

数学建模竞赛

模糊聚类分析在足球队排名中的应用*

曾文艺

(北京师范大学数学系, 北京 100875)

崔宝珍

(天水师范专科学校教务处, 天水 741000)

摘 要 本文针对 93 年全国大学生数学建模竞赛 B 题, 运用模糊聚类分析的方法, 讨论了足球队比赛的排名问题, 得到的结果是: $T_7, T_1, T_3, T_9, T_{10}, T_8, T_{11}, T_{12}, T_2, T_6, T_5, T_4$, 此排名结果合理、可信, 并且对参数在一定范围内的变化有良好的稳定性.

关键词 模糊聚类分析, 灵敏度分析

一、问题的分析

足球队排名次的问题是 93 年全国大学生数学建模竞赛 B 题^[1], 题中给出我国 12 支足球队在 1988—1989 年全国足球甲级联赛中的成绩, 通过研究成绩表, 我们发现: 该表所包含的数据量庞大, 而且表中的数据残缺不全, 队与队之间的比赛场数相差较大, 直接根据成绩表来排出它们的名次比较困难; 不过, 我们可以先看看每个队在它所参加的比赛中, 胜、负以及平的场数, 凭借它, 我们对每个队的实力有一个大概的了解, 于是得到表 1.

表 1 参赛各队胜负以及平的场数分布表

队 \ 场数	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}
胜的场数	10	5	8	1	2	2	13	6	7	6	1	2
负的场数	5	4	4	12	5	3	1	8	8	5	6	3
平的场数	4	6	3	6	2	0	3	3	2	6	2	4
总场数	19	15	15	19	9	5	17	17	17	17	9	9

本文 1997 年 8 月 26 日收到.

* 本文得到北京市普通高校教学改革试点资助

其次,我们也可以再看看各队在比赛中的平均每场进球数、失球数和进失球差数,这也有助于我们进一步了解各队的实力,通过成绩表,我们有:

表 2 参赛各队平均每场进失球数分布表

	T_1	T_2	T_3	T_4	T_5	T_6
进球数	1.412	0.8	1.333	0.632	1	0.6
失球数	0.941	0.667	0.8	1.684	1.444	1.2
进失球差数	0.471	0.133	0.533	-1.052	-0.444	-0.6
	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}
进球数	2.059	0.941	0.647	0.882	0.778	0.667
失球数	0.588	0.824	1	1	1.556	1
进失球差数	1.471	0.118	-0.353	-0.118	-0.778	-0.333

通过表 1 和表 2 的分析,我们有理由认为 T_7 是最好的, T_4 是最差的; $T_5, T_6, T_{10}, T_{11}, T_{12}$ 等队的成绩靠后, T_1, T_2, T_3, T_9, T_8 等队的水平居中,但它们之间的差距都不太大,考虑到数据的不对称性和残缺性,仅根据上述两表来确定其名次,则其合理程度显然值得怀疑.

为了使排名更趋合理和可信,我们应该综合考虑 T_i 与其余各队的比赛成绩,充分利用 12 组数据,则可能找出更为合理的排名算法. 考虑到有些队两两之间没有比赛,其成绩难以确定,并且评判人的思维具有大量的模糊信息,因此,我们的想法是:先制定一个规则,为各队定义一组特征数据,同时计算各队之间的水平相似程度(即模糊相似程度),利用模糊聚类分析方法,根据聚类结果,并综合表 1 和表 2 的信息来确定各队的名次.

二、模型假设

考虑实际情况和解决问题的方便,我们做如下的假设:

- 1) 如果 T_i 与 T_j 没有比赛,那么假设它们的比赛成绩为 $A : B$, 且只赛一场, 令 $Q = A - B$, 先取 $\theta = 0$;
- 2) 每场比赛对于排名同等重要, 每个进失球对于排名也同等重要;
- 3) 在确定各队的特征数据时, 仅计算进失球的差数, 则第 i 队的特征数据记为 $r_i = (r_{i1}, \dots, r_{in})$, 其中 $n = 12$.

根据常识, 比赛时甲以一场 $2 : 1$ 胜乙, 易于两场都以 $2 : 1$ 胜乙, 更易于三场都以 $2 : 1$ 胜乙, 故计算特征数据时应考虑加权因子, 如赛一场时, $R_{甲乙} = (2-1) \cdot S$, 赛两场时, $r_{甲乙} = \frac{(2-1)+(2-1)}{2} \cdot V$, 赛三场时, $r_{甲乙} = \frac{(2-1)+(2-1)+(2-1)}{3} \cdot U$, 并且 $U > V > S$, 我们先取 $U = 1.4, V = 1.2, S = 1.0$.

- 4) T_i 与 T_i 自身的特征数据为 $r_{ii} = 0$;

5) T_i 与 T_j 之间的模糊相似程度用绝对值减数法来确定 $x_{ij} = 1 - c \cdot \sum_{k=1}^{12} |r_{ik} - r_{jk}|$,

通过估算 $c = 0.038$;

6) 排名原则是: 越先聚为一类的队, 名次越靠近.

三、建模及求解

在模型假设下, 根据成绩表中的数据, 我们可以计算出各队的特征数据如下: 假设论域为 $T = \{T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9, T_{10}, T_{11}, T_{12}\}$, 于是有:

$$r_1 = (0, 0, -0.466, 2.334, 2, 1, -1.8, -0.6, 3, 0, 0, 0),$$

$$r_2 = (0, 0, -0.466, 0.934, 0, 1, 0, 0, 2, -2, 0, 0),$$

$$r_3 = (0.466, 0.466, 0, 0.934, 1, 3, 1, -1, 1, -1, 0, 0),$$

$$r_4 = (-2.334, 0, -1, 1, 0, -1, 0, 0, 0, 0, 0, -0.6),$$

$$r_5 = (-2, 0, -1, 1, 0, -1, 0, 0, 0, 0, 0, -0.6),$$

$$r_6 = (-1, -1, -3, 1, 1, 0, 0, 0, 0, 0, 0, 0),$$

$$r_7 = (1.8, 0, -1, 3.6, 0, 0, 0, 1.4, 2.334, 2.334, 2, 2),$$

$$r_8 = (0.6, 0, 1, 0.6, 0, 0, -1.4, 0, 0, 0, 2, 0),$$

$$r_9 = (-3, -2, -1, 0.6, 0, 0, -2.334, 0, 0, 1.866, 1, 1),$$

$$r_{10} = (0, 2, 1, 0.6, 0, 0, -2.334, 0, -1.866, 0, 1, 2),$$

$$r_{11} = (0, 0, 0, 0, 0, 0, -2, -2, -1, -1, 0, -0.466),$$

$$r_{12} = (0, 0, 0, 0, 0.6, 0, -2, 0, -1, -2, 0.466, 0).$$

利用绝对值减数法, 我们可以计算出 T_i 与 T_j 的模糊相似程度 x_{ij} , 于是有模糊相似矩阵 X .

$$X = \begin{pmatrix} 1 & 0.666 & 0.544 & 0.351 & 0.473 & 0.496 & 0.339 & 0.514 & 0.306 & 0.346 & 0.511 & 0.526 \\ & 1 & 0.641 & 0.397 & 0.65 & 0.597 & 0.389 & 0.59 & 0.351 & 0.392 & 0.587 & 0.678 \\ & & 1 & 0.275 & 0.483 & 0.506 & 0.182 & 0.493 & 0.184 & 0.296 & 0.531 & 0.516 \\ & & & 1 & 0.645 & 0.511 & 0.004 & 0.453 & 0.569 & 0.372 & 0.602 & 0.572 \\ & & & & 1 & 0.749 & 0.313 & 0.62 & 0.574 & 0.422 & 0.577 & 0.557 \\ & & & & & 1 & 0.26 & 0.567 & 0.521 & 0.369 & 0.488 & 0.549 \\ & & & & & & 1 & 0.405 & 0.303 & 0.237 & 0.129 & 0.179 \\ & & & & & & & 1 & 0.529 & 0.681 & 0.648 & 0.699 \\ & & & & & & & & 1 & 0.478 & 0.42 & 0.471 \\ & & & & & & & & & 1 & 0.572 & 0.623 \\ & & & & & & & & & & 1 & 0.828 \\ & & & & & & & & & & & 1 \end{pmatrix}$$

根据直接聚类方法,对12个队可得到其聚类结果,再结合表1和表2的分析,可知 T_4 是倒数第1名,由假设6,首先与 T_4 聚成一类的队是 T_5 ,因此, T_5 是倒数第2名,由此类推,最后与 T_4 聚成一类的队是 T_7 ,所以, T_7 是第1名,全部的排名结果如下:

名次	1	2	3	4	5	6	7	8	9	10	11	12
队名	T_7	T_3	T_1	T_9	T_{10}	T_8	T_{11}	T_{12}	T_2	T_6	T_5	T_4

四、参数的灵敏度分析

由于模型中含有较多参数,而且部分参数的取值又有一定的主观性,若要排名结果合理、可靠,令人信服,则其解对参数在一定范围内的变化来说,应该是稳定的.即当参数有较小的变化时,排名结果不应该有变化或者不应该有较大的变化,考虑到我们使用的是离散模型,故我们对参数作一些数值分析,得到的结果如下:

	名次											
	1	2	3	4	5	6	7	8	9	10	11	12
$Q=0, U=1.4$ $V=1.2, S=1$	T_7	T_3	T_1	T_9	T_{10}	T_8	T_{11}	T_{12}	T_2	T_6	T_5	T_4
$Q=0.1, U=1.4$ $V=1.2, S=1$	T_7	T_3	T_1	T_9	T_{10}	T_8	T_{11}	T_{12}	T_2	T_6	T_5	T_4
$Q=0.2, U=1.4$ $V=1.2, S=1$	T_7	T_3	T_1	T_9	T_{10}	T_8	T_{11}	T_{12}	T_2	T_6	T_5	T_4
$Q=0.3, U=1.4$ $V=1.2, S=1$	T_7	T_3	T_1	T_9	T_{10}	T_8	T_{11}	T_{12}	T_2	T_6	T_5	T_4
$Q=-0.1, U=1.4$ $V=-0.2, S=1$	T_7	T_3	T_1	T_9	T_{10}	T_8	T_{11}	T_{12}	T_2	T_6	T_5	T_4
$Q=-0.2, U=1.4$ $V=1.2, S=1$	T_7	T_3	T_1	T_9	T_{10}	T_8	T_{11}	T_{12}	T_2	T_6	T_5	T_4
$Q=-0.3, U=1.4$ $V=1.2, S=1$	T_7	T_3	T_1	T_9	T_{10}	T_8	T_{11}	T_{12}	T_2	T_6	T_5	T_4
$Q=0, U=1.5$ $V=1.3, S=1$	T_7	T_3	T_1	T_9	T_{10}	T_8	T_{11}	T_{12}	T_2	T_6	T_5	T_4
$Q=0.1, U=1.6$ $V=1.4, S=1.1$	T_7	T_3	T_1	T_9	T_{10}	T_8	T_{11}	T_{12}	T_2	T_6	T_5	T_4

从灵敏度分析表可以看出,当 $Q \in [-0.3, 0.3]$ 时,所得的排名结果不发生变化,这说明结果对 Q 的不灵敏性较好,同时当 U, V, S 发生变化时,其排名结果也没有变化,这也说明结果对 U, V, S 不灵敏,所以,我们的排名结果具有良好的稳定性,所得到的结果是合理的、可信的.

五、模型的推广及进一步讨论

我们的排名算法很容易推广,当队数不是12,而是任意确定的 N 时,利用计算机

来计算是很方便的，当然同时也要受到计算机的物理条件所限制。

另外，我们的算法还受以下条件所限制：

- 1) 当有两个队，它们的比赛成绩完全一样时，算法无法排出其名次；
- 2) 当根据比赛成绩不能判断哪些队成绩较好或较差时，算法无法排出其名次；
- 3) 当残缺数据或没有比赛的场数太多时，算法也将失效。可以设想，即使排出名次来，排名结果对参数的变化也将是灵敏的，结果的稳定性也将较差。

参 考 文 献

- [1] 蔡大用，关于球队排名问题的几点评注，数学的实践与认识，1994，2：95-96。
- [2] 罗承忠，模糊集引论，北京师范大学出版社，北京，1989。

Application of Fuzzy Cluster Analysis on the Football Team Ranking

ZENG WEN-YI

(Department of Mathematics, Beijing Normal University, Beijing, 100875)

CUI BAO-ZHEN

(Instruction Section, Tianshui Teacher School, Tianshui, 741000)

Abstract Aimed at the CUMCM — 93B, we use the method of fuzzy cluster analysis, discuss the football team ranking and obtain its result: $T_7, T_1, T_3, T_9, T_{10}, T_8, T_{11}, T_{12}, T_2, T_6, T_5, T_4$. The result is believable and stable when these parameters are variable on the certain range.

Key words Fuzzy cluster analysis, sensitivity analysis