is high. Once he entered 4 meters area from the rim, he barely misses shots, which is in the A area of the contour. According to the B area in the histogram, he also gets many shots from there. Klay is evaluated as a stable outside shooter, as he expected. The C area in the contour shows that around 7 meters to 9 meters, the shot quality is higher when Lebron is guarded than open. The SQVA suggests not to guard the best player in the world at three-point range. The D and E area of the histogram showing Lebron missing those open shots are there, compared to the F area of Klay Thompson's contour, showing that Klay can hit consistently when no one is guarding him on that range. So something must happen to Lebron's outside shooting, either mentally or mechanically. SQVA does not directly suggest any tactics. It only provides a view to helping the coaches to create strategies. However, this phenomenon exposes one of the limitations of SQVA. If we can show more context of the missing shots by Lebron, we can probably help the user find a pattern so they can explain why this is happening.

On the other hand, the coach is more interested in the point guard position (one of the five positions in the basketball game), the position that is often referred to as the coach on the court. He is willing to see the

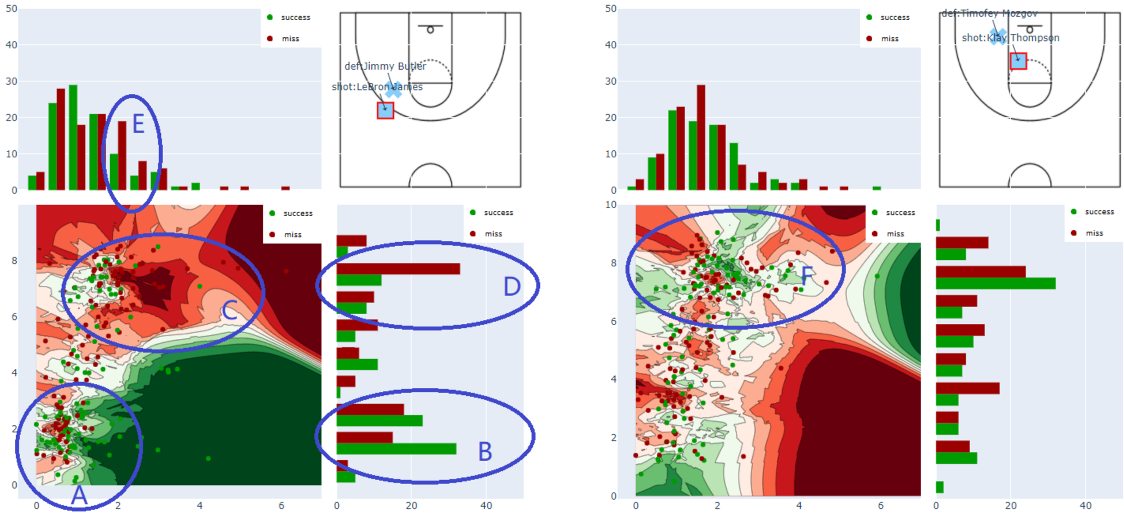


Figure 9.1: The shot quality compareison between Lebron James(left) and Klay Thompson(right)

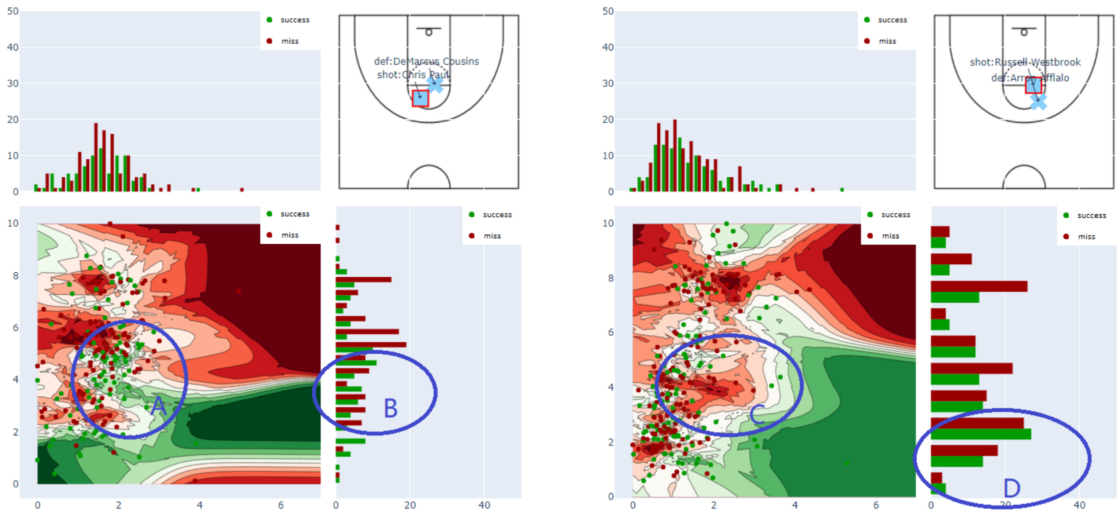
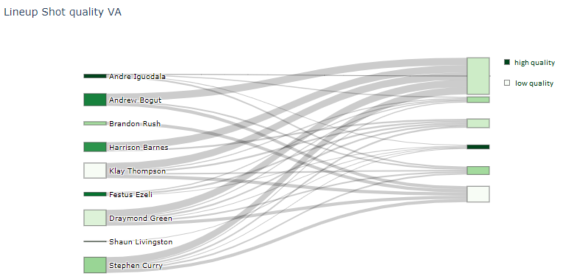


Figure 9.2: The shot quality comparison between Chris Paul(left) and Russel Westbrook(right)

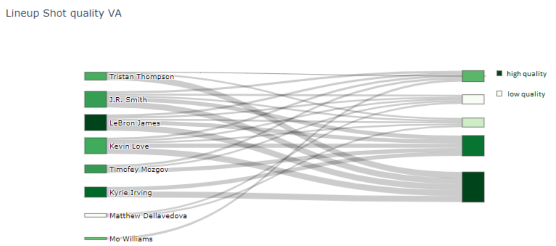
shot selection of the two best point guards in the NBA because he thinks like a point guard. Shot selection is essential because he is the player who dominates the ball all the time. In other words, he can shoot all the shots he wants if he thinks his team can benefit from it.

After exploring the shot quality of these two players shown as Figure9.2. The coach observed that Chris Paul is better at mid-range jump shots. The A area and the B area in the contour show he is a relatively better shooter between the 4 meters to 6 meters area. In addition, most of these shots are not so heavily guarded. The coach suggests most of the shot, probably due to his passing ability. He always finds his teammates either open at the three-point line or under the basket. Therefore his opponents cannot guard him for those mid-range shots. Russel Westbrook, from the C area of his contour, we can see he is less efficient on the mid-range shots. However, he is so strong and athletic to get inside the area around the rim. If we look at the D area in the histogram, we can find he gets plenty of shots from there and considered as a high quality shot even with a defender on him.

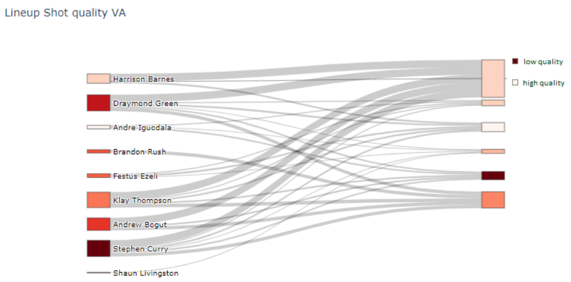
In the next session, we ask the player and coach of Tantalus B.V to study a case together. We pick two teams that played in the final game of the 2015-2016 season: Cleveland and Golden State. We use the SQVA system to evaluate these two teams and find out how they can win the NBA title. We make some strategies for both teams to overcome the opponent and get the championship. We look at the team shot quality overview as shown in Figure 7.3. The Golden State ranks itself at the top both in average shot qualities and wins rate percentage. There is no team even close to them. Cleveland, on the contrary, ranks relatively lower in the average shot qualities but still comfortably lies in the top five teams in terms of win rates. Therefore, the average shot quality does not directly determine if a team will win many games. However, it can be a factor, especially for those teams built based on smart shot selections like Golden states.



(a) Offence lineup Warriors



(b) Offence lineup Cavaliers



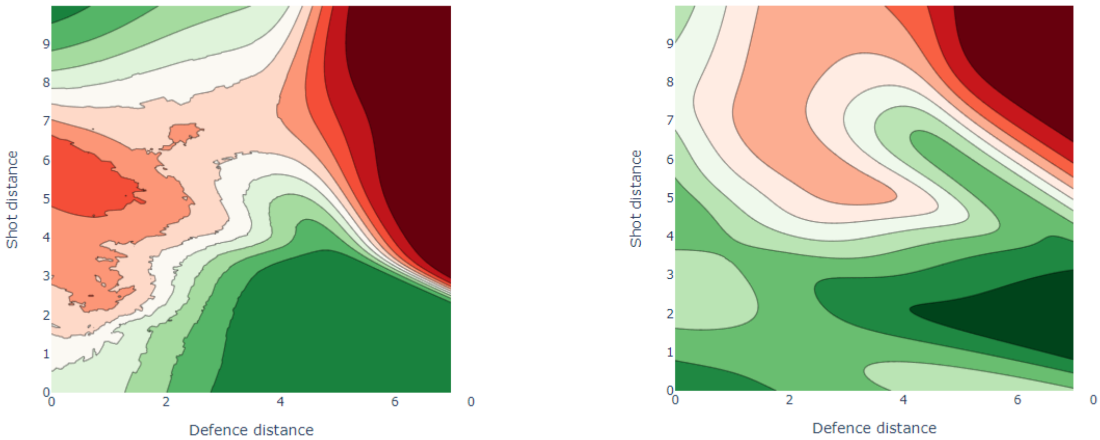
(c) Defence lineup Warriors

(d) Defence lineup Cavaliers

Figure 9.3: The lineups graph with shot qualities, (a) and (b) is in the offence point of view, (c) and (d) is in the defence offence point of view.

We take a closer look at the two team's lineups. In Figure **??** (a), we see that the lineup: Curry, Thompson, Barns, Green and Bogut play the majority of the time. As their core players, Curry, Thompson and Green played in almost all lineups that have meaningful playtime. One thing we noticed is that these three players average shot quality are relatively lower. This is exactly the opposite case when we observe the Cleveland team. Irving and James are their core players, James plays in all the Lineups with meaningful minutes, and Irving plays in two of them because of injury in the early season. We can see a completely different pattern in their shot quality distributions. James and Irving have a much better shot quality than their teammates. The coach explained this phenomenon to us. These patterns show us the different approaches of the two teams. Warriors' offense tends to let their star players take the difficult shots and leave the easy ones to their teammates. The Cavaliers simply give their best player the best shot opportunities and leave the rest to others. These are two basketball philosophies that co-exist. It is exciting to observe that this co-exists by using the SQVA system. We also find something noticeable in the defense point of view. In Figure **??** (c), although Stephen Curry is not famous for his defending, we can see he is actually a good defensive player in terms of taking away the shots his opponents good at. This can be reflected both in the Sankey diagram and the contour of Stephen curry's individual shots. What we can also observe is that Andrew Bogut is a big boost for their team defense. The lineups nodes with his own node connected are redder than those that do not connect to his node, showing he has a much better defensive performance improvement in terms of shot qualities. This is also what they think a nice feature of the SQVA.

Examine the lineups in the Sankey diagram help us know better for the players like Andrew Bogut, whose value is not shown much as an individual a part of a group. He definitely takes away the better options his opponents' and rather take some alternative ones and not counted on him as a defender. This effect is hardly revealed in similar Visual analytic tools. The Figure **??** (d) illustrates the Cavaliers lineups' defence performance. James's node is completely red, showing he is their best defensive force, and Irving is almost white, showing he is the weak link in all lineups. A similar effect of the Bogut case applies to Tristan Thompson. Although he did not give up high-quality shots from his own matchups, the lineups with him at present always seems not effective. This can imply he is not a good team defender in terms of taking away his teammates troubles.
After examining the lineups, we use the probability estimation to evaluate the lineups shots, try to find what is vulnerable in the team defence.



(a) The opponents shot qualities of the Warriors lineup: Curry, Thompson, Barns, Green, Ezeli

(b) The opponents shot qualities of the Warriors lineup: Irving, Smith, James, Love, T.Thompson

Figure 9.4: The probability estimation comparison between the worst defensive lineup of the two teams.

As an example, we show the worst defensive lineup's shot probability estimation in Figure 9.4. The Warriors are having their center(one of the five positions in the basketball game) Ezeli in their lineup. As shown in Figure 9.4 (a), The protection near the basket is below the average, and the three-point line is also vulnerable compare to other lineups in the league. While, on the other hand, the Cavaliers lineup with T.Thompson and Irving, their defence close to the basket is a disaster, it seems almost no difference in terms of successful probability with or without a defender around there. Even in the mid-range, their opponents shoot a much higher percentage against them.

After a two hours session with our basketball coach and player, we have evaluated our tools. We asked the coach and player the experience of analyzing with our SQVA, what is valuable, and what is instead not to do. We also asked for suggestions for future work. The feedback is given that features are helpful and not seen in similar products. NBA teams can benefit from this analysis. However, the SQVA system can be improved in usability. It is found complex for the users to navigate by themselves. The offence/defence switch should be implemented in a better way. Currently, as soon as the user switches the mode, the visualizations also switch to the corresponding ones. The user might want to compare the defence performance with the offence performance. The tool is much easier to use if we give the user flexibility to switch offence and defence view for each view. In addition, the performance is still not satisfying. Some interactions are latent, giving confusion to the users and providing a bad user experience.

Chapter 10

Conclusions

10.1 Conclusions

To conclude this thesis, to overcome the problem of subjective shot quality in basketball. We build a visual analysis tool, namely SQVA. Which provide a method to evaluate the shot quality, one of the most important yet subjective factors in basketball games. Using the historical as input, we filter and enhance the data set to obtain the data set of shots with D_d and D_b as parameters. With the KNN probability density function, we create a probability estimation of a successful shot, then classify the shots and construct visual analytic with it.

We build a visual analytic with the shot quality estimations. We achieved our goal for the tools. We have three views in our tool, can be used for analyzing teams' lineups' and players' shot quality. All the views also support both offense view and defense view, allowing the users analyzing the team performance in both way. Furthermore, the SQVA system provides the comparison view, allowing users compare the lineup or player's shot quality. Users such as coaches can make improvement to their game strategies and plans, e.g. change the playing time and combinations of lineups base on the evaluations. From there, teams can potentially gain the competitive advantages by applying more efficient strategies .

We have one coach and one player from the local basketball club, Tantalus B.V. To have an evaluation session together, evaluate our tool and go through a real-world case, analyze the re-match of the final games 2015-2016 season. They have some interesting findings while experience our tool, and most of the result match their basketball knowledge. They also give us some valuable advises on improving the system

Based on the advises we received, we make a plan on how to improve the SQVA tool in the future. The navigation in the user interface is going to be designed better. We should provide more flexibility to the user and customize the interface for their own usage. We also plan to add more features to make the tool more lively and bring more information for each shot. E.g. show the animated position of all players for each shot and accessible from the probability estimation. We also consider to use different colors to distinguish the shot quality and success/miss, due to the fact people may have problem with the coloring we have now.

At the end, we want to generalize our project to make it applicable not only in NBA but also in every basketball league or even other sports league around the world. We hope our work can bring more insight for the sport expert, and maybe change the game we used to.

Bibliography

[1] Nba statics. 5

[2] Sportvu. 5

[3] Tantalus basktball. 28

[4] Wei Chen, Tianyi Lao, Jing Xia, Xinxin Huang, Biao Zhu, Wanqi Hu, and Huihua Guan. Gameflow: narrative visualization of nba basketball games. *IEEE Transactions on Multimedia*, 18(11):2247–2256, 2016. 3

[5] D. Daly-Grafstein and L. Bornn. Rao-blackwellizing field goal percentage. *Journal of Quantitative Analysis in Sports*, 15:85 – 95, 2018. 3

[6] BC Elliott and Elizabeth White. A kinematic and kinetic analysis of the female two point and three point jump shots in basketball. *The Australian Journal of Science and Medicine in Sport*, 21(2):7–11, 1989. 3

[7] Evelyn Fix and Joseph L Hodges Jr. Discriminatory analysis-nonparametric discrimination: Small sample performance. Technical report, CALIFORNIA UNIV BERKELEY, 1952. 16

[8] Thomas Gilovich, Robert Vallone, and Amos Tversky. The hot hand in basketball: On the misperception of random sequences. *Cognitive psychology*, 17(3):295–314, 1985. 3

[9] Jonathan J Koehler and Caryn A Conley. The "hot hand" myth in professional basketball. *Journal of sport and exercise psychology*, 25(2):253–259, 2003. 3

[10] Antonio G Losada, Roberto Therón, and Alejandro Benito. Bkviz: A basketball visual analysis tool. *IEEE computer graphics and applications*, 36(6):58–68, 2016. 3

[11] Tal Neiman and Yonatan Loewenstein. Reinforcement learning in professional basketball players. *Nature communications*, 2(1):1–8, 2011. 3

[12] Hiroki Okubo and Mont Hubbard. Defense for basketball field shots. *Procedia Engineering*, 34:730–735, 2012. 3

[13] G. Peskir. *Optimal Stopping and Free-Boundary Problems*. Optimal Stopping and Free-Boundary Problems, 2006. 3

[14] Petruccelli and D. Joseph. *Secretary Problem*. Encyclopedia of Statistical Sciences, 2004. 3

[15] C. Rasmussen and C. K. Williams. Gaussian processes for machine learning. In *Adaptive computation and machine learning*, 2009. 16

[16] Brian J Reich, James S Hodges, Bradley P Carlin, and Adam M Reich. A spatial analysis of basketball shot chart data. *The American Statistician*, 60(1):3–12, 2006. 3

[17] MN Satern. Kinematic parameters of basketball jump shots projected from varying distances. In *ISBS-Conference Proceedings Archive*, 1993. 3, 14

[18] Rajiv Shah. Merging nba play by play data with sportvu data. 6

[19] Brian Skinner. The problem of shot selection in basketball. *PloS one*, 7(1):e30776, 2012. 3

[20] Alan Stein. 7 keys to good shot selection. 13, 14

[21] Timothy R Vollmer and Jason Bourret. An application of the matching law to evaluate the allocation of two-and three-point shots by college basketball players. *Journal of Applied Behavior Analysis*, 33(2):137–150, 2000. 3

[22] Robert L Wardrop. Simpson's paradox and the hot hand in basketball. *The American Statistician*, 49(1):24–28, 1995. 3