



運用隨機森林法預測NBA之勝率



指導老師: 蔡啟揚 教授

學生: 彭仲傑 韓富渝 吳奕賢

研究動機及目的

NBA的比賽精采多元，帶來商機非常大，光是周邊代言或是運動彩券更是一筆可觀的數目，讓預測球隊的勝負排名變得很有挑戰，需運用龐大的資料和因子去印證自己的看法，球迷之間更是會去猜測排名，因此怎麼預測的準確是一件重要的事情。而我們打算使用隨機森林此預測法來去預測新年度的勝率，而隨機森林的優點在於可以處理大量的輸入變數，因為影響球賽的因素通常不會只有少數幾個因子，所以隨機森林對於瞬息萬變的球賽事再適合不過了，而最終的目的則是透過隨機森林預測法去預測2022例行賽賽季的勝率。

研究方法

首先，先收集各球隊的各項因子數據

年度	勝	敗	勝率	比賽	投籃命中	三分球命中	罰球命中	進攻籃板	防守籃板	平均得分	平均籃板	平均助攻	平均失誤	平均抄截	平均阻礙	平均阻礙
2021	33	49	0.402	82	46.9	34.7	73.2	9.5	34.5	112.1	44	24	7.6	5.2	14.5	20.2
2020	42	30	0.583	72	47.2	35.4	73.9	9.7	34.6	109.5	44.2	24.7	7.8	5.4	15.2	19.1
2019	52	19	0.732	71	48	34.9	72.9	10.7	35.1	113.4	45.7	25.4	8.6	6.6	15.2	20.7
2018	37	45	0.451	82	47	33.3	69.9	10.2	36.4	111.8	46.6	25.6	7.5	5.4	15.7	20.7
2017	35	47	0.427	82	46.1	34.5	71.4	10.7	35.7	108.1	46.4	23.8	7.7	4.7	15.8	21.2
2016	26	56	0.317	82	45	34.6	75.4	11.4	32.1	104.6	43.5	20.9	8.2	3.9	15.2	20.7
2015	17	65	0.207	82	41.4	31.7	78.1	10.7	32.3	97.3	43	18	7.2	4.1	13.7	20.3
2014	21	61	0.256	82	43.5	34.4	74.1	11.6	32.3	98.5	43.9	20.9	7.1	4.5	13.2	21.2
2013	27	55	0.329	82	45	38.1	75.7	9.1	32	103	41	24.5	7.5	5.4	15.1	19.8
2012	45	37	0.549	82	45.8	35.5	69.2	11.5	33.3	102.2	44.8	22.2	7	5.2	15	17.9
2011	41	41	0.621	66	45.7	32.6	75.6	12.1	34.1	97.3	46.2	22.5	5.9	5.3	15.1	16.8

我們透過EXCEL迴歸分析來去篩選那些因子與勝率的顯著值較高，來做為隨機森林預測法中判別勝率的關鍵因子，最後得到以下5個因子為關鍵因子

投籃命中率	三分球命中率	進攻籃板	防守籃板	平均得分
-------	--------	------	------	------

隨機森林模型的建立是依靠輸入的數據集(訓練集)來完成的，而隨機森林預測的結果能夠與我們導入的測試集進行比對來查看準確度

```
train=pd.read_csv('2011-2020.csv')
test=pd.read_csv('2021.csv')
y_train=np.array(train['WIN'])
x_train=train.drop('WIN',axis=1)
y_test=np.array(test['WIN'])
x_test=test.drop('WIN',axis=1)
feature_columns=[ 'FG', '3PTFG', 'OREB', 'DREB', 'SCORE', ]
label_columns=[ 'WIN' ]
```

我們測試了使用3年、5年的資料，來去探討使用多少年的資料集當作訓練集資料預測效果最好，以下為各測試20次的平均準確度

	準確度平均值
3年	0.407039
5年	0.376498
10年	0.704395

由上表可知使用10年的資料準確度最高

我們將30對球隊2012-2021的數據當作訓練集，而測試集投入2022球季的數據(截止至12/12)

```
train=pd.read_csv('2012-2021.csv')
test=pd.read_csv('2022.csv')
```

再將各球隊按照東西區做出排名，以下就是我們的最終成果

	西區球隊	勝率	東區球隊	勝率	
1	金塊	0.736	賽爾提克	0.727	1
2	鵜鶘	0.677	籃網	0.663	2
3	國王	0.667	76人	0.615	3
4	勇士	0.617	騎士	0.591	4
5	拓荒者	0.609	公牛	0.576	5
6	太陽	0.599	尼克	0.547	6
7	快艇	0.583	黃蜂	0.5	7
8	灰熊	0.561	暴龍	0.498	8
9	爵士	0.559	公鹿	0.469	9
10	灰狼	0.555	溜馬	0.453	10
11	獨行俠	0.549	巫師	0.441	11
12	火箭	0.522	魔術	0.405	12
13	雷霆	0.445	活塞	0.401	13
14	湖人	0.427	老鷹	0.395	14
15	馬刺	0.405	熱火	0.34	15

結論

目前我們預測出2022-2023球季的各球隊勝率，該球季目前還在進行中，所以我們可以等待到明年得知各球隊勝率來看我們的預測排名是否準確，如果此預測法準確率相當高，未來可以將此方法應用於市場上，站在球團的立場，預測戰績好，周邊收益相對會提高，此時就能提前準備周邊商品以免供貨不足，在觀眾角度也能應用在運動彩卷上，將金錢投入在預測戰績好的球隊，會提升觀眾獲勝的機率。