CS170A Course Project

Bolun Hu

UID: 104056403

Professor D. S. Parker

NBA Game Schedule Analysis and Assessment

Motivation for this project

NBA is one of the most famous sport associations in the world, with thousands of millions of fans around the world, and I am one of the fanatics. That fact became especially true since the joining of Yao Ming into the Houston Rockets in 2002 (and as well Yi Jianlian joining the Milwaukee Bucks), and it was also the time I became obsessed with NBA computer games like NBA Live and NBA 2K series. At first I was only interested in the real game playing, and factors like graphics cards, CPU speeds, face patch, shoes patch were the most fascinating topics to me. However, when the dynasty mode was introduced in NBA Live 2004, which enables me to really manage a club, the issues relevant to data came into my mind: how to pick a good rookie among hundreds of candidates around the nation? How to design an efficient rotation plan so that not only my players can win the game but also have enough amount of recovering time? Is the season schedule really fair to all the teams? How much money should I pay for a player like Kevin Love or DeAndre Jordan in the free agent market? What is team chemistry? Can it be represented by data? All of these kinds of questions bothered me every time I played the game. And today I am really happy to be given such an opportunity to analyze an interesting topic that confuses but as well excites me for years.

Topics / Domain of Analysis

For gigantic organization like NBA, coming up with a complete, fair-to-all-team season schedule is really a huge and complex project, and the schedule would greatly affect the teams’ overall performance and records. And my topics for this paper is to use the method of mathematical modeling to analyze and assess the fairness of the schedule quantitatively. The season schedule I chose is 2008 – 2009, which is a little bit far from 2014, for the reason that it was the last season that Yao Ming, my idol, participated in. (After that season he was stricken by a serious ankle injury and unfortunately retired afterwards in 2011)

My analysis would be focusing on the following 3 main aspects:

1. Determine the factors or parameters that would be decisive or influential to the “rating” of a season schedule. And based on those factors, I would try to convert the schedule into digital format that could be analyzed mathematically and finally give out the mathematical expression which is valid for “rating” a certain schedule.
2. Based on the results in first part’s computation, I would try to analyze the “rating” (how good or bad) of the schedule of Houston Rockets for which Yao played. Also I would try to find out the team with most advantageous game schedule, and the team with the least advantageous schedule.
3. There is an interesting fact that confused me for quite a long time: After scanning through the schedule I found that every team in the NBA would play against other teams belonging to a different Conference for 2 times (e.g. LA Lakers would play against Boston Celtics for 2 times every season) and would play 4 times against teams within same Conference and as well the same Division (e.g. Golden States Warriors would play against LA Clippers 4 times every season because they both are in Pacific Division). But as for the teams that are in the same Conference but different Divisions, it would play 4 times against some of them, and 3 times against other ones (either 2 away, 1 at home or 2 at home, 1 away). Let’s call this situation as **3-or-4 policy**. With such policy, every team in the end would play 41 at home and 41 away, and teams within same Conference but different Divisions would all be in “balanced” state. (I.e. for teams in Pacific Division, say, LA Lakers, they would play the same number of games at home and away against teams in Northwest Division / Southwest Division. The same situation holds for all other teams.) *Based on this policy and schedule of 2008 – 2009 season, I would try to find out the algorithm for choosing teams that would play against each other for 3 times instead of 4. And I would try to assess the method to explore the possibility of improvement / optimization.*

Source of Data

All data associated with this project can be downloaded or observed from the following websites:

<http://www.hoopchina.com/teams/>

[www.basketball-reference.com](http://www.basketball-reference.com)

[www.sports.sina.com.cn/nba](http://www.sports.sina.com.cn/nba)

<http://espn.go.com/nba/statistics/player/_/stat/>

Also, the commands and scripts needed for this project, including extracting those data would be appended after this project. I will upload the initial raw data for your reference as well.

Analyzing the Topics and Questions raised

In order to design a of game schedule, there are multiple factors needed to consider: the time interval between different games, number of back to back games, miles traveled through out the season, strong / weakness of opponents, etc. Under such limits, the following 2 main challenges are what I plan to focus on:

1. Design programs to create a satisfying schedule according to the constraints
2. Given the schedule data, analyze its fairness to all teams, and make a detailed analyze on its effects on a particular team

First of all I would focus on the first two topics in which data is given. For the first two topics, I would try to deal with it using hierarchical approach. I can first find out the factors that would be influential to all teams, and then for every factor, I can analyze its effect on a team: is it a positive effect or negative effect? After that I would reason on a higher level: I would analyze the effects of different factors altogether on a team, and then try to compute out an appropriate ratio for each factor. Finally, based on the computation I can assess the “rating” of the schedule for each team, and find out the team that is most advantageous and team that is most “unlucky”.

For the detailed discussion of Houston Rockets, we can take step further based on our previous conclusions, by adding more specialized factors that may be discarded in the previous analysis, such as back to back games, the support rate of fans, etc. And I would analyze as detailed as possible to figure out the goodness or badness of the schedule from the perspective of Rockets, and draw a reasonable solution if possible.

Also for the final 3rd topic/question, I think it would be a good attempt to use greedy algorithm. First we can figure out the possible factors that may influence the choosing of 3 teams. After that I can digitalize those factors and build a mathematical model if possible. Finally I can compare the results with the given one, and formalize the solution for choosing which teams would play against each other for 3 times instead of 4.

Assumptions before Analysis

1. All teams would be on-time for all games, and there is no mid-season adjustment for any team.
2. When analyzing the effect of a single factor, I assume all other factors are mutually independent and thus would not affect each other.
3. Assume the strong / weakness of each team stays constant and could be represented by its winning percentage.

Notes on Symbols and Definitions

The following definitions are prepared for the later modeling. It might seem to be cumbersome. Skip this part if you want, because they would make more sense as you see the later reasoning.

*Aj* --------------data index for the time interval of team j，j=1，2，3，…30；

*k*1 --------------parameter representing the degree of effect of time interval；

*Bj* -------------- data index for the transition of arena for team j，j=1，2，3，…30；

*C j* -------------unified value for the power rating of opponents for team j，j=1，2，3，…30；

k2 --------------parameter representing the ratio of effect of opponents’ power；

*Wj* --------------index for the rate of advantageous for team j, j=1，2，3，…30；

*Q*1 --------------ratio representing the effect of time interval (0<Q<1) ；

*Q*2 -------------- ratio representing the effect of transition of arena (0<Q<1) ；

*Q*3 --------------ratio for the effect of opponents’ power (0<Q<1) ；

*Ei* --------------the difference of winning percentage between team I and team j；

*Pj* --------------representing team j’s winning percentage;

*Gi* --------------representing team i’s winning percentage against team j that is in the same Conference but different Division from team i；

*Li* --------------representing team j’s winning percentage against team i；

Build and Analysis of the Model

(a) Analysis of Model

Aiming at the negative and positive effect that a certain schedule has on different teams, I considered the following factors:

First factor to be considered is the time interval between two games. If the time interval is too short, players would feel exhausted and that would definitely downgrade their performance. On the other hand, if the time interval is too long, that would as well hinder player from playing high quality games, and thus affect the team’s overall performance. After some internet search and analysis, I think the effect of time interval should be major, and decide to choose it to be the main factor, with the ratio Q1.

Second factor is the changing of arenas (i.e. transition of home and away), it is easy to realize that if a team play multiple away games in a roll, it would cause players to travel continuously, and therefore affect their performance in a negative way. We set the ration of this factor to be Q2.

Finally is the effect of opponents’ power ranking. For instance, if Sacramento Kings play against LA Clippers today, Houston Rockets tomorrow, and Dallas Mavericks next Monday, such painful scheduling would greatly affect players’ mood, since a losing stream against opponents that are far stronger than you would hurt your confidence and give you much pressure which will restrict your future performance for sure. I set this ratio of this factor to be Q3.

Other factors including weather, foods, and transportation method would also objectively affected a team’s performance. But because the degree of those effect seems to be much smaller than what I have mentioned above, I will not include these into my model.

Therefore, I searched on different NBA data websites and gathered information about the factors I have mentioned, and computed out the ratio of effect for each of those factors. The following diagram shows the results of my computation:

Other 5%

Opponents’ power

20% 45% Time interval between games

30%

Change of arenas

(b) Building the Model

(1) Analyze each single factor’s influence on teams’ performance

* Model for Time Interval

Given the schedule, I analyzed the data with the help of PASCAL programming and Excel which helps me to coordinate and neaten the data. The resulting diagram for time interval is shown as follows:

(Program script for this project is appended after the paper)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Index | Team Name | Interval Time (days) | | | | | | |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 魔术 | 15 | 52 | 14 | 0 | 1 | 1 | 0 |
| 2 | 奇才 | 19 | 48 | 10 | 5 | 0 | 1 | 0 |
| 3 | 老鹰 | 23 | 41 | 11 | 7 | 0 | 1 | 0 |
| 4 | 山猫 | 22 | 40 | 17 | 3 | 0 | 1 | 0 |
| 5 | 热火 | 19 | 45 | 15 | 3 | 0 | 1 | 0 |
| 6 | 凯尔特人 | 19 | 48 | 13 | 1 | 1 | 0 | 1 |
| 7 | 猛龙 | 17 | 50 | 11 | 4 | 0 | 0 | 1 |
| 8 | 76 人 | 22 | 46 | 8 | 3 | 3 | 1 | 0 |
| 9 | 篮网 | 22 | 41 | 16 | 3 | 0 | 0 | 1 |
| 10 | 尼克斯 | 19 | 47 | 13 | 2 | 1 | 1 | 0 |
| 11 | 活塞 | 18 | 47 | 16 | 1 | 0 | 1 | 0 |
| 12 | 骑士 | 21 | 44 | 13 | 4 | 0 | 0 | 1 |
| 13 | 步行者 | 23 | 38 | 19 | 2 | 0 | 1 | 0 |
| 14 | 公牛 | 24 | 35 | 21 | 2 | 0 | 1 | 0 |
| 15 | 雄鹿 | 24 | 38 | 15 | 5 | 0 | 1 | 0 |
| 16 | 黄蜂 | 20 | 47 | 10 | 4 | 1 | 1 | 0 |
| 17 | 马刺 | 19 | 47 | 13 | 2 | 1 | 1 | 0 |
| 18 | 火箭 | 21 | 44 | 14 | 3 | 0 | 1 | 0 |
| 19 | 小牛 | 17 | 49 | 14 | 1 | 1 | 1 | 0 |
| 20 | 灰熊 | 23 | 40 | 14 | 5 | 0 | 1 | 0 |
| 21 | 爵士 | 21 | 41 | 18 | 2 | 0 | 1 | 0 |
| 22 | 掘金 | 22 | 45 | 9 | 5 | 1 | 0 | 1 |
| 23 | 开拓者 | 19 | 47 | 13 | 3 | 0 | 1 | 0 |
| 24 | 森林狼 | 23 | 37 | 21 | 1 | 0 | 0 | 1 |
| 25 | 超音速 | 19 | 45 | 16 | 2 | 0 | 1 | 0 |
| 26 | 湖人 | 20 | 46 | 12 | 3 | 1 | 1 | 0 |
| 27 | 太阳 | 20 | 46 | 13 | 1 | 2 | 1 | 0 |
| 28 | 勇士 | 16 | 52 | 11 | 3 | 0 | 1 | 0 |
| 29 | 国王 | 24 | 43 | 9 | 4 | 2 | 0 | 1 |
| 30 | 快船 | 22 | 44 | 11 | 4 | 0 | 2 | 0 |

Chart 1

Since the number of games for regular season is a constant, the total interval time (rest time) should also be constant. If the rest time is distributed evenly between every two games, then the team with that schedule should be the most advantageous team. In other words, the smaller the fluctuation of the time interval values is, the better the schedule is. Therefore, I use the concept of variance, and computed out the fluctuation of the time interval.

Based on the above data chart, I calculated out the variance for the time interval with help of Lingo Program (For the detailed program script, please check out appendix 2):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Index | Team | Variance | Data index | Index | Team | Variance | Data index |
| 1 | 魔术 | 0.647 | 517.6 | 16 | 黄蜂 | 0.785 | 628 |
| 2 | 奇才 | 0.943 | 754.4 | 17 | 马刺 | 0.799 | 639.2 |
| 3 | 老鹰 | 0.746 | 596.8 | 18 | 火箭 | 0.834 | 667.2 |
| 4 | 山猫 | 0.83 | 664 | 19 | 小牛 | 1.032 | 825.6 |
| 5 | 热火 | 0.908 | 726.4 | 20 | 灰熊 | 0.785 | 628 |
| 6 | 凯尔特人 | 0.651 | 520.8 | 21 | 爵士 | 0.894 | 715.2 |
| 7 | 猛龙 | 0.799 | 639.2 | 22 | 掘金 | 0.82 | 656 |
| 8 | 76 人 | 0.894 | 715.2 | 23 | 开拓者 | 0.859 | 687.2 |
| 9 | 篮网 | 0.785 | 628 | 24 | 森林狼 | 0.763 | 610.4 |
| 10 | 尼克斯 | 0.711 | 568.8 | 25 | 超音速 | 0.883 | 706.4 |
| 11 | 活塞 | 0.774 | 619.2 | 26 | 湖人 | 1.032 | 825.6 |
| 12 | 骑士 | 0.711 | 568.8 | 27 | 太阳 | 0.883 | 706.4 |
| 13 | 步行者 | 0.711 | 568.8 | 28 | 勇士 | 0.834 | 667.2 |
| 14 | 公牛 | 0.859 | 687.2 | 29 | 国王 | 0.661 | 528.8 |
| 15 | 雄鹿 | 1.12 | 896 | 30 | 快船 | 0.982 | 785.6 |

Chart 2

In order to unify the order of magnitude of each factor so that I can make comprehensive analysis, I decide to multiply all the variance value by a constant value K1=800, and got *Aj*.

* Model for Change of Arenas (Home/Away Transition)

According to different locations of different games, I roughly divide the situation into 4 main categories: (later I would build an automata based on this 4 categories)

**1. At Home**

**2. Away but in the same Division**

**3. Away in the same Conference but different Division**

**4. Away in different Conference and different Division.**

When transitioning between these 4 states, I give each of these transitions different values according to its geographical distance: the far the locations is from home arena, the worse this transition would be, and the larger the transition value of this transition would be. Namely:

1. Transition from Home to Away, within the same Division: Transition value = 4
2. Transition from Away to Home, within the same Division: Transition value = 2
3. Away to Away, within the same Conference, Transition value = 14
4. Home to Away, within the same Conference, different Division, Transition value = 6
5. Away to Home, within same Conference, different Division, Transition value = 5
6. Transition from away where is in the same Conference, but different Division, to another away where is in a different Conference, Transition value = 18
7. Transition from away where is in different Conference, to another away where is in the same Conference but different Division, Transition value = 18
8. Home to Away, different Conference, Transition value = 8
9. Away to Home, different Conference, Transition value = 7
10. Transition from away where away is in different Conference, to another away, where away is in the same Division, Transition value = 16
11. Transition from away where away is in the same Division, to another away where away is in a different Division, Transition value = 16

I used an automata to represent the transition as the following:

Different Conference ------------18----------🡪 Same Confer. Diff. Division

20 16

16

8 7 5 14 14

6

Home 2 Same Division

0 4 10

According to this transition diagram, I computed out the transition values *Bj* for all teams, data chart is shown as follows: (source code please refer to appendix 3)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| index | Team | Transition Value | index | Team | Transition Value |
| 1 | 魔术 | 562 | 16 | 黄蜂 | 557 |
| 2 | 奇才 | 552 | 17 | 马刺 | 567 |
| 3 | 老鹰 | 572 | 18 | 火箭 | 569 |
| 4 | 山猫 | 561 | 19 | 小牛 | 572 |
| 5 | 热火 | 564 | 20 | 灰熊 | 560 |
| 6 | 凯尔特人 | 569 | 21 | 爵士 | 565 |
| 7 | 猛龙 | 560 | 22 | 掘金 | 566 |
| 8 | 76 人 | 557 | 23 | 开拓者 | 590 |
| 9 | 篮网 | 540 | 24 | 森林狼 | 564 |
| 10 | 尼克斯 | 572 | 25 | 超音速 | 563 |
| 11 | 活塞 | 550 | 26 | 湖人 | 587 |
| 12 | 骑士 | 549 | 27 | 太阳 | 572 |
| 13 | 步行者 | 545 | 28 | 勇士 | 577 |
| 14 | 公牛 | 573 | 29 | 国王 | 569 |
| 15 | 雄鹿 | 560 | 30 | 快船 | 569 |

Chart 3

* Effect of Opponents’ Power

For any team, it would play 3 games with some opponents, and 4 games with other ones (i.e. 3-or-4 policy). So in such circumstance, there always exist some teams that would play more games with you than some other teams do. If luckily, those teams that play more times are weak teams (low power ranking/low winning percentage), then the schedule is a good one for your team right?

Therefore, using the Lingo program I calculated out the degree of effect that opponent’s power ranking has on each team. The resulting chart shows the values I got: (source code are appended in Appendix 3)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 球队 | result | unified | 球队 | result | unified |
| 魔术 | 2.437 | 487.44 | 黄蜂 | 2.527 | 505.47 |
| 奇才 | 2.797 | 559.4 | 马刺 | 2.215 | 443.07 |
| 老鹰 | 3.101 | 620.18 | 火箭 | 2.526 | 505.21 |
| 山猫 | 3.162 | 632.42 | 小牛 | 2.567 | 513.36 |
| 热火 | 4.471 | 894.11 | 灰熊 | 3.908 | 781.63 |
| 凯尔特人 | 2.215 | 443.07 | 爵士 | 2.667 | 533.32 |
| 猛龙 | 2.773 | 554.54 | 掘金 | 2.836 | 567.12 |
| 76 人 | 2.689 | 537.87 | 开拓者 | 3.017 | 603.45 |
| 篮网 | 2.944 | 588.86 | 森林狼 | 4.173 | 834.51 |
| 尼克斯 | 3.841 | 768.14 | 超音速 | 3.97 | 794.07 |
| 活塞 | 2.149 | 429.75 | 湖人 | 2.429 | 485.86 |
| 骑士 | 2.592 | 518.45 | 太阳 | 2.46 | 492.01 |
| 步行者 | 3.274 | 654.81 | 勇士 | 2.75 | 549.96 |
| 公牛 | 3.101 | 620.25 | 国王 | 3.058 | 611.67 |
| 雄鹿 | 3.453 | 690.7 | 快船 | 3.764 | 752.83 |

Chart 4

In order to unify the variable quantity, each resulting value is multiplied by a constant k2 = 200. And I got *C j.*

(2) Comprehensive Analysis of factors’ effects on teams’ performance

After analyzing each factor individually in part 1, the integrated equation for the overall effect on a team’s performance based on all 3 factors can be formalized as the following:

*Wj* = *Aj* \* k1 \* *Q*1 + *Bj* \* *Q*2 + *C j* \* k2 \* *Q*3

By using the Microsoft Excel, I computed out the final “rating” for each team’s schedule, and sorted them in the ascending order. Also it is worth noting that the smaller the W value is, the better the corresponding schedule is. Following is the resulting chart.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 排名 | 球队 | 最终利弊指标（w | ） 排名 | 球队 | 最终利弊指标（w） |
| 1 | 凯尔特人 | 493.674 | 16 | 太阳 | 587.882 |
| 2 | 魔术 | 499.008 | 17 | 山猫 | 593.584 |
| 3 | 骑士 | 524.35 | 18 | 76 人 | 596.514 |
| 4 | 活塞 | 529.59 | 19 | 爵士 | 598.004 |
| 5 | 国王 | 530.994 | 20 | 公牛 | 605.19 |
| 6 | 马刺 | 546.354 | 21 | 灰熊 | 606.926 |
| 7 | 步行者 | 550.422 | 22 | 开拓者 | 606.93 |
| 8 | 黄蜂 | 550.794 | 23 | 森林狼 | 610.782 |
| 9 | 篮网 | 562.372 | 24 | 奇才 | 616.96 |
| 10 | 老鹰 | 564.196 | 25 | 湖人 | 644.792 |
| 11 | 猛龙 | 566.548 | 26 | 超音速 | 645.594 |
| 12 | 火箭 | 571.982 | 27 | 小牛 | 645.792 |
| 13 | 掘金 | 578.424 | 28 | 快船 | 674.786 |
| 14 | 尼克斯 | 581.188 | 29 | 热火 | 674.902 |
| 15 | 勇士 | 583.332 | 30 | 雄鹿 | 709.34 |

Chart 5

From the above char, I can draw the claim that the most “lucky” team (with theoretically the most advantageous schedule) for the 2008-2009 season is the Boston Celtics, and the most “unlucky” team is the Milwaukee Bucks. Better luck next season Bucks.

Here comes my favorite part!

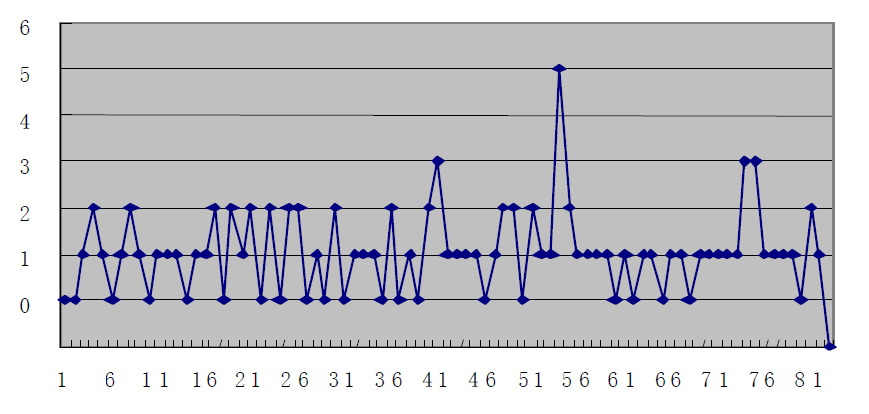
(c) Detailed Analysis of pros and cons of Houston Rockets’ schedule

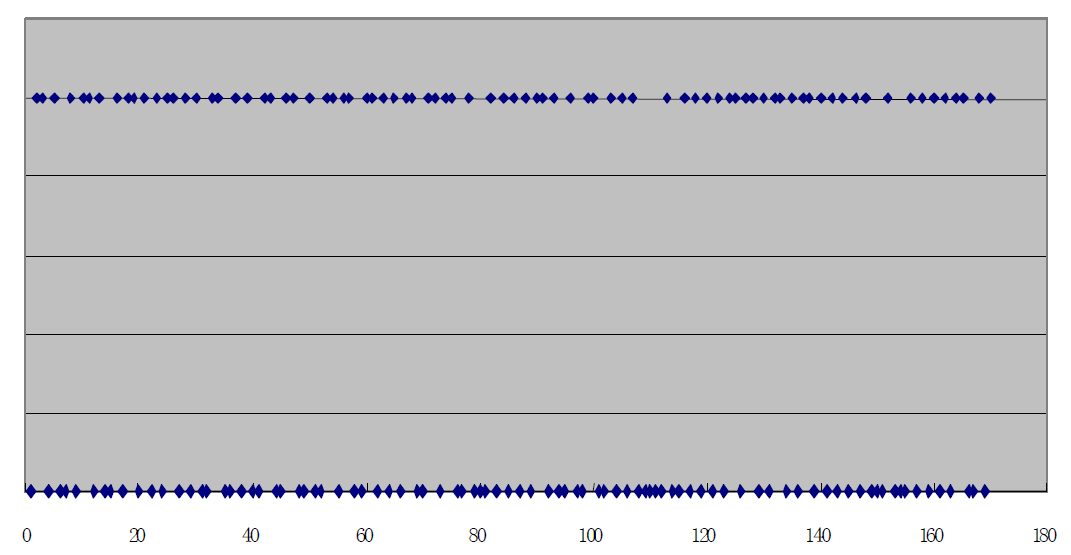
(1) Analyzing each factor individually

(a) Effect of time interval on Houston Rockets

By checking the variance chart above, I got the variance of time interval 0.834.

After data compilation, the broken line graph describing effect of time interval on Houston Rockets is drawn as follows:

 As we can see, the broken line fluctuates frequently, which indicates that the time interval is not distributed very uniformly. However the good news is, for the most part, the time interval fluctuates within a relatively small region, which indicates that the factor of time interval does not play a huge role in the overall effect. I also plotted out the scatter diagram for the games, as follows:



Note: the lower spotted line indicate the days without games. Upper spotted line indicates the opposite.

From the above scatter diagram we can see that the Rockets had a quite tough opening, which included a 6-game-in-9-days agenda. But the middle and later period is relatively easy compared to the beginning, and such schedule is beneficial for player’s recovery.

(2) The effect of transition between home game and away game on Rockets

From the schedule I found out that after in November after the game versus Boston Celtics, the Rockets went through a 5-game road trip, and the road trip was connected by a back-to-back home game versus the New Orleans Hornets (now Pelicans). This period of time was the toughest moment of the season. After that period, situation became a bit better in the mid-late season, since the Rockets had 23 home games. And this fact is especially true in February, during which Rockets had 9 home games out of 11 games in total, including a 6 home-game stream. Such home game stream is very beneficial for team’s recovery and facilitating team chemistry. After checking chart 3, I got the transition value for Rockets: 569, which is in the middle of all 30 teams.

(3) Effect of opponents’ power ranking on Rockets

In the first 18 games, there were 12 away games. More importantly, in those 12 away games, only 3 opponents were not among last season’s (07-08) playoffs. During the 5-game road trip, Rockets’ opponents are respectively: Portland Trail Blazers, LA Clippers, LA Lakers, Phoenix Suns, and San Antonio Spurs. After checking chart 4, I got the Rockets’ value C18 being 505.21, which is also around the middle of all teams.

(2) Comprehensive Analysis of Rockets’ schedule

After checking chart 5, I got the overall ranking of Rockets: 12, not a very top-notch rank. But considering its tough opening, it is a satisfying one in my opinion.

**Building Model for Topic No. 3 (3-or-4 policy)**

(a) Build of the Model

First let me remind you about what the topic is in case you already forget right now: Based on the schedule we can find that for teams that are in the same conference but different divisions, some teams will play against each other only 3 times, instead of the normal 4 times. My concern is to find out the method/model that could decide which teams will be playing with each other for only 3 times. I begin with the following 3 main aspects of analysis:

1. From the audience point of view

NBA is a league with strong commercial background. And from the stand point of being as much eye-catching as possible, the people scheduling the season definitely would take the audience perspective as a big factor into consideration. In other words, when people make schedule, they would take teams’ overall ability and power ranking into consideration. If two teams’ squads have similar overall ability and power ranking, then their difference would be minor, which means game between such teams would be much more competitive than those with larger power difference. Therefore, the task for finding teams that play 3 games against each other could be converted to the task for finding teams that have the largest power difference (pairing strong team and lesser team).

Therefore, to measure this type of effect, I used greedy algorithm. First