

# 信息检索第六次作业

2011012162 刘扬

## 实验目的

本次实验的目的是使用四种不同的模型完成对同样搜索询问的结果比对，这四种模型是

- tf\*idf Model
- Extended Boolean Model
- Probabilistic Model
- Language Model

## 实验设定

### 实验环境

本次实验我使用了python设计了四种模型的实现，代码可以在<https://github.com/largelymfs/IRModel> 中下载执行。

### 实验评价

本次实验我使用了定性和定量相结合的方法进行评价，首先对于四种方法定性进行研究，然后使用准确率的分析进行定量研究。

### 实验数据

这次实验我使用了sample之后的SinaNews数据，我一共sample出了四种不同的类别，分别是**体育**，**金融**，**政治**，**科技**，然后我选取了两种数据进行测试：

- 对于体育，我使用 **进球** **晋级** **胜利** 进行搜索。
- 对于金融，我使用 **股市** **盈利** **亏损** 进行搜索。

## 实验结果

为了方便观察，我只选择了最优的前5个进行观察，下面是对于体育搜索，各种模型给出的结果

- Tf-idf Model

78 火箭不指望自己能变成当年勇猛的小牛，也不想琢磨他们到底能拿到多少场胜利，他们想要的就是一场胜利。“我们需要一场胜利，不管下个对手是谁，不管他们的实力如何，我们要的就是一场胜利，”姚明说，“我迫切希望看到这场胜利，我的队友也一样，我们的球迷也一样。为了这场胜

利，我简直都快失眠了。”

1291 新浪体育讯北京时间11月8日消息，在经历了尴尬的11年最差五连败开局之后，火箭终于赢得了本赛季的第一场胜利，成为了联盟最后一支在常规赛中获得胜利的球队。

3644 凯文-杜兰特：这是我第一次的经验在NBA打季后赛，（即便面对）我完完全全不会感觉到紧张，只是尽力而为，希望我和我的队友得到胜利，我们真的是得到了胜利。

1381 新浪体育讯北京时间11月8日（休斯顿当地时间11月7日）消息，一场久违的胜利，这场胜利对于目前的火箭队来说就是一场及时雨。今天他们在主场击败了来访的明尼苏达森林狼队，结束了自己的5连败。

870 赛后的李非常的难过，洗完澡之后他就呆坐在椅子上一直沉默不语，一会儿拿着毛巾沮丧的捂着脑袋。当笔者试着问他那两个罚球时，李低沉的说：“那个机会本来很好的可以结束比赛，拉开两个进球的差距，但是我罚失了。没有什么好多说的，我也不知道怎么回事。”

- Extended Boolean Model

搜索树是(进球 and (晋级 or 胜利))

684 有记者开门见山的问道：“我知道火箭队亟需一场胜利，如果能在客场击败马刺那样的胜利会有什么不同呢？”巴蒂尔回答道：“如果是这样高质量的胜利，大概可以算成是一场半吧。因为马刺的客场是如此的难打。”

901 本赛季之前五次主场对骑士，奇才都站在胜利的一方。新赛季不用对付詹姆斯，但奇才反而打得不顺利。

1277 新浪体育讯北京时间11月8日，火箭在主场以120-94击败森林狼，终于取得了新赛季第一场胜利

1381 新浪体育讯北京时间11月8日（休斯顿当地时间11月7日）消息，一场久违的胜利，这场胜利对于目前的火箭队来说就是一场及时雨。今天他们在主场击败了来访的明尼苏达森林狼队，结束了自己的5连败。

1681 对于即将开始的东部客场之旅，马丁信心满满：“我们已经准备好了东海岸之旅，希望能取得更多的胜利。”

- Probabilistic Model

0 或许正源于此，商瑞华才在首战前3天召开了一堂完全针对“恐瑞症”的战术分析课。3日中午在喜来登大酒店中国队租用的会议室里，商瑞华给大家播放了瑞典队的比赛录像剪辑。这是7月6日瑞典在主场同美国队进行的一场奥运热身赛，当时美国队头牌射手瓦姆巴赫还没有受伤，比赛当中双方均尽遣主力，占据主场优势的瑞典队曾一度占据上风，但终究还是在定位球防守上吃了亏，美国队在下半场的一次角球配合中，通过远射打进唯一进球，尽管慢动作显示是打在瑞典队后卫腿上弹射入网，但从过程到结果，均显示了相同内容——“瑞典队的防守并非无懈可击”。

659 虽然科比今天超越昔日自己的“伯乐”韦斯特成为队史进球之王，可在进攻效率上飞侠这14年跟当年湖人教父的表现还是差了不少。韦斯特的9016球只用了932场，19032次出手，14年湖人生涯的命中率为47.4%；而科比的9019个进球，用了1027场，19833次出手，生涯“打铁”10811次，命中率为45.5%。另外多说一句，乔丹生涯平均命中率49.7%，也要明显强于科比。

735 阿里扎全场只进了五个球，但是却个个都十分重要。首节比赛还剩八分钟左右时，阿里扎连续接球命中空位三分球，这让黄蜂在比赛一开始就确立了今天整场比赛的基调。阿里扎的进球都是接连而来，这也让人特别的印象深刻。在第三节末段，保罗连续发动快攻，阿里扎也总是能及时的跟进，然后用扣篮完成进攻。

870 赛后的李非常的难过，洗完澡之后他就呆坐在椅子上一直沉默不语，一会儿拿着毛巾沮丧的捂

着脑袋。当笔者试着问他那两个罚球时，李低沉的说：“那个机会本来很好的可以结束比赛，拉开两个进球的差距，但是我罚失了。没有什么好多说的，我也不知道怎么回事。”

1280 由于打背靠背，姚明昨天休息，全力准备今天的比赛。火箭的第一次进攻由他来完成，但遭到达科-米勒西奇盖帽。状元与榜眼的较量，姚明不会示弱，本节还有6分59秒时，姚明篮下要到位，背对米勒西奇强打，转身后勾手命中，火箭以9-8超出。姚明刚找到点感觉，不过进球后35秒他就吃到第二次犯规，不得不下场休息，首节的表演到此为止。火箭以一波9-2结束本节，以26-14扩大优势，森林狼本节最后两分钟只投中一球。

• Language Model

1322 巴蒂尔：圣城一役火箭有血性若都如此胜利不难求

78 火箭不指望自己能变成当年勇猛的小牛，也不想琢磨他们到底能拿到多少场胜利，他们想要的就是一场胜利。“我们需要一场胜利，不管下个对手是谁，不管他们的实力如何，我们要的就是一场胜利，”姚明说，“我迫切希望看到这场胜利，我的队友也一样，我们的球迷也一样。为了这场胜利，我简直都快失眠了。”

684 有记者开门见山的问道：“我知道火箭队亟需一场胜利，如果能在客场击败马刺那样的胜利会有什么不同呢？”巴蒂尔回答道：“如果是这样高质量的胜利，大概可以算成是一场半吧。因为马刺的客场是如此的难打。”

3644 凯文-杜兰特：这是我第一次的经验在NBA打季后赛，（即便面对）我完完全全不会感觉到紧张，只是尽力而为，希望我和我的队友得到胜利，我们真的是得到了胜利。

1291 新浪体育讯北京时间11月8日消息，在经历了尴尬的11年最差五连败开局之后，火箭终于赢得了本赛季的第一场胜利，成为了联盟最后一支在常规赛中获得胜利的球队。

从上面的结果可以看出，四种模型给出的搜索结果至少在top-5的情况下都是可靠的。

# 定量分析

我接下来尝试了 **股市** **盈利** **亏损** 进行搜索的结果，选取top5, top 10, top20的结果比较准确度（手动观测），得到的结果是：

方法	Top5	Top10	Top20
tf*idf Model	1.0	0.9	0.85
Extended Boolean Model	1.0	1.0	0.95
Probabilistic Model	1.0	0.8	0.75
Language Model	1.0	0.9	0.9

由此可以看出来，Extended Boolean Model 和Language Model 比其他模型能够更加优秀一些，但是比如Probabilistic Model 模型也有可能是迭代次数不够引起的精度不高。