

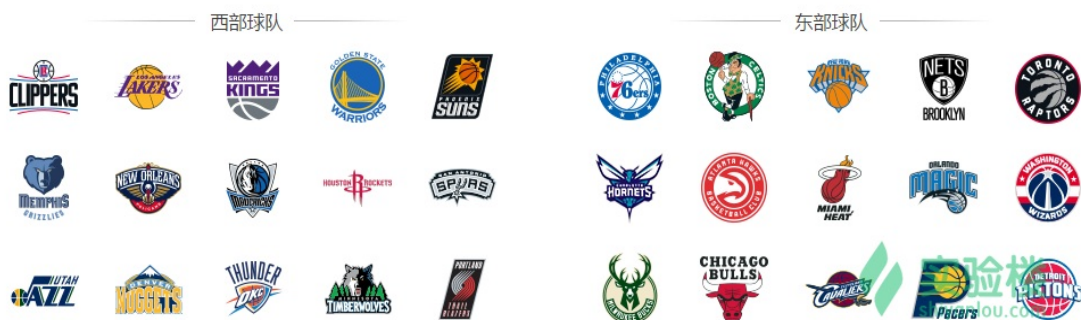
利用Python进行NBA比赛数据分析

一、介绍

1.1 内容简介

不知道你是否朋友圈被刷屏过nba的某场比赛进度或者结果？或者你就是一个nba狂热粉，比赛中的每个进球，抢断或是逆转压哨球都能让你热血沸腾。除去观赏精彩的比赛过程，我们也同样好奇比赛的结果会是如何。因此本节课程，将给同学们展示如何使用nba比赛的以往统计数据，判断每个球队的战斗力的，及预测某场比赛中的结果。

我们将基于2015-2016年的NBA常规赛及季后赛的比赛统计数据，预测在当下正在进行的2016-2017常规赛每场赛事的结果。



1.2 知识点

- nba球队的Elo score计算
- 特征向量
- 逻辑回归

1.3 环境

- python2.7
- pycharm

1.4 流程

我们将按照下面的流程实现NBA比赛数据分析的任务：

1. 获取比赛统计数据
2. 比赛数据分析，得到代表每场比赛每支队伍状态的特征表达
3. 利用机器学习方法学习每场比赛与胜利队伍的关系，并对2016-2017的比赛进行预测


二、获取 NBA比赛统计数据

2.1 比赛数据介绍

我们将采用Basketball Reference.com中的统计数据。在这个网站中，你可以看到不同球

员、队伍、赛季和联盟比赛的基本统计数据，如得分，犯规次数等情况，胜负次数等情况。而我们在这里将会使用2015-16 NBA Season Summary中数据。


Sports ReferenceBaseballFootball (college)Basketball (college)HockeySoccerBlogQuestions or Comments?Create AccountLogin



Enter Person, Team, Section, etc

Search

PlayersTeamsSeasonsLeadersScores1PlayoffsDraftPlay IndexFull Site Menu Below ▾



via Sports Logos.net
About logos

2015-16 NBA Season Summary

« 2014-15 Season Summary2016-17 Season Summary »

League Champion: [Cleveland Cavaliers](#)
Most Valuable Player: [Stephen Curry](#) (30.1/5.4/6.7)
Rookie of the Year: [Karl-Anthony Towns](#) (18.3/10.5/2.0)
PPG Leader: [Stephen Curry](#) (30.1)
RPG Leader: [Andre Drummond](#) (14.8)
APG Leader: [Rajon Rondo](#) (11.7)
WS Leader: [Stephen Curry](#) (17.9)

2015-16 NBA SeasonStandingsSchedule and ResultsLeadersPlayer Stats ▾Other ▾2016 Playoffs Summary

On this page:

Conference StandingsDivision StandingsPlayoff SeriesTeam Per Game StatsOpponent Per Game StatsTeam StatsOpponent StatsTeam Per Poss StatsTeam ShootingOpponent ShootingLeague AwardsPlayers of the Week & MonthLeague LeadersAll-NBAAll-DefensiveAll-RookieAll-Star Game RostersFull Site Menu

在这个2015-16总结的所有表格中，我们将使用的是以下三个数据表格：

- **Team Per Game Stats:** 每支队伍平均每场比赛的表现统计

Team Per Game Stats		* Playoff teams		Share & more ▾		Glossary																		
Rk	Team	G	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
1	Golden State Warriors*	82	242.4	42.5	87.3	.487	13.1	31.6	.416	29.4	55.7	.528	16.7	21.8	.763	10.0	36.2	46.2	28.9	8.4	6.1	15.2	20.7	114.9
2	Oklahoma City Thunder*	82	241.8	41.1	86.4	.476	8.3	23.7	.349	32.9	62.6	.524	19.7	25.2	.782	13.1	35.6	48.6	23.0	7.4	5.9	15.9	20.6	110.2
3	Sacramento Kings	82	241.5	40.0	86.4	.464	8.0	22.4	.359	32.0	64.0	.500	18.5	25.5	.725	10.6	33.7	44.2	24.5	8.9	4.5	16.2	20.4	106.6
4	Houston Rockets*	82	241.8	37.7	83.5	.452	10.7	30.9	.347	27.0	52.6	.514	20.4	29.4	.694	11.3	31.7	43.1	22.2	10.0	5.2	15.9	21.8	106.5
5	Boston Celtics*	82	241.2	39.2	89.2	.439	8.7	26.1	.335	30.5	63.1	.483	18.5	23.5	.788	11.6	33.3	44.9	24.2	9.2	4.2	13.7	21.9	105.7
6	Portland Trail Blazers*	82	241.5	38.6	85.9	.450	10.5	28.5	.370	28.1	57.4	.490	17.4	23.0	.754	11.6	33.9	45.5	21.3	6.9	4.6	14.6	21.7	105.1
7	Los Angeles Clippers*	82	241.8	38.3	82.4	.465	9.7	26.7	.364	28.6	55.7	.513	18.2	26.2	.692	8.8	33.3	42.0	22.8	8.6	5.6	13.0	21.3	104.5
8	Cleveland Cavaliers*	82	242.1	38.7	84.0	.460	10.7	29.6	.362	27.9	54.4	.514	16.3	21.7	.748	10.6	33.9	44.5	22.7	6.7	3.9	13.6	20.3	104.3
9	Washington Wizards	82	240.9	39.5	85.8	.460	8.6	24.2	.358	30.8	61.6	.501	16.5	22.5	.730	9.1	32.8	41.8	24.5	8.6	3.9	14.5	20.8	104.1
10	San Antonio Spurs*	82	240.3	40.1	82.9	.484	7.0	18.5	.375	33.2	64.4	.515	16.4	20.4	.803	9.4	34.5	43.9	24.5	8.3	5.9	13.1	17.5	103.5
11	Charlotte Hornets*	82	242.1	37.0	84.4	.439	10.6	29.4	.362	26.4	55.0	.479	18.7	23.7	.790	9.0	35.0	43.9	21.7	7.3	5.3	12.6	18.1	103.4
12	Atlanta Hawks*	82	241.8	38.6	84.4	.458	9.9	28.4	.350	28.7	56.1	.512	15.6	20.0	.783	8.3	33.8	42.1	25.6	9.1	5.9	15.0	19.1	102.8

数据名	含义
Rk -- Rank	排名
G -- Games	参与的比赛场数（都为82场）
MP -- Minutes Played	平均每场比赛进行的时间
FG--Field Goals	投球命中次数
FGA--Field Goal Attempts	投射次数
FG%--Field Goal Percentage	投球命中次数
3P--3-Point Field Goals	三分球命中次数
3PA--3-Point Field Goal Attempts	三分球投射次数
3P%--3-Point Field Goal Percentage	三分球命中率
2P--2-Point Field Goals	二点球命中次数
2PA--2-point Field Goal Attempts	二点球投射次数
2P%--2-Point Field Goal Percentage	二点球命中率
FT--Free Throws	罚球命中次数
FTA--Free Throw Attempts	罚球投射次数
FT%--Free Throw Percentage	罚球命中率
ORB--Offensive Rebounds	进攻篮板球
DRB--Defensive Rebounds	防守篮板球
TRB--Total Rebounds	篮板球总数
AST--Assists	助攻
STL--Steals	抢断
BLK -- Blocks	封盖
TOV -- Turnovers	失误
PF -- Personal Fouls	个犯
PTS -- Points	得分

- Opponent Per Game Stats: 所遇到的对手平均每场比赛的统计信息，所包含的统计数据与Team Per Game Stats中的一致，只是代表的该球队对应的对手的

Opponent Per Game Stats [Share & more](#) [Glossary](#)

Rk	Team	G	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
1	San Antonio Spurs*	82	240.3	35.7	81.8	.436	6.6	19.9	.331	29.1	61.9	.470	14.9	19.6	.758	9.1	31.4	40.5	20.8	7.2	3.9	14.8	19.5	92.9
2	Utah Jazz	82	243.4	35.6	79.9	.446	7.9	22.2	.357	27.7	57.7	.480	16.8	22.5	.746	9.3	30.8	40.1	19.1	8.0	4.7	14.0	19.9	95.9
3	Toronto Raptors*	82	241.2	36.5	82.1	.444	8.7	23.4	.373	27.7	58.7	.473	16.5	22.1	.748	9.5	31.2	40.8	21.7	6.5	5.4	13.3	22.0	98.2
4	Cleveland Cavaliers*	82	242.1	36.8	82.1	.448	7.9	22.7	.347	28.9	59.4	.487	16.8	22.6	.743	9.3	31.8	41.0	21.4	7.2	4.4	13.3	20.6	98.3
5	Miami Heat*	82	241.8	37.2	84.3	.442	7.4	21.2	.347	29.9	63.1	.474	16.5	21.5	.770	9.8	31.5	41.3	20.2	7.5	4.1	12.9	19.6	98.4
6	Atlanta Hawks*	82	241.8	37.1	86.1	.432	8.3	24.5	.338	28.9	61.6	.469	16.7	22.1	.755	11.5	35.0	46.5	22.0	8.6	5.0	16.1	18.3	99.2
7	Los Angeles Clippers*	82	241.8	36.8	84.7	.434	7.9	23.3	.338	28.9	61.4	.471	18.8	25.1	.751	11.8	34.9	46.7	21.2	7.1	3.2	15.4	22.5	100.2
8	Indiana Pacers*	82	242.4	37.4	84.9	.440	8.3	24.8	.334	29.1	60.1	.484	17.4	23.2	.751	10.7	33.8	44.5	20.8	7.7	4.5	15.8	20.4	100.5
9	Charlotte Hornets*	82	242.1	37.8	85.0	.444	8.9	25.4	.349	28.9	59.6	.485	16.3	21.2	.769	8.9	35.8	44.7	23.2	6.7	5.5	13.5	20.4	100.7
10	New York Knicks	82	241.5	38.0	85.8	.443	7.6	22.4	.341	30.3	63.4	.479	17.5	23.2	.754	10.9	33.3	44.2	20.8	7.2	4.2	11.3	18.5	101.1
11	Memphis Grizzlies*	82	241.8	35.9	78.8	.456	9.7	26.6	.365	26.2	52.1	.503	19.7	25.7	.768	10.1	33.1	43.2	21.9	7.0	5.7	16.2	21.1	101.3
12	Detroit Pistons*	82	242.4	38.8	84.2	.461	7.3	20.5	.355	31.5	63.7	.495	16.5	21.0	.783	8.8	33.7	42.5	21.5	7.1	4.5	13.4	21.6	101.4

- Miscellaneous Stats: 综合统计数据

Miscellaneous Stats [Share & more](#) [Glossary](#)

Rk	Team	Age	Offense Four Factors										Defense Four Factors										Attendance		
			W	L	PW	PL	MOV	SOS	SRS	ORtg	DRtg	Pace	FTtr	3PAR	TS%	eFG%	TOV%	ORB%	FT/FGA	eFG%	TOV%	DRB%		FT/FGA	Arena
1	Golden State Warriors*	27.4	73	9	65	17	10.76	-0.38	10.38	114.5	103.8	99.3	.250	.362	.593	.563	13.5	23.5	.191	.479	12.6	76.0	.208	Oracle Arena	803,43
2	San Antonio Spurs*	30.3	67	15	67	15	10.63	-0.36	10.28	110.3	99.0	93.8	.246	.223	.564	.526	12.4	23.0	.197	.477	14.1	79.1	.182	AT&T Center	756,44
3	Oklahoma City Thunder*	25.8	55	27	59	23	7.28	-0.19	7.09	113.1	105.6	96.7	.292	.275	.565	.524	14.0	31.1	.228	.484	11.7	76.0	.205	Chesapeake Energy Arena	746,32
4	Cleveland Cavaliers*	28.1	57	25	57	25	6.00	-0.55	5.45	110.9	104.5	93.3	.259	.352	.558	.524	12.7	25.1	.194	.496	12.6	78.5	.205	Quicken Loans Arena	843,04
5	Los Angeles Clippers*	29.7	53	29	53	29	4.28	-0.15	4.13	108.3	103.8	95.8	.318	.324	.556	.524	12.1	20.1	.220	.480	13.8	73.8	.222	STAPLES Center	786,91
6	Toronto Raptors*	26.3	56	26	53	29	4.50	-0.42	4.08	110.0	105.2	92.9	.328	.287	.552	.504	12.3	24.6	.255	.498	12.7	77.7	.201	Air Canada Centre	812,86
7	Atlanta Hawks*	28.2	48	34	51	31	3.61	-0.12	3.49	105.1	101.4	97.1	.237	.336	.552	.516	13.8	19.1	.185	.480	14.4	74.6	.194	Phillips Arena	690,15
8	Boston Celtics*	25.2	48	34	50	32	3.21	-0.37	2.84	106.8	103.6	98.5	.264	.293	.531	.488	12.1	25.1	.208	.487	14.6	74.6	.231	TD Garden	749,07
9	Charlotte Hornets*	26.0	48	34	49	33	2.72	-0.36	2.36	107.1	104.3	95.7	.280	.348	.545	.502	11.7	20.0	.222	.496	12.5	79.8	.191	Time Warner Cable Arena	716,89
10	Utah Jazz	24.2	40	42	46	36	1.79	0.05	1.84	105.9	103.9	91.0	.286	.297	.540	.501	14.2	25.9	.213	.495	13.5	77.7	.210	Vivint Smart Home Arena	791,48
11	Indiana Pacers*	26.9	45	37	46	36	1.71	-0.09	1.62	104.6	102.9	96.6	.268	.270	.536	.497	13.5	23.4	.205	.489	14.3	76.0	.205	Bankers Life Fieldhouse	690,73
12	Miami Heat*	28.4	48	34	46	36	1.65	-0.14	1.50	106.1	104.4	93.6	.282	.221	.545	.508	13.3	23.8	.210	.485	12.1	77.8	.196	AmericanAirlines Arena	809,35

数据项	数据含义
Rk (Rank)	排名
Age	队员的平均年龄
W (Wins)	胜利次数
L (Losses)	失败次数
PW (Pythagorean wins)	基于毕达哥拉斯理论计算的赢的概率
PL (Pythagorean losses)	基于毕达哥拉斯理论计算的输的概率
MOV (Margin of Victory)	赢球次数的平均间隔
SOS (Strength of Schedule)	用以评判对手选择与其球队或是其他球队的难易程度对比，0为平均线，可以为正负数
SRS (Simple Rating System)	3
ORTg (Offensive Rating)	每100个比赛回合中的进攻比例
DRtg (Defensive Rating)	每100个比赛回合中的防守比例
Pace (Pace Factor)	每48分钟内大概会进行多少个回合
FTr (Free Throw Attempt Rate)	罚球次数所占投射次数的比例
3PAr (3-Point Attempt Rate)	三分球投射占投射次数的比例
TS% (True Shooting Percentage)	二分之一球、三分球和罚球的总命中率
eFG% (Effective Field Goal Percentage)	有效的投射百分比（含二分之一球、三分球）
TOV% (Turnover Percentage)	每100场比赛中失误的比例
ORB% (Offensive Rebound Percentage)	球队中平均每个人的进攻篮板的比例
FT/FGA	罚球所占投射的比例
eFG% (Opponent Effective Field Goal Percentage)	对手投射命中比例
TOV% (Opponent Turnover Percentage)	对手的失误比例
DRB% (Defensive Rebound Percentage)	球队平均每个球员的防守篮板比例
FT/FGA (Opponent Free Throws Per Field Goal Attempt)	对手的罚球次数占投射次数的比例

毕达哥拉斯定律

$$win\% = \frac{runs\ scored^2}{runs\ scored^2 + runs\ allowed^2}$$

我们将用这三个表格来评估球队过去的战斗力，另外还需2015-16 NBA Schedule and Results中的2015~2016年的nba常规赛及季后赛的每场比赛的比赛数据，用以评估Elo score（在之后解释）。在Basketball Reference.com中按照从常规赛至季后赛的时间。列出了2015年10月份至2016年6月份的每场比赛的比赛情况。

October Schedule

Share & more ▼

Date	Start (ET)	Visitor/Neutral	PTS	Home/Neutral	PTS		Notes
Tue, Oct 27, 2015	8:00 pm	Detroit Pistons	106	Atlanta Hawks	94	Box Score	
Tue, Oct 27, 2015	8:00 pm	Cleveland Cavaliers	95	Chicago Bulls	97	Box Score	
Tue, Oct 27, 2015	10:30 pm	New Orleans Pelicans	95	Golden State Warriors	111	Box Score	
Wed, Oct 28, 2015	7:30 pm	Philadelphia 76ers	95	Boston Celtics	112	Box Score	
Wed, Oct 28, 2015	7:30 pm	Chicago Bulls	115	Brooklyn Nets	100	Box Score	
Wed, Oct 28, 2015	7:30 pm	Utah Jazz	87	Detroit Pistons	92	Box Score	
Wed, Oct 28, 2015	8:00 pm	Denver Nuggets	105	Houston Rockets	85	Box Score	
Wed, Oct 28, 2015	10:30 pm	Minnesota Timberwolves	112	Los Angeles Lakers	111	Box Score	
Wed, Oct 28, 2015	8:00 pm	Cleveland Cavaliers	106	Memphis Grizzlies	76	Box Score	
Wed, Oct 28, 2015	7:30 pm	Charlotte Hornets	94	Miami Heat	104	Box Score	
Wed, Oct 28, 2015	8:00 pm	New York Knicks	122	Milwaukee Bucks	97	Box Score	
Wed, Oct 28, 2015	8:00 pm	San Antonio Spurs	106	Oklahoma City Thunder	112	Box Score	
Wed, Oct 28, 2015	7:00 pm	Washington Wizards	88	Orlando Magic	87	Box Score	
Wed, Oct 28, 2015	10:00 pm	Dallas Mavericks	111	Phoenix Suns	95	Box Score	
Wed, Oct 28, 2015	10:00 pm	New Orleans Pelicans	94	Portland Trail Blazers	112	Box Score	

可在上图中，看到2015年10月份的部分比赛数据。在每个 *Schedule* 表格中所包含的数据为：

数据项	数据含义
Date	比赛日期
Start (ET)	比赛开始时间
Visitor/Neutral	客场作战队伍
PTS	客场队伍最后得分
Home/Neutral	主场队伍
PTS	主场队伍最后得分
Notes	备注，表明是否为加时赛等

在预测时，我们同样也需要在2016-17 NBA Schedule and Results中2016~2017年的NBA的常规赛比赛安排数据。

2.2 获取比赛数据

我们将以获取Team Per Game Stats表格数据为例，展示如何获取这三项统计数据。

1. 进入到用[Basketball Reference.com](#)中，在导航栏中选择Season并选择2015~2016赛季中的Summary：

Players Teams Seasons Leaders Scores ¹ Playoffs Draft

2016-17: [Summary](#) · [Leaders](#) · [Schedule & Results](#) · [Stats](#) ([Per G](#) & [Totals](#) & [Per 36 Min](#) & [Per Poss](#) & [Advanced](#)) ...


2015-16: [Summary](#) · [Leaders](#) · [Schedule & Results](#) · [Stats](#) ([Per G](#) & [Totals](#) & [Per 36 Min](#) & [Per Poss](#) & [Advanced](#)) ...

2014-15: [Summary](#) · [Leaders](#) · [Schedule & Results](#) · [Stats](#) ([Per G](#) & [Totals](#) & [Per 36 Min](#) & [Per Poss](#) & [Advanced](#)) ...

2013-14: [Summary](#) · [Leaders](#) · [Schedule & Results](#) · [Stats](#) ([Per G](#) & [Totals](#) & [Per 36 Min](#) & [Per Poss](#) & [Advanced](#)) ...

2012-13: [Summary](#) · [Leaders](#) · [Schedule & Results](#) · [Stats](#) ([Per G](#) & [Totals](#) & [Per 36 Min](#) & [Per Poss](#) & [Advanced](#)) ...

» [All NBA & ABA Seasons](#) · [D-League Seasons](#) · [NBL Seasons](#) · [Europe Seasons](#) · [WNBA Seasons](#)

	TOR* (3) F C \$ 51 31 WAS* (4) F C \$ 49 33 ATL* (5) F C \$ 43 39 MIL* (6) F C \$ 42 40	HOU* (3) F C \$ 55 27 LAC* (4) F C \$ 51 31 UTA* (5) F C \$ 51 31 OKC* (6) F C \$ 47 35
---	--	--

2. 进入到2015~2016年的Summary界面后，滑动窗口找到Team Per Game Stats表格，并选择左上方的Share & more，在其下拉菜单中选择Get table as CSV (for Excel)：

2015-16 NBA Season | Standings | Schedule and Results | Leaders | Player Stats ▼ | Other ▼ | 2016 Playoffs Summary | Back to top ▲

Team Per Game Stats * Playoff teams

Rk	Team	G	MP	FG	FGA	FG%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
1	Golden State Warriors*	82	242.4	42.5	87.3	.484	7.0	18.5	.375	33.2	64.4	.515	16.4	20.4	.803	9.4	34.5	114.9
2	Oklahoma City Thunder*	82	241.8	41.1	86.4	.477	6.9	18.0	.383	33.0	63.9	.512	16.3	20.3	.799	9.3	34.4	114.7
3	Sacramento Kings	82	241.5	40.0	86.4	.464	6.8	17.9	.380	32.8	63.6	.511	16.2	20.2	.798	9.2	34.3	114.6
4	Houston Rockets*	82	241.8	37.7	83.5	.451	6.7	17.8	.377	32.7	63.5	.510	16.1	20.1	.797	9.1	34.2	114.5
5	Boston Celtics*	82	241.2	39.2	89.2	.439	6.6	17.7	.373	32.6	63.4	.509	16.0	20.0	.796	9.0	34.1	114.4
6	Portland Trail Blazers*	82	241.5	38.6	85.9	.449	6.5	17.6	.370	32.5	63.3	.508	15.9	19.9	.795	8.9	34.0	114.3
7	Los Angeles Clippers*	82	241.8	38.3	82.4	.464	6.4	17.5	.369	32.4	63.2	.507	15.8	19.8	.794	8.8	33.9	114.2
8	Cleveland Cavaliers*	82	242.1	38.7	84.0	.461	6.3	17.4	.366	32.3	63.1	.506	15.7	19.7	.793	8.7	33.8	114.1
9	Washington Wizards	82	240.9	39.5	85.8	.460	6.2	17.3	.365	32.2	63.0	.505	15.6	19.6	.792	8.6	33.7	114.0
10	San Antonio Spurs*	82	240.3	40.1	82.9	.484	7.0	18.5	.375	33.2	64.4	.515	16.4	20.4	.803	9.4	34.5	114.9
11	Charlotte Hornets*	82	242.1	37.0	84.4	.439	6.6	17.7	.373	32.6	63.4	.509	16.0	20.0	.796	9.0	34.1	114.4
12	Atlanta Hawks*	82	241.8	38.6	84.4	.458	6.5	17.6	.370	32.5	63.3	.508	15.9	19.9	.795	8.9	34.0	114.3
13	New Orleans Pelicans	82	241.2	38.5	85.9	.448	6.4	17.5	.369	32.4	63.2	.507	15.8	19.8	.794	8.8	33.9	114.2

Share & more ▲ | Glossary

Modify & Share Table

Embed this
 Export table as suitable for use with excel (xl)

Get as Excel

Get table as CSV (for Excel)

Strip Mobile Formatting

Copy Link to Table to Clipboard

About Sharing Tools

Video: SR Sharing Tools & How-to

Video: Stats Table Tips & Tricks

3. 复制在界面中生成的csv格式数据，并复制粘贴至一个文本编辑器保存为csv文件即可。

三、数据分析

在获取到数据之后，我们将利用每支队伍过去的比赛情况和Elo 等级分来判断每支比赛队伍的可胜概率。在评价到每支队伍过去的比赛情况时，我们将使用到Team Per Game Stats, Opponent Per Game Stats和Miscellaneous Stats（之后简称为T、O和M表）这三个表格的数据，作为代表比赛中某支队伍的比赛特征。我们最终将实现针对每场比赛，预测比赛中哪支队伍最终将会获胜，但并不是给出绝对的胜败情况，而是预判胜利的队伍有多大的获胜概率。因此我们将建立一个代表比赛的特征向量。由两支队伍以往的以往比赛情况统计情况（T、O和M表），和两个队伍各自的Elo等级分构成。

关于Elo score等级分，不知道同学们是否看过《社交网络》这部电影，在这部电影中，Mark（主人公原型就是扎克伯格，FaceBook创始人）在电影起初开发的一个美女排名系统就是利用其好友Eduardo在窗户上写下的排名公式，对不同的女生进行等级制度对比，最后PK出胜利的一方。



这条对比公式就是Elo Score等级分制度。Elo的最初为了提供国际象棋中，更好地对不同的选手进行等级划分。在现在很多的竞技运动或者游戏中都会采取Elo等级分制度对选手或玩家进行等级划分，如足球、篮球、棒球比赛或LOL，DOTA等游戏。

在这里我们将基于国际象棋比赛，大致地介绍下Elo等级划分制度。在上图中Eduardo在窗户上写下的公式就是根据Logistic Distribution计算PK双方（A和B）对各自的胜率期望值计算公式。假设A和B的当前等级分为 R_A 和 R_B ，则A对B的胜率期望值为：

$$E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}}$$

B对A的胜率期望值为：

$$E_B = \frac{1}{1 + 10^{(R_A - R_B)/400}}$$

如果棋手A在比赛中的真实得分 S_A （胜1分，和0.5分，负0分）和他的胜率期望值 E_A 不同，则他的等级分要根据以下公式进行调整：

$$R_A^{new} = R_A^{old} + K(S_A - E_A)$$

在国际象棋中，根据等级分不同K值也会做相应的调整：

- ≥ 2400 , $K=16$
- $2100 \sim 2400$ 分, $K=24$
- ≤ 2100 , $K=32$

因此我们将会用以表示某场比赛数据的特征向量为（加入A与B队比赛）：[A队Elo score, A队的T,O和M表统计数据, B队Elo score, B队的T,O和M表统计数据]

四、基于数据进行模型训练和预测

4.1 开发环境准备

我们将会使用到python的pandas, numpy, scipy和sklearn库，不过实验楼中已经安装了numpy，所以在实验前，我们需要先利用pip命令安装另外三个Python库。

```
$ pip install pandas
```

```
$ pip install scipy
```

```
$ pip install sklearn
```

```
$ pip install cookiecutter
```

在安装完所需的实验库之后，初始化工程

```
cookiecutter https://github.com/audreyr/cookiecutter-pypackage.git
```

根据提示输入相关信息，我们把工程命名为nba_prediction。然后将我们下载的数据文件保存到nba_prediction/data下面。初始化完成后的工程目录如下：

```
├── AUTHORS.rst
├── CONTRIBUTING.rst
├── HISTORY.rst
├── LICENSE
├── MANIFEST.in
├── Makefile
├── README.rst
├── docs
│   ├── Makefile
│   ├── authors.rst
│   ├── conf.py
│   ├── contributing.rst
│   ├── history.rst
│   ├── index.rst
│   ├── installation.rst
│   ├── make.bat
│   ├── readme.rst
│   └── usage.rst
├── nba_prediction
│   ├── __init__.py
│   └── cli.py
```



```
| |— data
| | |— 15-16Miscellaneous_Stat.csv
| | |— 15-16Opponent_Per_Game_Stat.csv
| | |— 15-16Team_Per_Game_Stat.csv
| | |— 16-17Schedule.csv
| |— nba_prediction.py
|— requirements_dev.txt
|— setup.cfg
|— setup.py
|— tests
| |— __init__.py
| |— test_nba_prediction.py
|— tox.ini
|— travis_pypi_setup.py
```

在data文件夹中，包含了2015~2016年的NBA数据T,O和M表，及经处理后的常规赛和挑战赛的比賽数据2015~16result.csv，这个数据文件是我们通过在basketball-reference.com的2015-16 Schedule and result的几个月份比赛数据中提取得到的，其中包括三个字段：

- WTeam: 比赛胜利队伍
- LTeam: 失败队伍
- WLoc: 胜利队伍一方所在的为主场或是客场 另外一个文件就是16-17Schedule.csv，也是经过我们加工处理得到的NBA在2016~2017年的常规赛的比賽安排，其中包括两个字段：
- Vteam: 访问/客场作战队伍
- Hteam: 主场作战队伍

4.2 代码实现

建具体工程代码

五、总结

我们利用Basketball-reference.com的部分统计数据，计算每支nba比赛队伍的Elo socre，和利用这些基本统计数据评价每支队伍过去的比賽情况，并且根据国际等级划分方法Elo Score对队伍现在的战斗等级进行评分，最终结合这些不同队伍的特征判断在一场比赛中，哪支队伍能够占到优势。但在我们的预测结果中，我们没有给出绝对的正负之分，而是给出胜算较大一方的

队伍能够赢另外一方的概率。当然在这里，我们所采用评价一支队伍性能的数据量还太少（只采用了15~16年一年的数据），如果想要更加准确、系统的判断，有兴趣的你当然可以从各种统计数据网站中获取到更多年份，更加全面的数据。结合不同的回归、决策机器学习模型，搭建一个更加全面，预测准确率更高的模型。在[kaggle](#)中有相关的篮球预测比赛项目，有兴趣的同学可尝试一下。

六、参考阅读

- [知乎：在哪能看到最全面细致的NBA数据统计](#)
- [How I won my NCAA tournament bracket pool using machine learning](#)

七、思考

我们只是利用了scikit-learn提供的Logisitic Regression方法进行回归模型的训练，你可否尝试scikit-learn中的其他机器学习方法，或者其他类似于TensorFlow的开源框架，结合我们所提供的数据集进行训练。

