



● 运动训练学

K—Mean聚类分析对第 18届世界杯足球赛各参赛队进攻能力的研究

王朝信

摘要: 对第 18 届世界杯足球赛 32 支球队 64 场比赛的进球数、射进指数、助攻、射正球门次数、射门次数、角球、任意球、越位、反击次数、短传、长传、传中次数等 12 项进攻指标进行 K—Mean 聚类分析, 并对各指标的差异性进行方差分析, 结果显示各类间在场均进球数、射进指数、助攻、射正球门次数、射门次数、角球、短传次数等 7 项进攻指标上均存在显著性差异, 表明应用 K—Mean 聚类分析对球队的各项进攻指标进行量化评价具有较高的可靠性, 可以客观地反映出各队的进攻能力, 揭示出现代足球进攻的发展趋势, 并探索一种较为合理有效的量化评价球队进攻能力的方法。

关键词: 第 18 届世界杯; 足球赛; 进攻能力; K—Mean 聚类分析

中图分类号: G843 文献标识码: A 文章编号: 1008—1909(2009)01—0110—04

K—Mean Gathers a Kind of Analysis on the Attack Abilities in the 18th Session of World Cup Soccer Game Each Participating Team

WANG Chao-xin

(Physical Education Department of Heze College Heze 274015 China)

Abstract: Enters the ball numbers to the 18th session of World Cup soccer game 32 teams 64 competitions, shoots the index, the secondary attack, shoots the goal number of times, shoots the number of times, the corner kick, the free kick, the offside, the counter-attack number of times, passes in short distance, in the long pass, the biography, the number of times and so on 12 attack targets carries on K—Mean to gather a kind of analysis, and carries on the variance analysis to various targets difference, finally demonstrated during each kind enters the ball number in the field, shoots the index, the secondary attack, shoots the goal number of times, shoots the number of times, the corner kick, passes in short distance, the number of times and so on in 7 attack targets to have the difference. Indicated gathers a kind of analysis using K—Mean to carry on the quantification appraisal to teams each attack target to have a higher reliability may objectively reflect various teams the attack ability, promotes the modern soccer attack the development tendency, and explores one kind of more reasonable effective quantification appraisal team attack ability the method.

Key words: the 18th session of World Cup soccer games; attack ability; K—Mean gathers a kind of analysis

由于现代足球全攻全守的整体性打法趋势使足球比赛的攻守矛盾日趋激化, 给进球增加了更大的难度, 导致近几届世界杯足球赛的进球率呈下降趋势, 体现出球队越来越注重防守的发展方向, 所以很有必要对球队的进攻体系进行系统地研究与分析。但目前国内、外关于足球进攻体系的研究大多数停留在各指标间两两比较的研究与分析上, 缺少对足球比赛多因素综合分析, 尚未从比赛活动的整体角度对球队的进攻能力的运用效果进行评价。为此, 本文拟选择进球数、射进指数、助攻、射正球门次数、射门次数、角球、任意球、越位、反击次数、短传、长传、传中次数等 12 项主要进攻指标, 应用 K—Mean 聚类分析法对第 18 届世界杯足球赛各参赛队的进攻质量给予综合分析, 试图通过研究发现问题和差距, 找出一种较为合理有效的量化评价球队进攻质量的方法, 为尽快提高我国的足球水平, 提供理论基础和科学的依据。

1 研究对象与方法

1.1 研究对象

以参加 2006 年第 18 届世界杯足球赛 32 支球队的 64 场比赛的进球数、射进指数、助攻、射正球门次数、射门次数、角球、任意球、越位、反击次数、短传、长传、传中次数等 12 项进攻指标为研究对象。

1.2 研究方法

主要采用录像观察法、文献资料法、数理统计法、专家访谈法以及逻辑分析法等方法来进行研究。

1.3 统计概念与尺度

射进指数: $\text{进球数} / \text{射门次数} \times 100$

助攻次数: 场上队员传球给队友并且使队友进球得分的传球的次数。

收稿日期: 2008—10—13 修回日期: 2008—12—22

作者简介: 王朝信 (1979—), 男, 山东莘县人, 硕士, 讲师, 研究方向为足球教学与训练。

作者单位: 菏泽学院 体育系, 菏泽 274015

越位次数: 是指进攻队员在进攻时由于处在越位位置而被判罚越位的次数。

反击次数: 是指一个队在获得控球权后, 趁对方重新组织严密防守之前, 以迅雷不及掩耳之势, 将球传递给中前场有利位置的队员, 或个人在中、前场抢、断球后, 快速运球直捣球门, 创造有利的射门机会的次数。

传中次数: 是指进攻队员在边路有目的地把球踢向中路预定方位的战术方法的次数。

2 数据处理

2.1 统计方法与原则

K-Means聚类分析是用户制定类别数的大样本资料的逐步聚类分析, 它先对数据进行初始分类, 然后逐步调整, 得到最终分类, 该聚类方法又称为快速聚类方法。

2.2 K-Means聚类分析的计算过程

首先制定聚类分类数定义为 5 该系统默认值; 迭代次数为 10 收敛参数为 0.02 其次 SPSS for windows 11.5 统计软件确定 5 个类的初始类中心点。计算所有样本数据点到 5 个类中心点的欧氏距离, 其中欧氏距离 (Euclidean Distance) 的计算公式为:

$$EUCLID = \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

其中, k 表示每个样本有 k 个变量; x_i 表示第一个样本在第 i 个变量上的取值; y_i 表示第二个样本在第 i 个变量上的取值。

然后 SPSS for windows 11.5 统计软件重新确定 5 个类的中心点。最后重复上面的两步计算过程, 直到达到指定的迭代次数或者终止迭代的判断要求为止。

2.3 K-Means聚类分析的数据处理过程

2.3.1 初始类中心点

表 1 K-Means聚类分析的初始类中心点表

类别	1	2	3	4	5
进球数	1.67	2.00	2.25	0.33	0.00
射进指数	10.20	14.71	13.04	2.56	0.00
助攻次数	0.67	1.60	1.00	0.33	0.00
射正球门	7.67	7.40	8.75	4.67	2.33
射门次数	16.33	13.60	17.25	13.00	7.33
角球次数	7.67	6.20	9.25	3.67	1.67
任意球次	0.67	1.20	1.25	0.67	0.00
越位次数	3.67	3.00	0.75	2.33	2.00
反击次数	9.00	5.00	5.75	0.67	2.33
短传次数	334.00	407.60	453.50	265.30	185.70
长传次数	90.70	74.40	93.00	93.00	92.70
传中次数	26.30	17.60	24.50	15.30	15.70

表 1 是 SPSS 指定的初始类中心点。由于需要快速聚类成 5 类, 因此指定了 5 个初始类中心点。

2.3.2 迭代过程中类中心的变化量

表 2 是 K-Means 聚类分析的迭代历史过程。从表中可以看出总共进行了 3 次迭代 (Iteration)。第一次迭代后形成的类中心点和初始类中心点的距离 (Change in Cluster Centers) 分别为: 12.659 18.415 0.000 8.642 14.953 直到第 3 次迭代后类中心点没有发生变化或者变化很小, 距离为 0.000 其中初始

类中心点之间的最小距离为 50.336 即 K-Means 聚类分析经过 3 次迭代完成。

表 2 K-Means 聚类分析的迭代历史过程表

类别	1	2	3	4	5
1	12.659	18.415	0.000	8.642	14.953
2	2.519	11.843	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000

2.3.3 最终的类中心变量值

表 3 K-Means 聚类分析最终的类中心点位置表

类别	1	2	3	4	5
进球数	1.18	1.73	2.25	0.64	0.33
射进指数	9.73	14.15	13.04	6.97	4.15
助攻次数	0.79	1.44	1.00	0.40	0.22
射正球门	5.73	6.63	8.75	4.05	2.89
射门次数	12.22	12.83	17.25	9.75	8.55
角球次数	5.28	6.03	9.25	4.16	3.22
任意球次	1.02	0.98	1.25	0.87	0.89
越位次数	3.12	3.50	0.75	2.83	2.56
反击次数	4.80	4.03	5.75	3.80	3.89
短传次数	323.82	385.37	453.50	262.45	197.13
长传次数	97.41	90.90	93.00	93.14	85.37
传中次数	22.49	19.97	24.50	20.37	19.43

表 3 为 K-Means 聚类分析最终的类中心点位置。该表格和 SPSS 输出的第一个表格相比, 中心点位置有了一些变化, 表明迭代过程中, 中心点位置有了转移。

2.3.4 最终的类中心点之间的距离

表 4 K-Means 聚类分析最终的类中心点之间的欧氏距离表

类别	1	2	3	4	5
1	0.000				
2	62.126	0.000			
3	130.036	68.657	0.000		
4	61.712	123.238	191.493	0.000	
5	127.526	188.699	256.991	65.874	0.000

表 4 是 K-Means 聚类分析最终的类中心点之间的欧氏距离。从中可以看出第 3 类类中心点和第 4 类类中心点之间的距离最大, 为 191.493 第 1 类类中心点和第 2 类类中心点之间的距离最小, 为 62.126

2.3.5 各进攻指标间的方差分析

表 5 K-Means 聚类分析后形成的各样本之间的方差分析表

类别	组间平方和	自由度	组内平方和	自由度	F	Sig.	P
进球数	1.567	4	0.121	27	12.960	0.000	$P < 0.05$
射进指数	53.754	4	12.413	27	4.331	0.008	$P < 0.05$
助攻次数	0.855	4	0.094	27	9.107	0.000	$P < 0.05$
射正球门	12.954	4	1.563	27	8.286	0.000	$P < 0.05$
射门次数	25.486	4	5.413	27	4.708	0.005	$P < 0.05$
角球次数	9.676	4	2.102	27	4.604	0.006	$P < 0.05$
任意球次	0.062	4	0.224	27	0.275	0.891	$P > 0.05$
越位次数	1.699	4	1.608	27	1.056	0.397	$P > 0.05$
反击次数	2.239	4	4.287	27	0.522	0.720	$P > 0.05$
短传次数	24891.49	4	330.252	27	75.371	0.000	$P < 0.05$
长传次数	108.578	4	144.780	27	0.750	0.567	$P > 0.05$
传中次数	13.915	4	23.832	27	0.584	0.677	$P > 0.05$

表 5 是 K-Means 聚类分析后形成的各样本之间的方差分析表。从中可以看出: 各类间在进球数、射进指数、助攻、射正

球门次数、射门次数、角球次数、短传次数等 7 项进攻指标上均存在 $P < 0.05$ 结果为差异具有显著性,说明弱队和强队之间的差距主要集中在该 7 项指标上。由此可见,教练员应该在这 7 项指标上多下功夫,特别是在平时的训练和比赛中重点强调和分析,来提高球队的进攻能力。因此从表 5 中各类的因素方差分析来看,将样本分成 5 类的 K—Mean 聚类分析基本上是

成功的,聚类效果比较理想。同时也表明 K—Mean 聚类结果具有较高的可靠性和科学性。

3 结果与分析

3.1 结果

表 6 聚类结果统计一览表

队名	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	类别
意大利	1.71	14.46	1.29	6.86	11.86	6.29	1.14	4.86	5.71	322.4	101.6	20.0	1
法国	1.29	11.69	0.71	5.00	11.00	5.86	0.29	4.12	4.00	324.7	79.0	23.6	1
德国	1.86	11.50	1.14	7.04	16.14	4.72	1.14	3.00	5.57	341.7	117.3	28.9	1
葡萄牙	1.00	7.00	0.71	7.29	14.29	6.29	1.14	2.71	5.29	363.9	96.3	23.7	2
阿根廷	2.20	20.75	2.00	5.20	10.60	5.60	0.60	4.80	1.80	384.6	102.0	18.6	2
巴西	2.00	14.71	1.60	7.40	13.60	6.20	1.20	3.00	5.00	407.6	74.4	17.6	2
英格兰	1.00	7.46	0.60	5.80	13.40	6.00	0.80	2.00	2.60	320.0	111.0	26.8	1
乌克兰	1.00	8.93	0.60	5.20	11.20	3.80	0.60	2.60	3.40	288.8	118.4	23.8	4
西班牙	2.25	13.04	1.00	8.75	17.25	9.25	1.25	0.75	5.75	453.5	93.0	24.5	3
厄瓜多尔	1.25	12.5	1.25	4.50	10.00	4.25	1.25	2.75	2.25	339.0	93.8	14.8	1
墨西哥	1.25	10.87	1.25	5.25	11.50	5.50	1.00	1.25	3.25	318.8	95.8	26.5	1
澳大利亚	1.25	9.26	0.50	6.25	13.50	5.00	1.00	2.25	2.00	334.8	90.0	26.8	1
瑞士	1.00	8.70	0.75	6.25	11.50	5.50	1.50	2.00	4.75	355.0	101.3	23.0	1
加纳	1.00	6.56	0.50	5.75	15.25	4.25	0.50	6.25	6.75	327.8	88.3	19.3	1
荷兰	0.75	6.00	0.50	6.50	12.50	4.75	1.75	3.00	3.75	318.8	102.8	20.0	1
瑞典	0.75	6.12	0.75	6.00	12.25	7.50	0.75	1.75	4.25	268.0	95.0	31.5	4
科特迪瓦	1.67	10.20	0.67	7.67	16.33	7.67	0.67	3.67	9.00	334.0	90.7	26.3	1
哥斯达黎加	1.00	10.71	0.67	3.67	9.33	3.00	1.00	3.00	4.67	303.7	97.3	16.3	1
捷克	1.00	8.57	0.67	5.67	11.67	5.00	0.67	4.67	5.33	302.3	108.0	22.0	1
韩国	1.00	8.33	0.67	5.34	12.00	3.67	1.33	2.00	3.67	315.0	106.3	21.0	1
突尼斯	1.00	15.79	0.33	2.67	6.33	2.33	1.00	4.00	3.33	248.7	98.0	12.0	4
沙特	0.67	7.69	0.67	3.00	8.67	3.33	2.00	2.00	1.67	268.3	78.3	12.7	4
波兰	0.67	9.09	0.67	4.33	7.33	7.67	1.33	2.00	8.67	299.3	78.0	22.0	1
伊朗	0.67	6.67	0.33	6.33	10.00	3.00	0.33	2.33	1.67	247.0	78.0	23.0	4
日本	0.67	7.41	0.33	3.33	9.00	3.67	1.33	2.67	6.00	292.7	89.7	16.7	4
塞黑	0.67	9.52	0.33	2.67	7.00	3.67	1.67	1.00	3.00	189.0	73.7	22.3	5
克罗地亚	0.67	6.06	0.00	4.00	11.00	7.33	0.67	4.00	6.67	236.7	81.7	25.7	4
安格拉	0.33	2.94	0.33	3.67	11.33	4.33	1.00	4.67	6.33	216.7	89.7	20.3	5
巴拉圭	0.33	2.56	0.33	4.67	13.00	3.67	0.67	2.33	0.67	265.3	93.0	15.3	4
多哥	0.33	3.70	0.33	4.00	9.00	3.00	1.00	4.33	5.00	231.0	90.0	18.0	4
美国	0.33	4.76	0.33	1.33	7.00	4.00	0.33	2.33	5.33	278.0	109.3	25.0	4
特立尼达和多巴	0.00	0.00	0.00	2.33	7.33	1.67	0.00	2.00	2.33	185.7	92.7	15.7	5

注: x_1 = 进球数; x_2 = 射进指数; x_3 = 助攻次数; x_4 = 射正球门次数; x_5 = 射门次数; x_6 = 角球次数; x_7 = 任意球次数; x_8 = 越位次数; x_9 = 反击次数; x_{10} = 短传次数; x_{11} = 长传次数; x_{12} = 传中次数; 以上指标均为场均次数。

表 6 为 K—Mean 聚类最终结果,从中我们可以看出:本次聚类结果第 1 类有 15 支球队,占总参赛队的 46.875%;第 2 类共有 3 支球队,占总参赛队的 9.375%;第 3 类仅有 1 支球队,占总参赛队的 3.125%;第 4 类球队共有 10 支,占总参赛队的 31.25%;第 5 类共有 3 支球队,占总参赛队的 9.375%。在前 4 名球队中有前 3 名队伍意大利、法国和德国在第 1

类,第4名葡萄牙在第2类;在第5—8名队伍中第7名英格兰在第1类,第5名阿根廷以及第6名巴西在第2类;第8名乌克兰在第4类;在第9—16名队伍中,只有第9名西班牙在第3类以及第16名瑞典在第4类外,其余均在第1类;16强之后的球队只有科特迪瓦、捷克、韩国以及波兰在第一类外,其余都在第4—5类。以上K-Mean聚类结果表明,聚类结果所反映出的进攻趋势与比赛结果也基本吻合,说明快速攻防转化,进攻能力强的球队取得的成绩相对较好。相反,进攻能力差的球队则难以取得好成绩,但是在比赛中攻防是相互转化的,在这里我们研究更多的是进攻,对防守不做过多的分析。虽然各个国家队都有自己的技战术风格,足球比赛也有很大的偶然性,但是攻势强大、射门次数多、射进指数高、助攻次数多以及射正球门次数越多的球队其胜率也就相应越高。进攻之优劣主要反映在进球数、射进指数、助攻、射正球门次数、射门次数、角球次数、短传次数等7项进攻指标上,次数多、成功率高者占有优势。就进攻结果而言,射门得分则具体体现在进攻队员创造机会、把握射门机会的射门能力上,但这种效果最终还是全队整体进攻量的积累、质的升华,是球队进攻能力的集中体现。

3.2 分析

表6表明聚在第1类的15支球队,以意大利队、法国队、德国队等为代表,统计表明,这些球队的攻击力是比较强大的,进攻质量非常高,前锋善于把握射门的机会,也能自己制造或为同伴创造射门机会;前卫的助攻能力强,在中前场就开始组织稳固的防守,就地进攻而压迫对手,同时也加快了攻防转换速度,提高了进攻质量;后卫队员在稳固好自己的防守区域的同时,加强了与队友之间的相互协防,寻找恰当的时机进行插上进攻,与前卫队员、前锋队员进行相互之间的配合,为中前场队员创造进攻机会,这样三条线之间攻防形成了一体,缓解了后防线的压力,给全攻全守的足球发展趋势注入了新的活力。从总体数据统计上也显示,这些球队的进攻能力是非常有特点的,攻击力也是比较强的。尤其是冠军队——意大利队,他的防守体系非常稳固、坚实,进攻犀利且善于快速反击。

第2类的3支球队,这三支球队为葡萄牙队、阿根廷队以及巴西队,分别位居第4名、第5名以及第6名,从进攻数据的总体上来看,这些球队的攻击力是非常强大的,其进攻能力也是整体性的,同时也具有稳固的防守能力。聚在该类的球队都是世界传统的强队,球员的个人技术非常娴熟细腻,有良好的控球能力,尤其是巴西队和阿根廷队,它们的队员的过人突破能力、随机应变能力、短传渗透能力威胁很大,属于典型的技术流派。

第3类的1支球队西班牙队,从进攻数据的来看,这在本届世界杯上全部球队的进攻能力上来说是很强的,这与他在小组赛中的对手比较弱有一定的关系,由于过早的遇上了法国队,没有进入前八强也在情理之中。统计结果也表明,聚在该类的西班牙队的进攻质量非常高,前锋把握射门机会的能力强,也能自己制造或为同伴创造射门机会;但是中后场的防守不够稳固,防守漏洞太多,经不住法国队的猛烈冲击,从西班牙和法国队这场比赛的全过程来看,西班牙在进攻能力上还占据一定的优势,但是法国的防守更为坚固,强强对抗不止看谁的进攻能力强,更看谁的防守破绽多,可以说本场比赛谁输谁赢都很正常。

第4类的10支球队,以乌克兰队、瑞典队、沙特队为代表,这类球队中多为足球的不发达国家,从聚类结果来看,它们的进攻有一定的优势,进球较少且射进指数较低,说明进攻质量不高。录像观察发现,该球队的进攻多集中在中场附近,很难攻入对方防守的危险地带,缺乏有效的得分手段。

第5类的3支球队,除了塞黑队之外,安哥拉队与特立尼达和多巴队两队都是第1次进入世界杯,属于足球的落后国家,但是能打进世界杯足球赛,这也说明他们还是有一定的实力的,塞黑队本来还是有一定的足球基础的,但是在小组赛中就遇到了阿根廷队和荷兰队,使得进攻能力得不到完全发挥,过早的被淘汰也不足为怪。从数据上看,整体进攻能力较弱,进球不多,射进指数也不高,进攻质量低下,与世界足球强队有一定的差距,很难与之抗衡。

4 结论

1)本文提出的利用K-Mean聚类分析,对32支球队的进球数等12项进攻能力指标进行的聚类分析,与比赛的结果也有吻合之处,这表明应用K-Mean聚类分析法对综合评价球队的整体进攻能力具有很大的使用价值,是一种可靠的、合理有效的、较为客观的、快速量化评价球队进攻能力的方法,且这种方法也很容易推广到其他体育项目的综合评价中。

2)对第18届世界杯足球赛32支球队64场比赛的进球数等12项进攻指标进行K-Mean聚类分析,结果显示各类间在场均进球数、射进指数、助攻、射正球门次数、射门次数、角球、短传次数等7项进攻指标上差异具有显著性,弱队和强队之间的差距主要体现在该7项指标上。

3)以意大利队、法国队为代表的第1类球队进攻质量非常高,进攻时非常犀利且善于快速反击;第2—3类的球队他们的进攻能力是非常强大的,但是重攻轻守,很难走到最后;第4类的球队的进攻能力有一定的优势,但进攻质量不高,缺乏有效的得分手段;第5类的球队整体进攻能力较弱,与世界强队有一定的差距。

参考文献:

- [1] 刘涛. 体育统计学[M]. 北京: 体育学院出版社, 1990: 63—97.
- [2] 余建英, 何旭宏. 数据统计分析与SPSS应用[M]. 北京: 人民邮电出版社, 2003: 251—282.
- [3] 张力为. 体育科学研究方法[M]. 北京: 高等教育出版社, 2004: 57—158.
- [4] 奚天明. 第十三届亚洲杯足球赛进攻行动的统计与分析[J]. 北京体育大学学报, 2005: 133—136.
- [5] 张冉, 张雪临. 第16届世界杯足球赛进攻指标的Q型聚类分析[J]. 中国体育科技, 1999: 35(11): 16—18.
- [6] 麻雪田, 王崇喜. 现代足球运动高级教程[M]. 北京: 高等教育出版社, 2003: 1—7.
- [7] The official site for the 2006 FIFA World Cup Germany [EB/OL]. <http://fifaworldcup.yahoo.com>
- [8] 李平. 从第11、12届欧洲足球锦标赛析现代足球进攻战术发展趋势[J]. 成都体育学院学报, 2005(5): 77—80.
- [9] 亚洲足球联合会. 亚洲足球教练员C级培训教程(第二版)[M]. 北京: 人民体育出版社, 2005: 115—118.
- [10] 许在华, 袁野. 对第13届亚洲杯足球赛日本队传球特点及进攻节奏分析[J]. 中国体育科技, 2005(5): 85—87.