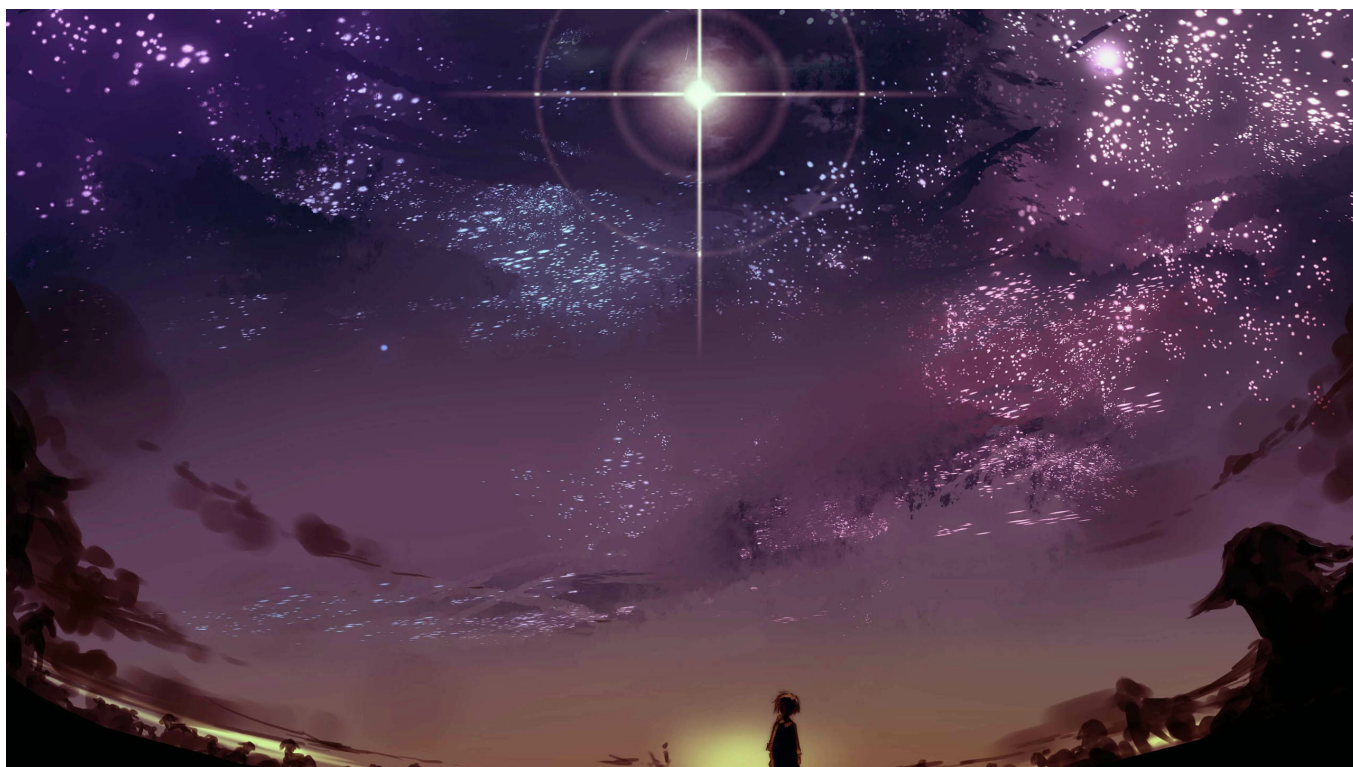


利用统计学习方法预测NBA比赛结果--软件设计

bbbb实用软件工程期末

- **Author** : 李英民 | **Henry**
- **E-mail**: li_yingmin@outlook dot com
- *Beijing U of Posts and Telecom(BUPT)*
- **Home**: <https://liyingmin.wixsite.com/henry>

- 快速了解我: [About Me](#)
- 转载请保留上述引用内容，谢谢配合！



项目Github地址: https://github.com/Li-Yingmin/nba_prediction

一、需求分析


1、市场分析

在百度搜索 **NBA预测** 关键字，出现如下搜索结果：

[今日NBA比赛结果智能预测 让分胜负预测 新浪彩通](#)

新浪彩通**NBA**智能结果**预测**采用大数据和模型算法,对今天进行的**NBA**赛事的胜负进行分析,提供基于基本面、赔率面和数据面的全面参考。

[ai.lottery.sina.com.cn...](#) - 百度快照

 为您推荐: [今日竞彩nba预测推荐](#) [nba预测推荐和比赛分析](#) [竞彩篮球专家每天推荐](#)

[【竞彩篮球预测】篮彩预测 篮球彩票预测 美职男篮篮彩预测-新浪爱彩](#)

竞彩篮球 | 全部 分析**预测** | 焦点赛事 | 中奖新闻 | 分析**预测**平安夜快乐,小姨妈总结**NBA**前俩月数据西部的火箭本赛季打的异常好,而勇士近期逐步回勇高居第一,马刺倒是...

[zx.aicai.com/cpzx/1_50...](#) - 百度快照

[特辑:2016-17赛季预测之得分后卫排行榜 NBA中国官方网站](#)

2016年10月21日 - 特辑:2016-17赛季**预测**之得分后卫排行榜首页图集**NBA**官网Moon2016-10-21 14:16 0 新赛季即将开始,在赛季开始之前,**NBA**官网依据球员数据和个人表现给出了...

[china.nba.com/a/201610...](#) - 百度快照

[竞彩篮球推荐,NBA预测,篮彩比赛分析 - 一定牛彩票网](#)

一定彩票网提供最新竞彩篮球推荐、**NBA预测**、篮彩比赛分析,帮助彩民购买竞彩篮球中大奖。

<https://www.ydniu.com/info/jcl...> - 百度快照

[NBA预测的最新相关信息](#)

[预测NBA还得看于嘉!篮彩再中冷门 7连中赚到手软](#)



迷尔**NBA**连中3串1+2串1,近6场比赛全中,近11中10。接下来,央视名嘴杨健、于嘉,前CBA总冠军李克还会对明日的比赛做出分析**预测**,感兴趣的朋友千万不要错...

其中，**80%** 以上的搜索结果都是关于预测彩票的。而且，关于**NBA**预测彩票的市场已经非常成熟，并且有非常多的专家参与彩票预测中来：

NBA预测的最新相关信息

预测NBA还得看于嘉!篮彩再中冷门 7连中赚到手软



迷尔NBA连中3串1+2串1,近6场比赛全中,近11中10。接下来,央视名嘴杨健、于嘉,前CBA总冠军李克还会对明日的比赛做出分析预测,感兴趣的朋友千万不要错...

网易 12月20日

因此，NBA预测彩票市场的特点如下：

- 有专家参与
- 用户数量非常庞大
- 市场成熟并且稳定
- 彩票厂商足够多

2、需求分析

通过如上的市场分析，我们可以从以下几个需求入手：

- 针对专家，可能想知道基于统计的预测结果作为判断依据
- 针对普通用户，预测结果作为购买彩票的参考
- 针对彩票厂商，提供预测技术支持，以作为营销的新方法

3、应用场景

考虑到专家和用户之间的利益关系，我们可以设计以下应用场景：

- 针对专家，可以提供预测结果，并由专家控制局部的预测错误。
- 针对普通用户，提供预测结果作为参考
- 针对VIP用户，我们可以提供经过专家和预测模型综合分析之后的预测结果
- 针对彩票厂商，我们可以提供预测技术支持

二、软件设计

1、开发与运行环境

开发环境

- Bash on Ubuntu on Windows--内嵌于Windows10的Ubuntu系统（update 1709 开发者模式）（Ubuntu 14.04）（无图形化界面）
- Windows10（借助Windows10的图形化界面）

运行环境

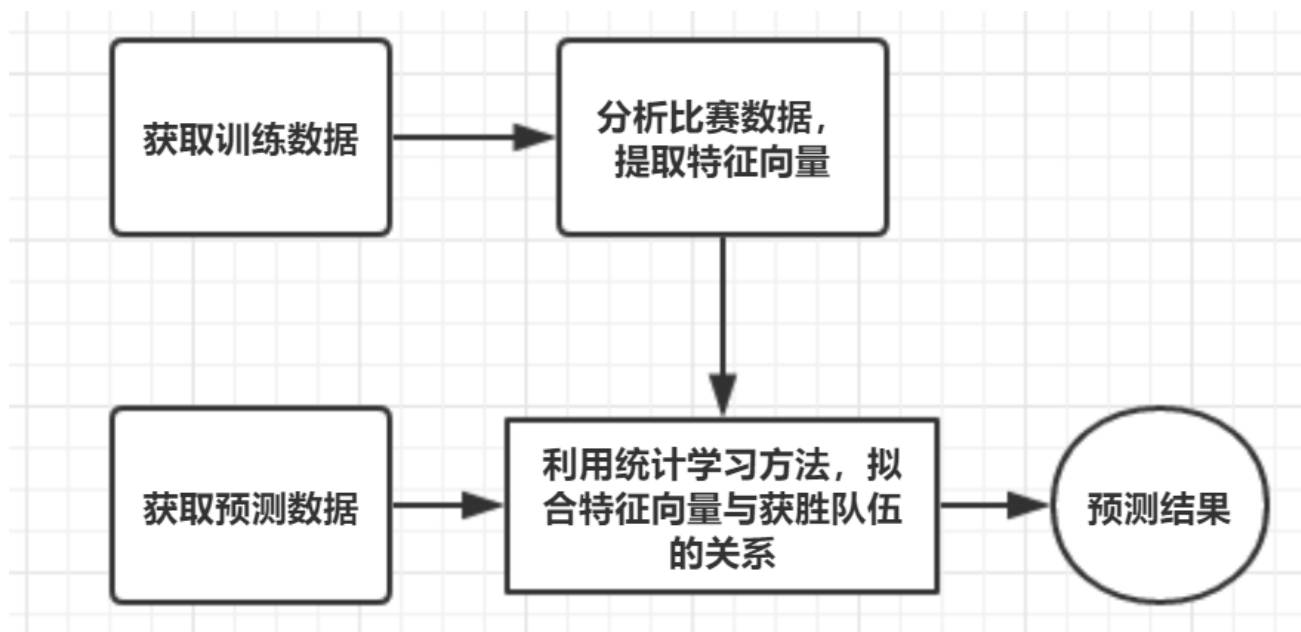
所有能安装 `virtualenv` 的操作系统，包括但不限于：

- Ubuntu
- Mac
- Bash on Ubuntu Windows

以上三种平台都经过测试运行成功。

2、总体处理流程

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. 利用统计学习方法分析每场比赛与胜利队伍的关系，并对2017-2018年的比赛进行预测



4、用户接口

4.1、修改预测数据与训练数据的接口

目前只能手动修改训练集和预测集文件，没有提供一键式接口。

修改预测数据：

修改以下目录的 `17-18Schedule.csv` 文件即可：

```
目录/nba_prediction/nba_pre_demo/proj/data/17-18Schedule.csv
```

如不需要修改，可以保留原文件，原文件中存储的是12月份和1月份的比赛日程。

4.2、执行预测脚本接口

使用python命令执行以下可执行文件：

```
$ python 目录/nba_prediction/nba_pre_demo/proj/prediction.py
```

执行结束之后会在 `prediction.py` 所在目录生成 `17-18Result.csv`

三、代码实现

1、数据结构设计

采用[Basketball Reference.com](https://www.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。

- Team Per Game Stats：每支队伍平均每场比赛的表现统计

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

数据名	含义
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

- Miscellaneous Stats : 综合统计数据

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、三分球和罚球的总共命中率
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

数据项	数据含义
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

- Schedule

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

2、编程语言的选择

编程语言选择 `Python`，因为 `python` 的很多库中实现了机器学习和统计学习的算法，不用重复造轮子。比如本软件使用的 `numpy` 和 `sklearn` 库。

3、核心算法设计

- Elo Score等级分生成特征向量

ELO等级分制度是由匈牙利裔美国物理学家Elo创建的一个衡量各类对弈活动选手水平的评分方法，是当今对弈水平评估的公认的权威方法。被广泛应用于国际象棋、围棋、足球等运动，以及很多网游与电子竞技产业。游戏界比较著名的应用有：WOW（魔兽世界）、DOTA、

LOL。

ELO计算方法:

Ra : A玩家当前的积分

Rb : B玩家当前的积分

Sa : 实际胜负值, 胜=1, 平=0.5, 负=0

Ea : 预期A选手的胜负值, $Ea = 1 / (1 + 10^{[(Rb - Ra) / 400]})$

Eb : 预期B选手的胜负值, $Eb = 1 / (1 + 10^{[(Ra - Rb) / 400]})$

因为E值也为预估, 则 $Ea + Eb = 1$

$R'_a = Ra + K (Sa - Ea)$

R' a : A玩家进行了一场比赛之后的积分

其中 K 值是一个常量系数, 按照国际象棋里的标准, K 值对于大师选手为16, 对于一般选手是32。K值的大小直接关系到一局游戏结束, 根据胜负关系计算出的积分变化值。

- Logistic Regression方法建立回归模型

关于逻辑回归可以参考这个教程：

[逻辑回归--简明教程](#)

*四、总结（与软件设计本身无关）

网络上关于预测模型的建立有非常多的案例, 也不乏NBA比赛的预测模型。但是, 软件工程的关键不在于实现的结果, 而更重要的是实现的过程, 并且, 注重用户对软件的使用。

因此, 我在设计该软件的过程中, 坚持用最先进的代码管理机制与有效的可移植性：

- Git版本控制
- Virtualenv虚拟运行环境

这样, 不论是开发者还是用户, 对软件开发过程中的版本控制和运行平台都不用过多的担心！