

基于机器学习的足球比赛数据特征研究

高 婕, 李 煜, 毛姝婷, 陈子涵, 叶瑶欣

杭州师范大学, 浙江 杭州
Email: 1162209997@qq.com

收稿日期: 2021年8月14日; 录用日期: 2021年9月6日; 发布日期: 2021年9月16日

摘 要

为研究足球比赛数据特征, 本文通过机器学习对足球赛场数据进行功能性分析, 特别是通过社会网络分析, 量化各球员在每场球赛的表现以及团队合作的效果。本文先考虑个人对团队的贡献, 对13个子指标归纳为四个大指标, 综合优序图法和独立性权重法构建网络球员个人贡献模型并提出建议。为分析团队合作的效果, 本文选取15个相关指标进行筛选并进行回归分析, 最终确定影响比赛结果的6个显著性因素, 并给出结论和建议。

关键词

足球比赛表现研究, 机器学习, 社会网络分析

Research on Data Features of Football Matches Based on Machine Learning

Jie Gao, Yu Li, Shuting Mao, Zihan Chen, Yaoxin Ye

Hangzhou Normal University, Hangzhou Zhejiang
Email: 1162209997@qq.com

Received: Aug. 14th, 2021; accepted: Sep. 6th, 2021; published: Sep. 16th, 2021

Abstract

In order to study the characteristics of football game data, this paper uses machine learning to perform functional analysis of football matches data, especially through social network analysis, to quantify the performance of each player in each match and the effect of teamwork. This paper first considers the individual's contribution to the team, summarizes the 13 sub-indices into four major indicators, integrates the optimal sequence diagram method and the independence weight method to construct the individual contribution model of the network player and makes recom-

mentations. In order to analyze the effect of teamwork, this paper selects 15 relevant indicators for screening and regression analysis, and finally determines 6 significant factors that affect the results of the matches, and gives conclusions and suggestions.

Keywords

Football Matches Performance Research, Machine Learning, Social Network Analysis

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 研究背景

1.1. 问题的提出

团队合作随着社会之间的联系越来越紧密,寻找最佳的团队策略也越来越复杂。随着传统学科、产业的数字化转型,通过机器学习的方式对团队过程进行功能性分析具有较大潜力。在足球领域中,持续的数字化和传感器技术领域的发展导致数据量迅速增加。

在一场足球比赛中,可以得到球赛的场进失球数,球员的射门数、射正率、进球率、绝佳机会、GOA 等数据,这些数据可以反映比赛时团队合作的绩效指标,例如:输赢、分数、球员和团队表现之间的平衡程度、球队在一段时间的协调能力等。根据直接这些指标可以更准确的分析一场球赛的比赛策略、团队配合、球员状态的不足,和预测一场比赛的团队配合、某个球员的行为等。我们的工作主要有选取不同联赛的球队比赛数据,基于机器学习分析足球比赛的数据特征,并结合比赛的结果、球员的身价等进行分析。

通过分析比赛数据,可以探索一场球赛中球员之间的复杂互动如何影响他们的成功(得分、团队积极性和协调能力等)。团队的成功不仅仅是个人能力的总和,而是基于许多其他因素。通过分析大量比赛数据,建立复杂团队网络与绩效指标之间的关系,可以量化各球员在每场球赛的结构和动态特征,确定每场各球员综合能力评分,从而改善球队在团队合作的策略。

1.2. 研究方法

1) 数据采集

要想对比赛结果进行预测,肯定要先获取比赛球队的各种数据信息,利用网络爬虫是通过编程语言的实现对网站发出请求,并且用迭代或者递归的函数去对网络上的数据,图像,文字以及其他内容进行抓取。本文利用 2020 年美赛 D 题哈士奇队(英超埃弗顿足球俱乐部)爬取的相关数据[1]。

2) 数据可视化

利用 NetDraw 等软件对各时间节点传球网络等进行可视化。网络中的节点表示参与者和的数量,节点对应于球员号码。节点圆的大小取决于其节点度,即传球数量越多,圆越大。此外,节点度对应于节点的度中心性,表示的重要性球员与传球次数呈正相关。因为通过是单向的,链接是单向的,意味着从一个球员传递到另一个球员。宽度表示权重,即此链接中包含的传递次数。次数越多,联系越广。每个节点的位置取决于它的平均位置的平均传球球员的坐标[2]。

以第一场比赛为例,我们可以给出哈士奇和对手队的传球网络图,如图 1 和图 2。

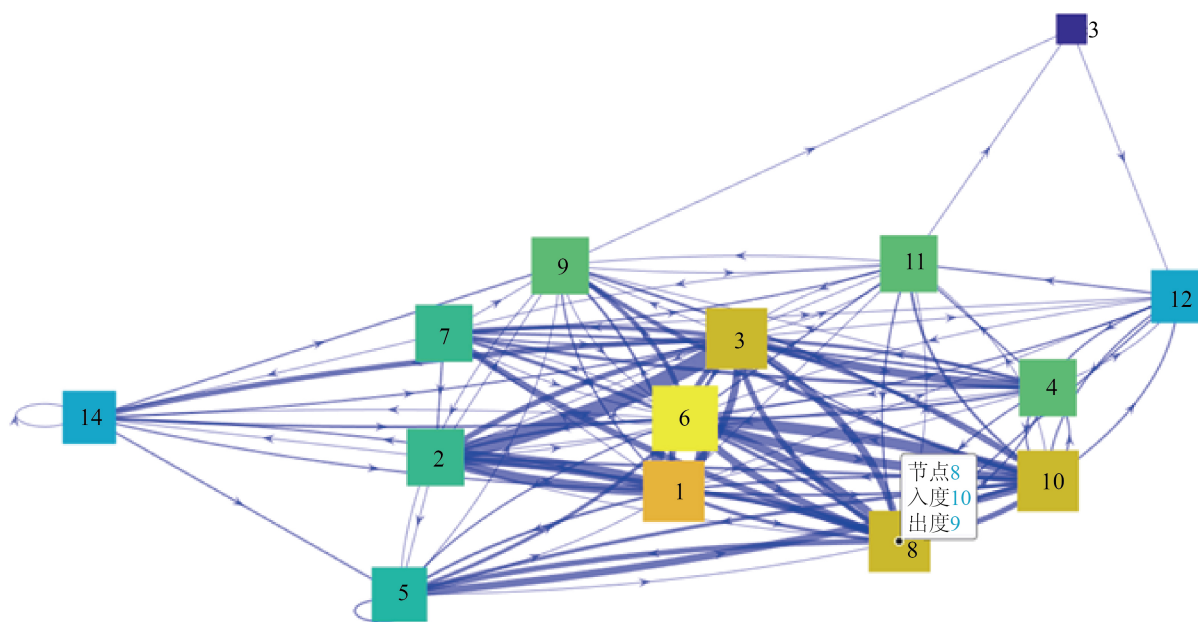


Figure 1. Huskies passing network map
图 1. 哈士奇队传球网络图

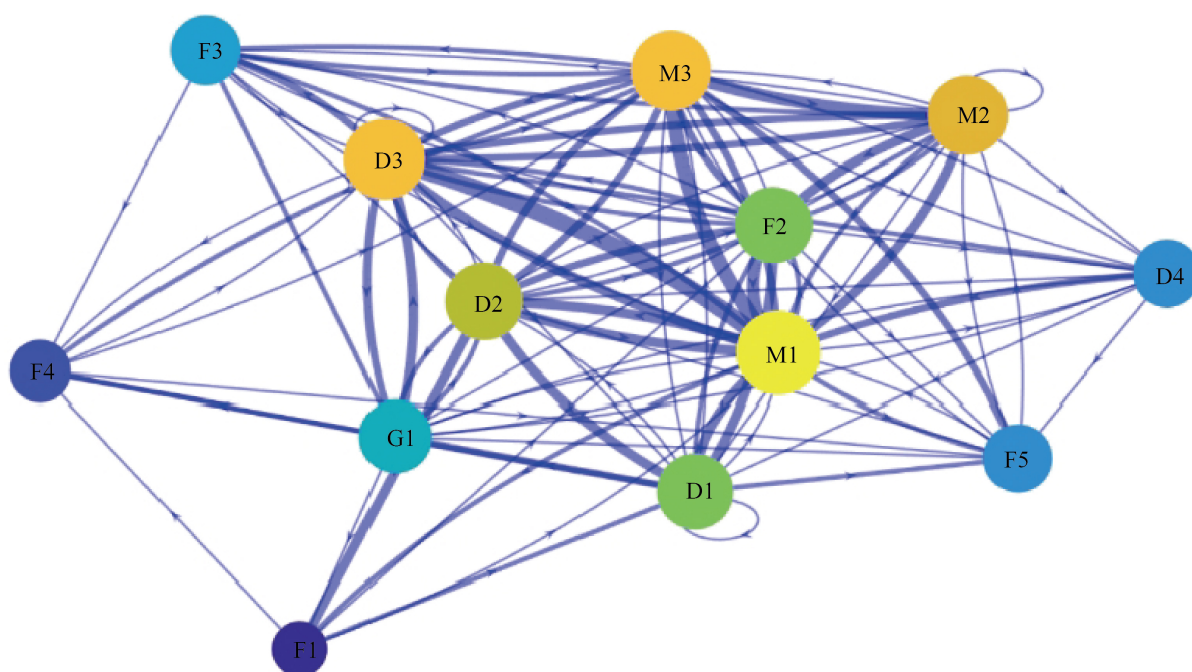


Figure 2. Opponent pass network diagram
图 2. 对手传球网络图

3) 社会网络分析

① 确定整体网，在球队的球员之间构建一个“球流”的定向网络。在该网络中，节点代表球员，并且根据两个球员之间成功完成的通过次数对弧线进行加权。球员的节点通过根据传球次数加权的弧线连接到这两个节点。

- ② 分析具体球赛的数据，计算各时间各球员节点的加权邻接矩阵，中心性和最大半径等。
- ③ 确定网络模式，包括核心、外围设备，进行分析以获得核心球员和外围球员及其关系，以及分析构造三节点子网的配置以获得类型和其图案发生的频率，即二元和三元的特定配置。
- ④ 根据以上指标分别量化球员个人特征以及团队合作的成效[3]。

2. 球员个人贡献模型

2.1. 团队配置

团队配置有二元和三元配置。二元配置是指两个玩家之间的交互，三元配置是指三个玩家之间的交互。为了更好地分析玩家之间的具体传球模式，考虑三个节点的子网。子网中三个节点成对形成二元配置，一起形成三元配置[4]。

2.2. 指标

1) 度中心性

此指标可用来衡量每个球员活动量。球员的中心性越高，与队友的联系就更多。

设 n_i 是加权有向图 G 的 n 个球员位置顶点，标准化中心度指标 $C_{(D-out)}^w(n_i)$ 是比例与相邻的顶点的权重百分比，计算公式为：

$$C_{(D-out)}^w(n_i) = \frac{k_i^{w-out}}{\sum_{i=1}^n \sum_{j=1, j \neq i}^n a_{ij}}$$

其中 k_i^{w-out} 是顶点的度中心指数， n_i 和 a_{ij} 是 G 的加权邻接矩阵的元素。

2) 亲近中心性

此指标量化了球员之间的接近程度。中心得分越高，表明球员可以在更少的传球中接触更多的队友。

给定一个带 n 个顶点的加权有向图 G 。顶点 n_i 的标准化接近度指数的计算公式如下：

$$C_{(C)}^w(n_i) = (n-1) \times \left[\sum_{\substack{j=1 \\ j \neq i}}^n d^w(n_i, n_j) \right]^{-1}$$

其中 $d^w(n_i, n_j)$ 是 n_i 和 n_j 之间的测地距离。

3) 压力中心性

此指标量化团队运动分析中节点的压力，例如向队友的传球，可以表明球员在功能上能够将互动球员聚集在一起的相关性。压力中心值越高，球员在连接进攻方面的相关性越高。

给定一个加权有向图 $G=(V, E)$ ，其中 $n_i, n_j, n_k \in V, i, j, k=1, \dots, n$ 。应力中心指数的计算公式为：

$$C_b(n_k) = \sum_{\substack{n_i, n_j \in V \\ i \neq n_j, j \neq k}} g_{ij}(n_k)$$

其中 $g_{ij}(n_k)$ 是 n_i 和 n_j 之间通过 n_k 的最短路径数。

4) 中介中心性

在传球网络中，BC 得分较高的球员可能是那些更经常处于队友之间的球员。例如，BC 得分高的球员在将球传给其他人时可能很重要。

给定一个加权有向图 $G=(V, E)$ ，其中 $n_i, n_j, n_k \in V, i, j, k=1, \dots, n$ 。中介中心性的计算公式为：

$$C'_b(n_k) = \frac{1}{(n-1)(n-2)} \sum_{\substack{n_i, n_j \in V \\ i \neq n_j, j \neq k}} \frac{g_{ij}(n_k)}{g_{ij}}$$

其中 $g_{ij}(n_k)$ 是 n_i 和 n_j 之间通过 n_k 的最短路径数, g_{ij} 是 n_i 和 n_j 之间的最短路径数。

5) 偏心中心性

此指标可视为队球员网络中距离最远的队友之间的距离。相较于网络的平均偏心率, 一个高偏心率的球员将更容易受到其他队友的影响, 或者相反, 很容易影响其他几个球员们。

给定一个具有 n 个顶点的加权有向图 G , 顶点 n_i 的偏心率中心性 $C_{ecc}(n_i)$ 由下式确定

$$C_{ecc}(n_i) = \frac{1}{\max\{d(n_i, n_j)\}}$$

其中 $d(n_i, n_j)$ 是顶点 n_i 和 n_j 之间的最短路径。

6) 度权威

中心性得分越高, 表明队友更愿意向此球员传球。他们在球队进攻时比其他球员更经常接到球。

给定一个具有 n 个顶点的加权有向图 G , 令 n_i 为图 G 的一个顶点, 标准化程度声望指数 $P_{(D-in)}^w(n_i)$ 是与 n_i 相邻的顶点的权重比例, 计算公式为

$$P_{(D-in)}^w(n_i) = \frac{k_i^{w-in}}{\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}}$$

其中 k_i^{w-in} 是顶点 n_i 的度威望指数, a_{ij} 是 G 的加权邻接矩阵的元素。

7) 邻近权威

邻近权威表明球员与其他队友的距离。如果一个球员更接近他们的队友, 这意味着在要传球的情况下, 队友更倾向于和他一起比赛。

给定一个具有 n 个球员位置的加权有向图 G , 令 n_i 为图 G 的一个顶点。顶点 n_i 的邻近权威指数 $P_p(n_i)$ 是可以到达 n_i 的顶点与这些顶点到 n_i 的平均距离的比例, 由下式确定:

$$P_p(n_i) = \frac{\frac{I_i}{n-1}}{\sum_{j=1, j \neq i}^n d(n_i, n_j)} \cdot I_i$$

其中 I_i 是直接或间接连接到 n_i 的顶点数; $d(n_i, n_j)$ 是顶点 n_i 和 n_j 之间的最短路径。

8) 特征向量中心性

特征向量较高的球员, 表明经常与几个重要的队友互动, 是一个中心调节角色。一个低特征向量的玩家, 比如一个外围队友, 只与少数队友互动, 而不是中心队友。

给定一个具有 n 个顶点的加权有向图 G 。顶点 n_i 的特征向量中心性指数定义为特征向量的第 i 个分量, \mathbf{x} 对应于下列特征方程 $A\mathbf{x} = \mathbf{x}\lambda$ 的最大特征值, 其中 A 是 G 的邻接矩阵。

9) 子图中心性

此指标表明了网络中所有子图中每个球员的参与度。子图中心性根据在特定球员处开始和结束的封闭步数来评定球员的重要性。步数最短的球员对中心性的影响最大。

设 G_s 是 n 个顶点的加权有向图 G 的子图。顶点 n_i 的子图中心性定义为

$$C_{G_s}(n_i) = \sum_{k=0}^{\infty} \frac{u_k(i,i)}{k!}$$

其中 $u_k(i,i)$ 是顶点 n_i 参与 G 的长度为 k 的封闭步数。

10) 定向信息中心

此指标可衡量球员分析传球能力，定向信息中心得分越高，球员分析传球能力越强。

给定一个具有 n 个顶点的加权有向图 G 。顶点 n_i 的有向信息中心性指数是所有 n_i 上平均有效阻力 $r(n_i, n_j)$ 的倒数，计算公式如下：

$$DI_C(n_i) = \left(\frac{1}{n} \sum_{j=1}^n r(n_i, n_j) \right)^{-1}$$

11) 算法中心性

算法中心性表明每个球员在合理数量的传球后他将获得球的概率，它遵循“如果一个球员从其他受欢迎的球员那里得到传球就会受欢迎”的原则。

给定一个具有 n 个顶点的加权有向图 G 。顶点 n_i 的中心性指数 $PR(n_i)$ 由下式确定

$$PR(n_i) = p \sum_{j \neq i} \frac{a_{ij}}{k_j^{out}} PR(n_j) + q$$

其中 a_{ij} 是邻接矩阵的元素， p 是一个启发式参数，表示这个顶点与其他顶点连接的概率， q 是一个参数，为每个顶点授予“免费”流行度。

12) 权力中心性

权力中心性表明与更多队友互动的球员被认为在网络活动期间更具有相关性。给定一个具有 n 个顶点的加权有向图 G 。顶点 n_i 的幂中心性指数 $P_C(n_i)$ 由下式确定

$$w_{ij} = \delta \sum_{m=1}^{\infty} b^m (a^{m+1})_{ij}$$

$$P_C(n_i) = \sum_j w_{ij}$$

其中 w_{ij} 是 G 和 δ 的邻接矩阵的元素； b 是常数。

13) 质心

具有较高平均值的网络质心值的球员，组织能力越强。

给定一个具有 n 个顶点的加权有向图 G 。顶点 n_i 的质心中心指数 $C_{Ce}(n_i)$ 由下式确定

$$C_{Ce}(n_i) = \min \{ f(n_i, n_j) : n_j \in V - \{n_i\} \}$$

其中 $f(n_i, n_j) = \gamma_{n_i}(n_j) - \gamma_{n_j}(n_i)$ ， $\gamma_{n_i}(n_j)$ 是离 n_i 比离 n_j 更近的顶点数，也就是，

$$\gamma_{n_i}(n_j) = \left| \left\{ n_k \in V : d(n_i, n_k) < d(n_j, n_k) \right\} \right|。$$

通过数据清洗、组合处理与挖掘，将数据应用至上述指标计算公式中，得到 14 个指标的对应结果。考虑每个指标的衡量角度及依据，综合各项指标的计算公式，本文将 13 个小指标等权归类于四个大指标，分别为导向性、凝聚性、协调性、受欢迎性，分类结果如表 1 所示。

由于同类小指标存在信息重复，故本文将每类小指标进行标准化后等权计算，作为大指标的值。

此表默认行为前者，列为后者。

能够计算出四项指标得分及权重值如表 3 所示：

Table 3. Index score and weight table

表 3. 指标得分及权重表

	导向性	凝聚性	协调性	受欢迎性
指标得分和	1.5	2	1.5	1.5
权重值	0.231	0.307	0.231	0.231

2.3.2. 独立性权重法在模型中的应用

每个球员的动作对结果的差异，在一定程度上取决于球员在某一位置执行同类动作的差异。基于此原理，独立性权重法可以客观评价评分规则。独立性权重法步骤如下：

多元回归分析计算每项指标与其他指标的复相关系数 R 。

$$R = \frac{\sum (y - \bar{y})(\hat{y} - \bar{y})}{\sqrt{\sum (y - \bar{y})^2 \sum (\hat{y} - \bar{y})^2}}$$

权重系数 R 越大，某指标与其它指标的相关性很强，说明信息有着较大的重叠，意味着该指标的权重会比较低。取复相关系数的倒数 R 作为得分，再经过归一化处理得到权重系数。

$$\left(\frac{1}{R}\right)^* = \frac{\frac{1}{R} - \left(\frac{1}{R}\right)_{\min}}{\left(\frac{1}{R}\right)_{\max} - \left(\frac{1}{R}\right)_{\min}}$$

经过上述复相关系数计算与归一化处理，以 forward 位置的权重为例，如表 4 所示：

Table 4. Independence weight method forward position player weight table

表 4. 独立性权重法 forward 位置球员权重表

权重	导向性	凝聚性	协调性	受欢迎性
forward	0.158	0.271	0.393	0.178

2.3.3. 评分权重的结合

我们分别得到了两个不同的评分权重。优序图法为主观，独立性权重法为客观，为增强模型的准确性，将其结合，对两种方法得到的权重进行等权平均计算。能够得出 forward 位置球员的权重评分方法，如表 5 所示：

Table 5. Forward position player weight score table

表 5. Forward 位置球员权重评分表

	权重
导向性	0.1945
凝聚性	0.289
协调性	0.312
受欢迎性	0.2045

为了帮助教练更好地了解和析每个球员的能力，我们使用上述模型，通过整个赛季的数据构建除守门员外的个人能力综合分析(由于球队只有一名守门员，并且鉴于守门员位置的特殊性，我们没有测量守门员的能力)。

因此，基于上述四个指标的权重值，能够计算出 forward 位置每位球员的贡献值，记 S_1 为 forward 位置计算出的球员贡献值， x_1 为球员导向性， x_2 为球员凝聚性， x_3 为球员协调性， x_4 为球员受欢迎性。易得计算公式为：

$$S_1 = 0.1945x_1 + 0.289x_2 + 0.312x_3 + 0.2045x_4$$

运用上述计算公式能够计算出 forward 位置球员的贡献值，部分数值如表 6 所示：

Table 6. Forward position player contribution table
表 6. Forward 位置球员贡献表

	贡献值
Huskies_F1	0.3462
Huskies_F2	0.5987
...	...
Huskies_F5	0.3747
Huskies_F6	0.7475

2.4. 网络球员个人贡献模型结论与建议

在现有的分析中，本文在球员网络模型中构建了网络属性与球员个人贡献指标。为了给哈士奇队的教练提出改进建议，我们将再次对上述指标进行分析。能够得出以下结论及相应的建议。

1) 我们给出了不同标准下所有球员的能力指标值，考察不同位置的球员，发现 F6 是最好的前锋，M13 是最好的中场，M3 是您球队中的明星球员 D8 是最好的后卫。他们可以构成球队的主力。其他球员可以考虑轮换，或者作为主力球员的替补上场，这也是比较理想的人员安排。因此，我们的能力值图为教练员的选择提供了有益的建议和评价。

2) 哈士奇队不同球员接到的传球次数差别很大，球员的位置相差也很大。注意球员之间的平衡，应该平等对待每一位球员。

3) 反击和长传不足，使得球员间不能够形成很好的链接。加强反击和长传技巧，发挥其优势。

4) 在训练中加入体能训练项目，帮助球员提高身体素质，支持更大范围的运动。

3. 团队模型

3.1. 指标

1) 总链接数

总链接数衡量的是比赛期间队友之间进行的总互动的绝对数量。因此，一场比赛总链接数越高，团队成员合作越紧密。该指数也可能与球员互动的频率相关，这可能会表现出长时间控球，以及强大的进攻或防守的团队合作[5]。给定一个带权的有向图 G ，其中 n 个顶点。 G 的总链接指数 L 计算公式为

$$L_D^w = \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}$$

2) 网络密度

团队网络密度是一个相对指标,也是衡量整体队友之间的合作的亲密度。在图论中,(有向)图的密度是节点之间存在的最大可能链接的比例。即给定一个带权有向图 G , 有 n 个顶点。 G 的密度指数计算公式为

$$\Delta_D^w = \frac{L_D^w}{n(n-1)}$$

3) 平均距离指数

球员之间的平均距离指数是一个重要的宏观特征。如果平均距离指数较大,可能是传球需要较长时间,也可能是一些球员没有参与传球,并受其他人的影响。球员与其他球员之间距离的差异可能是团队合作差异的原因。与队友更接近的球员可能比距离更远的球员能够发挥更大的力量。给定一个加权有向图 G , 其中 n 个顶点[6]。 G 的平均距离指数计算公式为

$$\overline{d^w} = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n d^w(n_i, n_j)$$

其中 $d^w(n_i, n_j)$ 是 n_i 与 n_j 之间的最短路径。

4) 网络直径

图的直径与玩家之间的距离有关。在图论中,图的直径是任何两个连接的玩家之间的最大距离(最大测地距离)。给定一个加权有向图 G , 其中 n 个顶点。 G 的网络直径计算公式为

$$D^w = \max_{i,j} d^w(n_i, n_j)$$

其中 $d^w(n_i, n_j)$ 是 n_i 与 n_j 之间的最短路径。

5) 网络异质性

异质性可用于衡量球员之间的连通性变化。已经发现许多复杂足球网络表现出近似的无标度拓扑,这意味着这些网络非常不均匀。在足球分析的情况下,更大的异质性值揭示了队友之间的非凝聚力互动过程。

令 G 为 n 个顶点的加权有向图表。顶点度指数 VAR 的变异系数计算为

$$CVD = \frac{\sqrt{VAR}}{\bar{k}}$$

VAR 是顶点度指数的方差, \bar{k} 是度数的平均值。

6) 中心化

中心化是衡量网络是否具有星形拓扑或网络节点是否平均具有相同连接性的度量。中心化数值越接近 1, 网络越有可能具有星形拓扑结构,更倾向于为同一球员进行传球。越接近于 0, 球员之间平均具有相同连接性就越大,这表示球员之间有更多的联系。

令 G 为 n 个顶点的加权有向图表。 G 的集体度中心性指数的变异系数 $CVGC_D$ 由下式获得:

$$CVGC_D = \frac{\sqrt{S_D^2}}{C_D} = \frac{\sqrt{\left[\sum_{i=1}^n (C_D(n_i) - \overline{C_D})^2 \right] / n}}{\sum_{i=1}^n C_D(n_i) / n}$$

令 G 为 n 个顶点的加权有向图表。 G 的集体中介中心性指数 GC_b 由下式获得

$$GC_b = \frac{\sum_{i=1}^n [C'_b(n^*) - C'_b(n_i)]}{n-1}$$

$C'_b(n^*)$ 是顶点 $n^* \in V$ 的最大标准化介数。

7) 传球速度

从时间的角度出发, 本文选取了在一场比赛中达到前 30 分钟传球次数来代表该球队的传球速度。时间越短, 球队在比赛开始后持球和快速传球的次数就越多。在一段时间后, 由于球队逐渐习惯了比赛, 传球速度会受到两队的各种因素影响, 因此本文不考虑后续的影响。本文将该指标表示为 V_{30} 。

8) 教练因素

教练可能会对球队比赛结果产生影响。coach 1 指导 9 场比赛, coach 2 指导 5 场比赛, coach 3 指导 24 场比赛。

9) 主客场因素

主客场可能会影响球队比赛结果。主场共有 19 场, 客场共 19 场。

10) 球队阵型选择

该场比赛中涉及多种球队阵型, 不同的球队阵型反映出不同球队的进球策略, 可能会影响队员进球的概率。因此本文将阵型数据化分析, 主要涉及到 4-5-1 阵型、4-4-2 阵型、4-3-3 阵型和 4-2-3-1 阵型, 用 A_1 、 A_2 、 A_3 、 A_4 来表示, 若该球队选择使用该阵型, 则即为 1, 否则即为 0。

11) 传球能力

一支球队的传球能力依靠每场比赛中中场人数除以核心球员人数的比值来衡量, 记为 COR 。

3.2. 筛选和检测指标

3.2.1. 相关指标分析

由于团队活动的成效受到了比赛时团队内部各因素影响, 选取比赛有关指标, 与比赛结果联系起来, 以此来筛选能够反映团队合作成效的指标。由于本模型建立在多项数据之上, 相较于根据赛场比赛模拟, 通过本指标结果建立分析更加有效。为消除各类指标(即自变量)对数据处理的影响, 本文将所有变量数据进行标准化。

通过数据分析表明, 主客场 S 、阵型选择 A_1 、 A_2 、 A_3 、 A_4 、总传球次数 L_D^w 、网络密度 Δ_D^w 、网络最大直径 D^w 、平均距离指数 $\overline{d^w}$ 和连接配置比 CVD , 而 COR 呈弱正相关。在所有 15 个自变量中, 只有在一场比赛中达到前 30 分钟传球次数 V_{30} 与因变量呈显著负相关。另外三个变量: 教练数据 C 、团队平均贡献 $CVGC_D$ 和加权标准来反映团队成员在整个比赛中的分散程度 GC_b 没有表现出明显的趋势, 说明它们与因变量之间存在线性相关。

由于 $CVGC_D$ 和 GC_b 来反映团队成员在整个比赛中的分散程度没有显著影响, 因此将它们从自变量中移除。建立多元线性回归模型来探究因变量与自变量之间的关系:

$$Goal = \alpha_1 + \alpha_2 L_D^w + \alpha_3 D^w + \alpha_4 \Delta_D^w + \alpha_5 \overline{d^w} + \alpha_6 CVD + \alpha_7 V_{30} + \alpha_8 COR + \alpha_9 A_1 + \dots + \alpha_{13} A_4 + \alpha_{14} S + \delta$$

由于数据代表两支足球队在同一场比赛中的差异, 我们采用 Robust 方法消除可能的误差。如表 7 (部分) 所示。

多元线性回归结果表明, 变量 S 、 A_1 、 A_3 、 V_{30} 和 CVD 通过了 t 检验, 而 L_D^w 、 Δ_D^w 、 D^w 的 p 值较大。其中 Δ_D^w 和 L_D^w 的回归系数为负, 与实际显著性不一致。考虑到因变量的胜利球与总传球次数的关系比较复杂, 两者之间可能没有明显的因果关系且可能存在相关性导致系数估计错误, 因此舍弃 L_D^w 和 Δ_D^w 变量。同时, A_2 、 A_4 、 $\overline{d^w}$ 、 COR 的 p 值大于 0.9, 因此本文将其舍去。

Table 7. Table of results of various indicators under linear regression**表 7.** 线性回归下各类指标结果表

得分	参数值	估计标准误	T 值	P 值($p > t $)	R 检验 = 0.4553	
					95%的置信区间	
L_D^w	-0.997915	0.619722	-1.53	0.133	-2.23245	0.323008
Δ_D^w	-0.795486	0.872644	-2.72	0.253	-6.83275	0.673023
D^w	0.397724	0.243772	1.65	0.112	-0.097557	0.875368
$\overline{d^w}$	-0.286743	0.724234	1.98	0.923	-0.098577	-0.757248
CVD	1.31886	0.554632	2.54	0.018	0.193442	2.44293
COR	-1.15489	0.337415	-3.5	0.012	-1.84901	-0.466476
V_{30}	0.011669	0.293356	0.05	0.078	-0.564976	0.569402
δ	-0.368421	0.245338	-1.69	0.153	-0.867992	0.131082

3.2.2. 相关指标检测

将 A_2 、 A_4 、 L_D^w 、 $\overline{d^w}$ 、 Δ_D^w 和 COR 去除后, 再次进行回归检验。可以验证得到第一次筛选出变量均通过了 t 检验。整体来看, F 值显著低于 0.05, 拟合优度值为 0.4422, 处于较高水平。计算三个变量的 VIF (方差膨胀因子), 结果如下表所示, 说明模型中自变量不存在多重共线性问题。因此可以进一步认为, 该模型中的指标可以从不同方面反映团队的合作水平。

由表 8 分析所示, 该模型整体回归效果良好。模型均方根误差值为 1.42 可以说明该模型可以较好分析球队的胜利球。也就是说, 哈士奇对在上赛季 38 场比赛的胜负可以通过该模型中的 6 个选定的指标进行有效预测。通过数据分析中 p 值表示变量大小的显著程度, 所以对于入选的六个指标, V_{30} 是最有效的预测团队合作过程的绩效指标, 其系数为负表示开场快节奏的传球往往对团队合作的全过程产生负面影响。上述结果均与直观数据相一致。

Table 8. Model's VIF calculation result table**表 8.** 模型的 VIF 计算结果表

变量	VIF	1/VIF
A_1	1.04	0.96
A_2	1.02	0.98
D^w	1.12	0.89
V_{30}	1.10	0.91
S	1.05	0.95
CVD	1.13	0.88
平均	1.076	

一个团队的策略主要包括人员的安置和沟通的安排。在所选指标中, D^w 代表团队的岗位安排, 连通性配置比例代表团队之间的沟通, 均有显著积极效果, 体现该模型合理性较高。因此, 该策略总体上是有效的, 无需考虑对手的反策略现象。比赛失败的原因可以用以上指标数据的不足来解释。

我们将阵型从四类扩展到六类, 并根据位置坐标、时间的精细程度进行细分。可以看出, 上下回归

的结果都通过了显著性检验。换句话说，逻辑回归模型是稳健的。

3.3. 策略与结论

1) 传球节奏

传球节奏是以时间为维度来观察在每场比赛中球队传球的表现。从微观层面上看，球队在比赛开始后前 30 分钟平均传球 108 次，在 20 支球队中排名第 7。哈士奇队在开场时通常会选择更快的步伐。但是从宏观层面上看，哈士奇队在 38 场比赛中场均传球排名第 10。这说明开场节奏对分差有负面影响，使得哈士奇队在下半场的传球次数上有所下滑，进而导致比赛失利的概率增加。

2) 球队阵型

不同阵型会造成比赛结果的不同。选用 4-5-1 的防守阵型和 4-3-3 的攻击阵型的时候，比赛更容易获得成功。研究发现，哈士奇队更善用防守反击的策略，但在中前场会遭到对手压制，因此需要提高进攻能力。

利用逻辑回归进一步考虑对手的阵型对哈士奇队队形选择和比赛结果的影响。当对手主要采用攻守平衡的阵型时，哈士奇队应选择 4-5-1 防守型，同时赢得比赛的概率是选择进攻型、攻守平衡型的 16.2%、4.3%。当对手主要采用攻守平衡的阵型时，哈士奇队更应选择平衡型阵型。

3) 主客场影响

主场更容易给球队带来积极的影响。

4) 教练指导

三位教练的指导对比赛结果的影响没有显著地差异，因此教练在比赛过程中的指导不能有效指挥球员的表现。教练应该在平时训练和赛前准备阶段对球员进行更加详尽的指导，促使球员在比赛场上能自主调整比赛状态和策略。

5) 团队紧密度

哈士奇队紧密度较高，会增加团队间传球的灵活性，因此可以提高在比赛中获胜的概率。聚类系数可以体现球队队员之间的合作效率，即提高队员之前传球的质量，而非仅提高传球次数。

6) 团队配合度

哈士奇队有大量的越位犯规，团队配合不足，使得进球效率和进攻效率不高，进攻方式较单一。需在后期比赛中，加强团队协作能力，减少犯规现象发生，从而提高比赛成功率。

综上所述，有多种因素会影响足球比赛的结果。

4. 模型推广

根据对足球比赛的结果，其中团队相关策略可以推广至其他的团队合作项目，列举如下：

4.1. 个人

1) 不同位置的球员有不同的计分系统，可以扩展到不同的绩效考核处理方式。采用不同的标准来评判团队中不同分工的人，即有利于组建更高效的团队。

2) 成绩的方差可以推广到个人贡献的差异。团队中每个人的贡献往往是异质的，有利于形成更有效的团队。

3) 我们可以参考模型的构建方法，建立综合评价体系，分析生产人员的相应绩效指标、预算完成率等，从而辅助管理决策优化系统的团队结构模型。为了测试模型，也可以对相关性进行统计分析。

4) 一个团队需要有一个核心成员来发挥核心领导作用，从一个团队的整体方向来看。此外，每个子

组还需要一个核心负责子组的工作。这样，团队可以更有效地工作，提高整个团队的协调性。

4.2. 团队

1) 团队紧密度的分析可以发现，球队队员之间的合作效率越高，会使得球队进球效率增加。因此，在其他团队项目中，成员之间须加强内部协作效率。

2) 对于集中式组织，人们长期适应某种管理和决策是，突然转变为其他管理形式是不利的到组织的绩效。因此团队需要事先适应好不同的场所，或增强对突发状况的适应能力。

3) 对内成员的犯规行为会影响团队得分情况，因此在团队比赛中需要增强规则意识，避免不必要的犯规行为发生。

4) 在比赛前期，教练需要做好比赛战略部署，要加强队员的配合度。并且对于一个团队来说，需要在平时培养内部信任度和凝聚力，因此需要增强领导者以及核心队员在团队中的威信。

4.3. 数据的处理和收集

我们不仅要收集与团队中绩效指标直接相关的人员或活动的数据，还要收集团队或团体整体运作的详细数据，例如其他人员的绩效和动态活动，从而准确把握企业整体战略的实施程度和效果。对人员和活动的把握要细致，便于后续分析。同时，通过对不同因素的敏感性分析，可以识别和关注建设中团队能力提升的重要方面，从而为团队建设和运营提供指导。

参考文献

- [1] 张岩. 社会网络分析在足球运动表现中的应用及研究进展[C]//中国体育科学学会. 第十一届全国体育科学大会论文摘要汇编. 北京: 中国体育科学学会, 2019: 3.
- [2] 李博, 王雷. 社会网络分析法研究足球比赛传球表现的可行性分析[J]. 北京体育大学学报, 2017, 40(8): 112-119.
- [3] 徐成博, 朱宏淼. 基于复杂网络的足球比赛中串联球队关键球员研究——以曼城俱乐部 2018-19 赛季数据为分析案例[J]. 体育与科学, 2020, 41(6): 101-105+110.
- [4] Zhao, T.G., Cui, N.N., Chen, Y.L., Li, M. and Li, J.X. (2020) Efficient Strategy Mining for Football Social Network. *Complexity*, 2020, Article ID: 8823189. <https://doi.org/10.1155/2020/8823189>
- [5] Clemente, F.M., Martins, F.M.L. and Mendes, R.S. (2016) *Social Network Analysis Applied to Team Sports Analysis*. Springer, Berlin. <https://doi.org/10.1007/978-3-319-25855-3>
- [6] Lusher, D., Robins, G. and Kremer, P. (2010) The Application of Social Network Analysis to Team Sports. *Measurement in Physical Education and Exercise Science*, 14, 211-224. <https://doi.org/10.1080/1091367X.2010.495559>