数据分析课程设计论文

题目：预测LPL春季赛小组出线情况

姓名：刘逸尘 班级：信息与计算科学二班

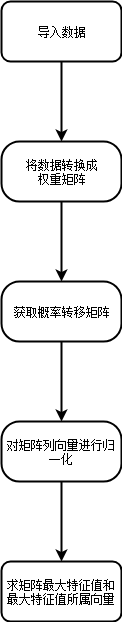
学号：12151209 指导老师：赵昕

摘要：运用PageRank算法对赛事情况进行预测。具体做法是综合考虑各队的比赛成绩, 为每支球队计算相应的等级分(Rank), 然后根据各队的等级分高低来确定名次。此方法优点在于强队战胜弱队所获得的rank分较低，不会因为赛事积分而误判强弱因而可以客观反应出队伍的综合实力。

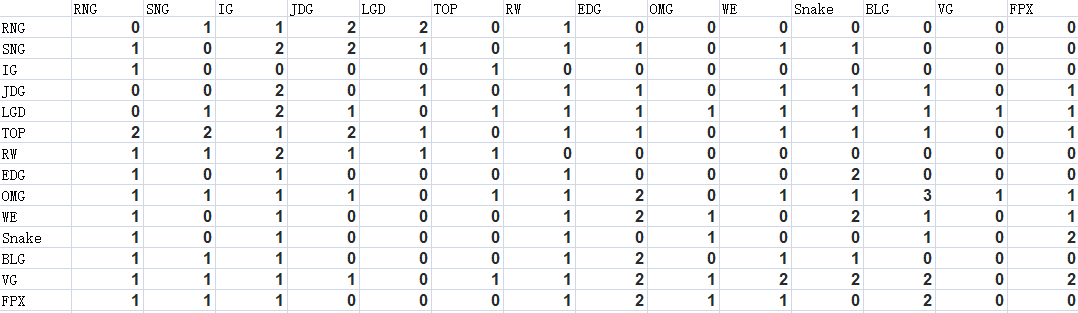
问题分析：

原先选择的题目是世界杯夺冠预测，但由于世界杯赛制的问题导致PR算法不适用。特地采用双循环赛制的LPL春季赛数据来对PR算法进行检验，给定队伍的等级分应该由它所战胜和战平的队伍的数量以及被战胜或战平的队伍的实力共同决定．具体来说，确定队伍Z的等级分的依据应为：一是看它战胜和战平了多少支队伍；二要看它所战胜或战平队伍的等级分的高低．这两条就是我们确定排名的基本原理。由于电竞比赛中不出现平局故而在本次实验中忽略掉这一可能但在其他赛事中可以用加权等级分来对rank分进行再平衡得到更接近实际情况的结果。

算法描述：



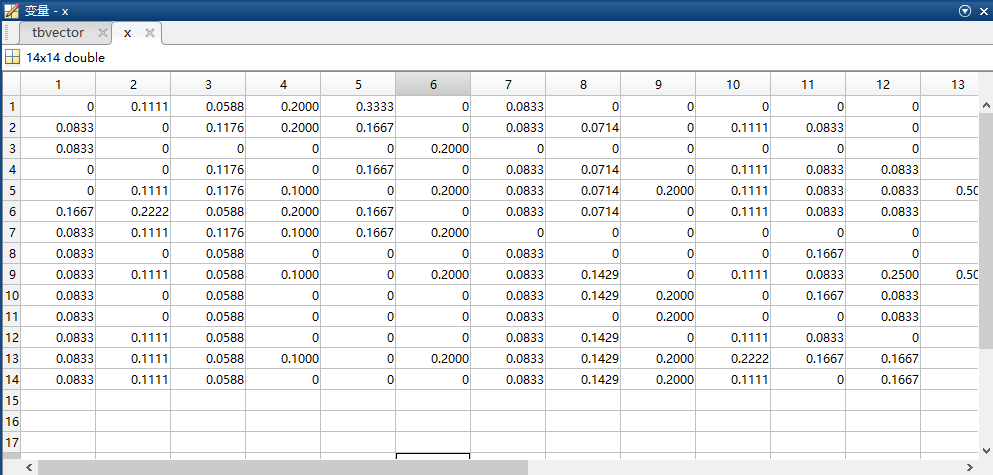
获取概率转移矩阵：



将表中数据转化为相同的权重矩阵，用图的顶点表示相应战队，用连接两个顶点的有向边表示两队的比赛结果。同时给边赋权重，表明占胜的次数。

例如, 表中 RNG与SNG比赛了两场, 各胜一场, 故T12= 1, T21 =1 .

权重矩阵转换成状态转移矩阵：



加权等级分：

本次实验中未出线平局情况不用讨论，但在足球比赛中会出现平局情况因此要用到加权等级分来对其进行进一步处理。

在比赛中出现平局情况应对平局情况进行记录同时对其进行同样的处理求得归一化后的状态转移矩阵，将胜负场的状态转移矩阵和平局的状态转移矩阵分别乘以权重再相加的到新的状态转移矩阵。

如：

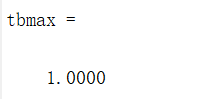
其中P为平局情况下的状态转移矩阵，S为胜负场的状态转移矩阵。

求矩阵特征值和特征向量：

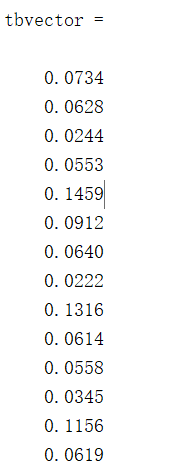
在MATLAB中，计算矩阵A的特征值和特征向量的函数是eig(A)，常用的调用格式有   
5种：   
 (1) E=eig(A)：求矩阵A的全部特征值，构成向量E。   
 (2) [V,D]=eig(A)：求矩阵A的全部特征值，构成对角阵D，并求A的特征向量构成V的列向量。   
 (3) [V,D]=eig(A,’nobalance’)：与第2种格式类似，但第2种格式中先对A作相似变换后求矩阵A的特征值和特征向量，而格式3直接求矩阵A的特征值和特征向量。   
 (4) E=eig(A,B)：由eig(A,B)返回N×N阶方阵A和B的N个广义特征值，构成向量E 。   
 (5) [V,D]=eig(A,B)：由eig(A,B)返回方阵A和B的N个广义特征值，构成N×N阶对   
角阵D，其对角线上的N个元素即为相应的广义特征值，同时将返回相应的特征向   
量构成N×N阶满秩矩阵，且满足AV=BVD。

这里采用第二种即可，求出矩阵的最大特征值和最大特征值所对应的特征向量即为权重。权重的排行即为战队综合实力排名从而打到预测的目的。

最大特征值：



相应等级分：



由此可以预测小组赛可以出线的是：T1、T2、T5、T6、T7、T9、T14这七支队伍。

源代码：

>> x=Book1(:,2:15);

>> x=table2array(x);

>> [m,n]=size(x);

o=sum(x);

for i=1:n

x(:,i)=x(:,i)/o(i);

end

>> [v,d]=eigs(x);

tbmax=max(d(:));

[m,n]=size(v);

sum = 0;

>> for i=1:m

sum = sum + v(i,1);

end

tbvector = v(:,1);

for i=1:m

tbvector(i,1)= v(i,1)/sum;

end

>> tbmax

tbmax =

1.0000

>> tbvector

tbvector =

0.0734

0.0628

0.0244

0.0553

0.1459

0.0912

0.0640

0.0222

0.1316

0.0614

0.0558

0.0345

0.1156

0.0619

>>

总结：

本次课程设计走了不少弯路，在最开始准备做预测世界杯夺冠时，在信息选择和采用上没有太进行思索花了很多时间搜集了不少用不到的数据。在算法的选择和算法的适用性上也存在大量误区，选择了PR算法多次进行改进后依旧得不到想要的数据，经老师点拨后才发觉世界杯赛制由小组淘汰制的积分赛和淘汰赛等赛制复合而成。想要进行局部预测而世界杯小组赛又是单循环赛制而PR算法仅适用于双循环故而只能忍痛放弃。课设的时间已经不多最终决定放弃原先选好的题目转而利用PR算法来解决一下双循环赛制的比赛问题，最终我选择了用pr算法预测LPL比赛中的小组赛出线问题。测试的结果与实际相符，成功完成预测。

现在的大型比赛为了公平和举办赛事的种种原因赛制在不断改革，简单的单一算法已经无法满足预测的需要。对应复合赛制的复合算法也应当不断改进才能逐渐变得精确。

文献参考：

1、Google 搜索引擎的数学模型及其应用----赵国, 宋建成 (西南民族大学计算机科学与技术学院, 四川成都 610041)

2、机器学习之PageRank算法的应用与C#实现