**摘要**

对于第一个问题，如何评价校园教练的执教水平。首先，本文采用**NCAA**比赛的数据，以确保评价对象为校园教练；选取了4个因素，执教年数、参赛历史（包括参赛次数、赢的次数、输的次数）、参加比赛获胜率、以及得到的奖项,作为评价教练执教水平的指标。其次，本文采用**熵权法**对各个指标进行赋权，再对每个教练进行**打分**，以分数的高低进行排名，分数高意味着教练的总体执教水平较高，即可找出前十名的校园教练。

对于第二个问题，在找出前十名总体水平较高的校园教练的基础上，进一步考虑时间因素的影响，时间因素的影响主要在两方面：一是不同年份获得相同奖项对教练水平的评价贡献不同，这通过构造荣誉系数，对不同年度的荣誉系数进行不同的赋权值来解决；二是教练不同年份参加比赛次数、赢的次数和输的次数对水平评价的参考意义不同，这个问题通过次数按比例折算进行解决。通过对第一个模型进行**时间修正**，在第一个模型找出的TOP10的基础上，选出TOP5的教练。TOP5的教练依次是：Mike Krzyzewski、Adolph Rupp、Jim Boeheim、Dean Smith、Bob Knight。

**关键词：** **熵权法** **时间修正 荣誉系数 标准化**

**I. 介绍说明  
1.1 问题背景**

体育画报是由媒体巨擘时代华纳所拥有的美国体育周刊。这个杂志正在寻找上个世纪最好的男性或女性大学教练。这个最好的大学教练无论是男性或者女性，无论是大学曲棍球还是场地曲棍球，橄榄球，棒球或垒球，篮球或是足球。

提出的问题：

1. 如何选择指标、建立模型，评估大学教练的水平？
2. 时间因素对教练水平评估有什么影响，如何解决？

**II.模型假设及符号说明**

**2.1模型假设**

假设1：NCAA参赛的教练均为校园教练。

假设2：本文采用的四个因素对教练水平的评价至关重要，并且本文所忽略的因素对教练水平的评价几乎不起作用

假设3：本文所搜集的数据是准确的且足够进行分析评价。

**2.2符号说明**

|  |  |
| --- | --- |
| **符号** | **符号说明** |
|  | 第*i*个教练的第*j*项指标的值 |
|  | 标准化后的数据 |
| *n* | 指标个数 |
|  | 第j个指标的信息熵 |
|  | 第j个指标的权重 |
|  | 第一次评分第i个教练的分数 |
|  | 第*i*个教练的第*j*项指标汇总数据的值 |
|  | 指标汇总数据标准化后的数据 |
|  | 第j个指标的权重 |
|  | 第二次评分第i个教练的分数 |
|  | 第*i*个教练的第*j*项指标的值 |
|  | 标准化后的数据 |

**III.模型建立与求解**

**3.1问题分析**

为了评价不同球类不同校园教练的水平，需要不同教练参加相关比赛的资料。本文以篮球为例，以各个教练参加美国大学体育总会（NCAA）篮球比赛的相关数据为基础。之所以选用NCAA的数据，是因为采用此数据可以保证参加比赛的教练都是大学教练。查找资料确定可以显著反映教练水平的指标，采用科学合适的方法确定其权重，进行打分，选出得分最高10名校园教练。

**3.2模型建立**

**3.2.1选出评价指标**

通过查找资料，发现教练的执教年数、参赛历史（包括参赛次数、赢的次数、输的次数）、参加比赛获胜率、以及得到的奖项能够较好地反映教练的水平。

执教年数 参赛历史 获胜率 奖项

GREG

参赛次数 CTRN

赢的次数 NCAA

输的次数 FF

NC

**3.2.2采用熵权法确定权重**

**Step1：**数据标准化

将指标数据进行标准化。设第*i*个教练的第*j*项指标的值为,标准化后的数据记为。

 （3.2.2.1）

**Step2：**求各个指标的信息熵

设第j个指标的信息熵为，指标个数为*n*。

 （3.2.2.2）

其中，

 (3.2.2.3)

若，则定义

Step3：确定各指标权重

根据信息熵的计算公式，计算出各个指标的信息熵为。

通过信息熵确定各指标的权重。设第j个指标的权重为。

 （3.2.2.4）

**3.2.3对各个教练进行打分**

设第i个教练的分数。则：

 （3.2.3.1）

**3.2.4对各个教练进行排名**

得分越高，证明教练水平越高。

**3.3** 结果分析

表1是打分前十名的校园教练及其排名

表1

|  |  |  |
| --- | --- | --- |
| 排名 | 教练 | 打分 |
| 第1名 | Mike Krzyzewski | 0.7332 |
| 第2名 | John Wooden | 0.7137 |
| 第3名 | Dean Smith | 0.6758 |
| 第4名 | Adolph Rupp | 0.6591 |
| 第5名 | Jim Calhoun | 0.5530 |
| 第6名 | Denny Crum | 0.5208 |
| 第7名 | Rick Pitino | 0.4929 |
| 第8名 | Roy Williams | 0.4826 |
| 第9名 | Bob Knight | 0.4474 |
| 第10名 | Jim Boeheim | 0.4315 |

**IV.模型拓展——时间的影响**

4.1 问题分析

从1913到2016年，NCAA的赛制发生了几次变化，参赛队伍不断增加，不同年份参加比赛次数、赢的次数和输的次数对评价的参考意义不同，在不同年份获得相同奖项的难度是不一样的，这就是时间因素对上文所建立的模型的影响。所以，需要进行时间修正，即对不同年份的参赛次数进行按比例折算，对不同年份的相同奖项赋不同的值，使得在评价分析中不同年份的相同奖项对教练水平的评价贡献不同。

4.2 模型建立

4.2.1 NCAA时间段划分

本文根据NCAA参赛队伍的多寡和阶段性特征，对NCAA的历史进行分段。分段结果如表2所示。

表2

|  |  |
| --- | --- |
| 时间段 | 特点 |
| 1913-1939 | NCAA未正式形成 |
| 1939-1951 | 8支参赛队伍 |
| 1951-1975 | 16支参赛队伍 |
| 1975-1980 | 32支参赛队伍 |
| 1980-1985 | 48支参赛队伍 |
| 1985-2016 | 64支参赛队伍 |

4.2.2 指标构造——荣誉系数

通过构造荣誉系数，一方面对教练获得的奖项进行量化处理，另一方面以不同的荣誉系数的值，使得在评价分析中不同年份的相同奖项对教练水平的评价贡献不同。

荣誉系数的值是通过衡量不同年份获得该奖项的难易程度来决定的。比如，1986年的全国冠军的难度大，荣誉系数为64，而1940年全国冠军的难度较1986年的小，荣誉系数为8。获得各个奖项的荣誉系数如表3所示。

表3

|  |  |  |
| --- | --- | --- |
| 时间段 | 奖项 | 荣誉系数 |
| 1939-1951 | Win national final | 8 |
| Lost national final | 4 |
| Lost regional final | 2 |
| Lost regional semifinal | 1 |
| 1951-1975 | Win national final | 16 |
| Lost national final | 8 |
| Lost national semifinal | 4 |
| Lost regional final | 2 |
| Lost regional semifinal | 1 |
| Lost first round | 0.5 |
| 1975-1980 | Win national final | 32 |
| Lost national final | 16 |
| Lost national semifinal | 8 |
| Lost regional final | 4 |
| Lost regional semifinal | 2 |
| Lost second round | 2 |
| Lost first round | 1 |
| 1980-1985 | Win national final | 48 |
| Lost national final | 24 |
| Lost national semifinal | 12 |
| Lost regional final | 8 |
| Lost regional semifinal | 4 |
| Lost second round | 2 |
| Lost first round | 1 |
| Lost opening round | 0 |
| 1985-2016 | Win national final | 64 |
| Lost national final | 32 |
| Lost national semifinal | 16 |
| Lost regional final | 16 |
| Lost regional semifinal | 8 |
| Lost third round | 4 |
| Lost second round | 2 |
| Lost first round | 1 |
| Lost opening round | 0 |

4.2.3 对次数进行按比例折算

Step1:对每一年度不同教练的相关指标的具体数据进行标准化。

记每一年度的第i个教练第j个指标的数据为，标准化数据为。则：

（4.2.3.1）

Step2：将每一教练在其执教年间的标准化数据进行汇总。

Step3：将10个教练的指标汇总数据进行标准化。

记第i个教练第j个指标的汇总数据为，汇总数据的标准化数据为。则：

（4.2.3.2）

4.2.4 采用熵权法确定权重、进行打分

熵权法的步骤与上文一致，所得到的权重记为.

设第i个教练的分数。则：

 （4.2.4.1）

4.2.5 对教练进行排名

教练得分越高，证明其水平越高。

4.3 结果分析，如表4可得1913-2013校园教练总体水平排名前五分别是Mike Krzyzewski、Adolph Rupp、Jim Boeheim、Dean Smith、Bob Knight。

表4

|  |  |  |
| --- | --- | --- |
| 排名 | 教练 | 打分 |
| 第1名 | Mike Krzyzewski | 0.7332 |
| 第2名 | Adolph Rupp | 0.7137 |
| 第3名 | Jim Boeheim | 0.6758 |
| 第4名 | Dean Smith | 0.6591 |
| 第5名 | Bob Knight | 0.5530 |

**附录：**

1、初步筛选

close all;clear;clc;

shuju=xlsread('ÀºÇò3000½ÌÁ· Á¿»¯.xlsx','A1:M3394');

for ii=11:13

for i=1:length(shuju)

if isnan(shuju(i,ii))==1

shuju(i,ii)=0;

end

end

end

backup=shuju;

backup(:,7)=max(shuju(:,7))-shuju(:,7);

%%%%%%±ê×¼»¯ÊýÖµ

for i=1:length(shuju)

for j=4:13

shuju(i,j)=(backup(i,j)-min(backup(:,j)))/(max(backup(:,j))-min(backup(:,j)));

end

end

%%%%%%

p=zeros(length(shuju),10);

for i=1:length(shuju)

for j=1:10

p(i,j)=shuju(i,j+3)/sum(shuju(:,j+3));

end

end

for j=1:10

pt=p(:,j);

for i=1:length(pt)

if pt(i)==0

lim(i)=0;

else

lim(i)=pt(i)\*log(pt(i));

end

end

e(j)=-1/log(length(shuju))\*sum(lim);

end

w=(1-e)/sum(1-e);

w=w';

point=zeros(length(shuju),1);

for i=1:length(shuju)

point(i)=shuju(i,4:13)\*w;

end

[B,ind]=sort(point,'descend');

Bind=[B,ind];

2、时间修正筛选

close all;clear;clc;

coach=xlsread('ÎÞ sinceÐÞÕý.xlsx','A1:N388');

session=xlsread('1931-2016',2,'A2:E87')

conclusion=zeros(10,7);

for No=1:10

[x,y]=find(coach(:,4)==No);

correcting=coach(x,:);

corrected=correcting;

for i=1:length(x);

year=correcting(i,2)

corrected(i,8)=correcting(i,8)-session(year-1929,3)/(session(year-1929,2)-session(year-1929,3));

corrected(i,9)=correcting(i,9)-session(year-1929,5)/(session(year-1929,4)-session(year-1929,5));

end

Yrs=length(x);

G=sum(corrected(:,8))+sum(corrected(:,9));

W=sum(corrected(:,8));

L=sum(corrected(:,9));

WL=sum(correcting(:,8))/(sum(correcting(:,8))+sum(correcting(:,9)));

honor=sum(corrected(:,14));

sumarry=[No,Yrs,G,W,L,WL,honor];

conclusion(No,:)=sumarry

end

backup=conclusion;

%%%%%%±ê×¼»¯ÊýÖµ

for i=1:length(conclusion)

for j=2:7

conclusion(i,j)=(backup(i,j)-min(backup(:,j)))/(max(backup(:,j))-min(backup(:,j)));

end

end

conclusion(:,5)=1-conclusion(:,5);

%%%%%%

p=zeros(length(conclusion),10);

for i=1:length(conclusion)

for j=1:6

p(i,j)=conclusion(i,j+1)/sum(conclusion(:,j+1));

end

end

for j=1:6

pt=p(:,j);

for i=1:length(pt)

if pt(i)==0

lim(i)=0;

else

lim(i)=pt(i)\*log(pt(i));

end

end

e(j)=-1/log(length(conclusion))\*sum(lim);

end

w=(1-e)/sum(1-e);

w=w';

point=zeros(length(conclusion),1);

for i=1:length(conclusion)

point(i)=conclusion(i,2:7)\*w;

end

[B,ind]=sort(point,'descend');

Bind=[B,ind];