

基于机器学习的乒乓球混双技战术评估模型构建与应用

曹烨程^{1,2}, 张千轶³, 陈辉⁴, 刘敏¹

(1.山西大学 体育学院,太原 030006;2.石家庄学院 体育学院,石家庄 050035;3.温州医科大学 体育科学部,浙江 温州 325035;4.重庆第二师范学院 体育与健康管理学院,重庆 400065)

摘要:在运动项目技战术分析领域引入机器学习算法,以场胜负为标签,以各段与轮次的得分率与使用率构造的得分效率密度和全局得分能力两个新指标作为特征,使用机器学习模型中的随机森林模型、LightGBM模型、XGBoost模型对乒乓球混双比赛28个技战术指标的特征重要性进行排序,提取了2个二级评价指标和8个三级评价指标,最后应用构造的评估模型对王楚钦/孙颖莎与林高远/王曼昱的2场混双决赛进行了技战术分析与评价,验证了机器学习方法构建的乒乓球技战术评估模型具有较高的实用性和准确性。

关键词:机器学习;乒乓球;技战术;评估模型

中图分类号:G846;TP181

文献标志码:A

文章编号:1000-2367(2025)03-0149-08

竞技体育优异运动成绩的取得离不开科技支持.提高运动员在比赛中的竞技水平是训练的出发点和最终目标^[1].对于乒乓球项目来说,对运动员在比赛中的技战术评价对实现训练目标、纠正训练偏差、改进比赛策略有着重要作用.在乒乓球技战术评价模型构建的研究方面,具有代表性的是“三段指标”评价模型^[2],该模型在对运动员技战术进行评价方面,具有简洁和高效的特点,是其他技战术评价模型建构的基础.随着乒乓球运动的快速发展以及其他学科知识的融入,相继有学者提出了乒乓球技战术评价模型^[3-8].这些乒乓球技战术评价模型在某种程度上可以对运动员比赛中的技战术环节特征进行评价,发现技战术存在的优势和劣势,在相应的实践当中也具有一定的应用性和可行性.

但这些成果中所应用的乒乓球技战术评估模型是经过各个特征的样本差异计算出的权重予以赋权并得出评价模型,决定乒乓球比赛胜负的因素复杂且多元,在不参考比赛真实胜负的情况下,很难通过特征内部的差异性去计算和推断运动员技战术发挥对比赛胜负的影响程度.换句话说,已有研究成果的加权方法没有考虑评价指标与比赛胜负之间的关系,并没有在指标与被关注的衡量优劣的胜负之间建立映射关系,只依据指标内部各个样本的差异大小主观地赋予权重,也无法推断哪些评价指标更为优越,导致相同的比赛采用不同的技战术评估方法会得出不同的结果.为了解决这些问题,本研究依托不断革新的人工智能分析,借助机器学习的算法,建立乒乓球技战术评价指标与胜负的映射关系,在八轮次-三段法^[9]的基础上构建乒乓球混双技战术评价模型.这是首次尝试将人工智能引入竞技体育项目的技战术分析中,这也将是未来的乒乓球技战术分析领域乃至竞技体育领域研究的前沿与重点方向.

收稿日期:2024-07-10;**修回日期:**2024-10-09.

基金项目:国家社会科学基金(21XTY001);河北省体育局体育科技项目(2024JT09).

作者简介:曹烨程(1984—),男,河北石家庄人,山西大学在站博士后,石家庄学院副教授,博士,研究方向为球类比赛数据分析、运动技战术智能诊断,E-mail:caoyecheng66@163.com.

通信作者:刘敏,女,河北行唐人,山西大学教授,博士,博士生导师,研究方向为体育教育训练学,E-mail:lium@sxu.edu.cn.

引用本文:曹烨程,张千轶,陈辉,等.基于机器学习的乒乓球混双技战术评估模型构建与应用[J].河南师范大学学报(自然科学版),2025,53(3):149-156.(Cao Yecheng,Zhang Qianyi,Chen Hui,et al.Construction and application of table tennis mixed dual skill tactical evaluation model based on machine learning[J].Journal of Henan Normal University(Natural Science Edition),2025,53(3):149-156.DOI:10.16366/j.cnki.1000-2367.2024.07.10.0002.)

1 数据来源及指标选择

1.1 数据来源

从国际乒联官方网站和咪咕视频官方网站下载观看比赛视频,选取 2018—2022 年间奥运会、世锦赛、WTT 系列赛事等世界高水平混双比赛共 222 场,用于乒乓球技战术评估模型的建立,具体收集数据为胜负双方每回合的得失分数据.此外,选取王楚钦/孙颖莎和林高远/王曼昱的两场比赛的得失分数据作为评估模型的应用研究.

1.2 评价指标及计算方法

得分率代表一项比赛行为(包括技术或战术)的质,而使用率则代表一项比赛行为的量^[10].在三段评估法中,段使用率=(段得分+段失分)/(总得分+总失分),段得分率=段得分/(段得分+段失分).发接发轮次:轮次使用率=(轮次得分+轮次失分)/(总得分+总失分),轮次得分率=轮次得分/(轮次得分+轮次失分).本研究借鉴了技术效益指标^[11]的构建思路,以三段评估法与八轮次三段法中的评价指标为基础,构造了 2 个新指标:段(轮次)得分效率密度与段(轮次)全局得分能力.

$$\text{得分效率密度} = \lg \frac{\text{得分率}}{\text{使用率}}, \tag{1}$$

$$\text{全局得分能力} = (\text{得分率} - 0.5) \times \text{使用率}, \tag{2}$$

其中,得分效率密度(以下简称得分效率)代表的是段(轮次)在使用的过程中是否做到了合理使用,若数值过大则说明该段(轮次)得分率高,但是并没有足够的使用率将这种得分优势转化为比赛优势;反之则说明该段(轮次)无法得分但是却多次使用导致被对手得分.对数化一方面是因为该算法能将数值控制在一个范围内,有利于比较;另一方面是防止使用率为零的极端情况出现.若得分效率密度指标数值适中则说明技战术使用合理,此外若该指标得分在一定范围内偏高说明该段(轮次)对整局比赛有一定积极作用,反之则存在消极作用,这个指标主要体现了运动员在技战术使用安排上是否合理.

段(轮次)全局得分能力(以下简称得分能力)指标最直观地表现出段(轮次)对整局比赛的贡献.要衡量贡献最直观的做法是得分率乘以使用率,但有可能在某段中出现使用率较高而得分率较低的情况,于是将得分率减去 0.5 以后再乘以使用率就会有较好的效果.比如在第 1 场孙颖莎与伊藤美诚的比赛中,孙颖莎的发抢段得分率为 47%而使用率为 35%.在第 2 场孙颖莎与伊藤美诚比赛中,孙颖莎的发抢段得分率和使用率分别为 54%和 30%.如果直接用得分率乘以使用率来看得分能力的结果,第 1 场孙颖莎能力是 0.165,而第 2 场她的得分能力是 0.162,从分数上看孙颖莎第 1 场比赛得分能力高于第 2 场,但按实际效果来说,第 2 场孙颖莎接抢段的表现更为优秀.所以为了避免使用率高而得分率低而产生错误判断的现象出现,本研究将得分率减去 0.5 后再与使用率相乘,这样一方得分率不满足 50%的情况下,所得结果为负数,再次比较大小就可以得出不同结果.重新计算第 1 场比赛孙颖莎的得分能力为-0.010,而第 2 场的得分能力数值为 0.012,从数值上即可以说明孙颖莎在第 2 场比赛中接抢段表现更优.因此该指标以 0 为基准,数值越大则段(轮次)对比赛获胜的贡献越大,若小于 0 就说明该段(轮次)对比赛获胜起到反作用,若等于 0 则表示该段(轮次)对整局比赛的胜负作用不明显.它主要体现了运动员在某一段(轮次)的得分质量.

本研究将构造的 2 个新指标与三段评估法以及八轮次三段法相结合,按“男/女发男/女”的“得分效率/得分能力”构造 16 个八轮次指标;按“男子/女子”“发抢段/接抢段/相持段”的“得分效率/得分能力”构造 12 个三段指标.以此作为乒乓球混双技战术评价指标(附录表 S1).

2 应用机器学习方法构建技战术评价模型

2.1 乒乓球技战术评价模型的构建流程

基于机器学习模型的乒乓球技战术评价构建主要流程如下:数据预处理,模型训练,模型评价与选择,特征重要性计算.常见的数据预处理的方法有对数化、标准化、归一化、过采样和欠采样等.本文选用的数据集在构造的时候已经考虑过避免常见的数据特征异常问题,正负样本均匀且对称不需要再做预处理工作.接下

来进入模型训练过程,为评价模型在全新的数据上的性能.首先将存在 222 条观测的数据集划分为训练集和测试集.其次将数据集划分为包含 157 和 65 条观测的 2 个数据集,并将后者作为测试集.对数据集中的各个指标进行了描述统计和频数分析.其中描述统计主要用于观察得分效率密度中有没有出现异常值,即某项战术得分率不高但是使用率过高,或得分率很高但是使用率很低.频数分析则分析该样本中得分效率密度与全局得分能力在 111 场高水平乒乓球赛事的胜负双方中处于什么样的水平(优秀、良好、中等、较差),本研究机器学习模型主要流程图如图 1.

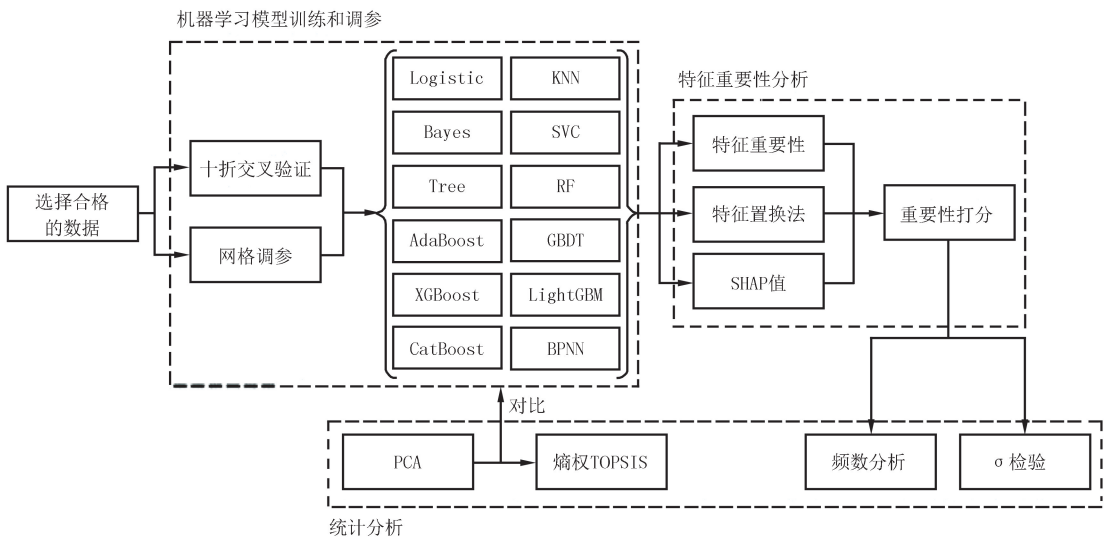


图1 机器学习模型主要流程图

Fig.1 The machine learning model building flowchart

2.2 模型的构建结果与分析

本研究的数据集选用监督学习的机器学习算法.将常用的该类机器学习模型按照 7 : 3 的比例划分训练集和测试集,采用十折交叉验证,对混双比赛三段和八次指标的最终模型准确率如表 1 所示.

表 1 各机器学习模型的准确率

Tab. 1 The accuracy of each machine learning model

模型	准确率	精确率	召回率	F_1	模型	准确率	精确率	召回率	F_1
Logistic 回归	0.867	0.870	0.867	0.867	AdaBoost	0.970	0.985	0.970	0.970
KNN	0.978	0.979	0.978	0.978	GBDT	0.985	0.972	0.985	0.985
朴素贝叶斯	0.911	0.914	0.911	0.911	XGBoost	0.985	0.985	0.985	0.985
SVC	0.978	0.979	0.978	0.978	LightGBM	0.985	0.985	0.985	0.985
决策树	0.956	0.956	0.956	0.956	CatBoost	0.955	0.955	0.955	0.955
随机森林	0.896	0.899	0.896	0.896	BP 神经网络	0.970	0.972	0.970	0.970

由数据可看出集成树模型整体准确性较高,集成树模型包含随机森林、XGBoost、LightGBM 等.接下来就运用集成树模型的 SHAP 值结果进行特征重要性分析,从 SHAP 分析的结果看,每行代表 1 个特征,横坐标为 SHAP 值.特征的排序是按照 SHAP 的平均绝对值,即重要特征排在前面.宽的地方表示有大量的样本聚集.1 个点代表 1 个样本,颜色越红说明特征本身数值越大,颜色越蓝说明特征本身数值越小.结果展示了各个特征和预测结果之间的关系(更重要的特征更能将红蓝色区分开).例如:男接女得分能力的值越大,其对预测为“胜”的正向贡献就越大.而男子接抢段得分效率的值越大,其对预测为“负”的正向贡献就越大,如图 2 所示.

根据图 2,男子接抢段得分效率最高,且远超其他特征.对比右侧的 SHAP 值区分,可以看出数值较大(红色)和数值较小(蓝色)的点区分得十分明显.这说明拟合的随机森林模型认为男子接抢段得分效率是对

区分胜负最为重要的特征,即为对胜负影响最大的能力.与之相比,男子接抢段得分能力、女子相持段得分能力和女子相持段得分效率的重要性较高,对于红色和蓝色的点的区分度较为明显.对于男接女得分能力和女子发抢段得分能力等特征,红色和蓝色的点越加重合,对胜负的影响逐步降低.

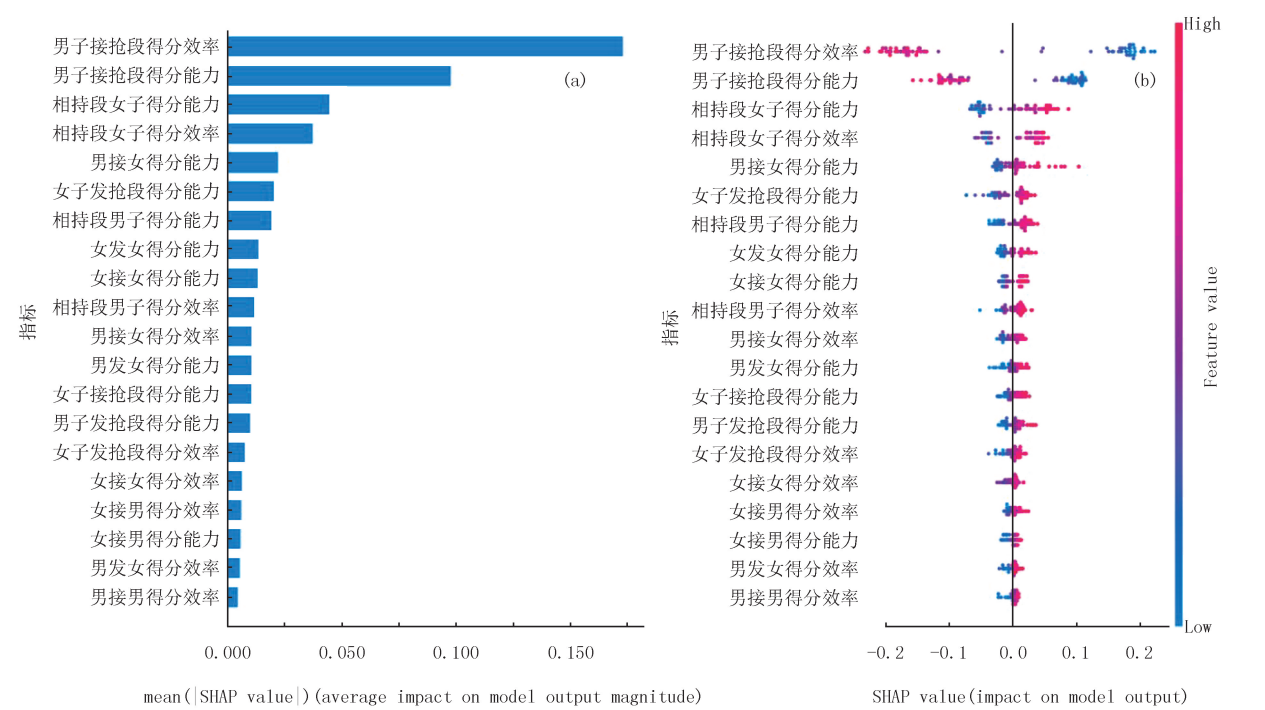


图2 SHAP值特征重要性排序和展示区分情况(随机森林)

Fig.2 Ranking the importance of SHAP-valued features and demonstrating differentiation(Random forest)

随后应用同样的方法计算出 LightGBM 和 XGBoost 的 SHAP 值,并将每个特征在随机森林、LightGBM 和 XGBoost 模型中对应的特征重要性进行排序,给每个模型中特征重要性前 10 的特征记 1 分后,统计每个特征的总得分,重要性得分为 2 分及以上的指标如表 2 所示.

表 2 特征的重要性得分

Tab. 2 The importance scores for features

特征	随机森林	LightGBM	XGBoost	重要性得分	特征	随机森林	LightGBM	XGBoost	重要性得分
女子发抢段得分能力	1	1	1	3	女子相持段得分能力	1	1	1	3
男子接抢段得分效率	1	1	1	3	男接女得分能力	1	1	1	3
男子相持段得分能力	1	1	1	3	男子相持段得分效率	1	1	0	2
女子相持段得分效率	1	1	1	3	女接女得分能力	1	1	0	2

由表 2 可以看出女子发抢段得分能力等 6 个特征对 3 个模型的预测都起到了重要的作用,此外男子相持段得分效率和女接女得分能力也具有较重要的作用,因此本研究主要应用这 8 个指标对比赛的应用进行技战术分析.

此外,在通过多个机器学习模型特征重要性分析后,本文还对三段和八轮次技术指标进行了频数分析.通过对应的频数确定了具体指标的评价标准(附录表 S2),将待分析的比赛中的胜方和负方的每项指标对应计算,然后和 4 分位数进行比较.首先,对某个队伍的某项指标为优秀,则说明这场比赛该指标反映的运动员技战术表现优异;但若是为较差,需要教练员和运动员格外关注;若指标得分接近中等,则技战术表现属于中规中矩.其次,应用该指标可以对比胜负双方运动员的技战术使用情况,可以通过指标的比较找出胜方赢在哪个环节,负方输在哪个环节,这也是这个评价指标的应用重点.同时,考虑到得分效率密度指标并不是数值越大就一定越好,数值过大和过小都是战术安排的异常,故通过描述性统计(附录表 S3)找到每个指标的均值和方差并且认为偏离均值 1 个方差的得分效率密度指标存在战术安排问题.故将 3 个重要的得分效率进行

指标区间计算(附录表 S4).最后将得分效率指标在三段和八轮次进行描述性统计,如一场比赛中得分效率在 1 倍标准差的范围内,即数据处在 $[\mu-\sigma,\mu+\sigma]$ 区间之中,则说明得分效率处在正常区间,用 0 表示.反之,则说明得分效率异常,用 1 表示,说明该指标的得分率和使用率使用有异常情况,需格外关注.

3 乒乓球混双技战术评价模型的应用研究

应用构建的机器学习模型对乒乓球混双比赛进行分析与评价.本研究所选的比赛双方运动员都是我国高水平乒乓球运动员,其中一组混双组合是参加巴黎奥运会的混双组合王楚钦/孙颖莎,因此有着较好的研究价值.本研究选取了王楚钦/孙颖莎组合与同一对手林高远/王曼昱在不同年份的全国比赛,分别是 2018 年全国锦标赛混双决赛和 2022 年全国锦标赛混双决赛,2 场比赛双方各赢 1 场,其中 2018 年全国锦标赛打满了 7 局,2022 年全国锦标赛也打了 6 局,所以无论是从研究对象的选择,还是比赛的激烈程度来看,以这 2 场比赛作为模型验证的案例非常适宜.

2018 年全国锦标赛混双决赛,王楚钦/孙颖莎以 4 : 3 战胜了对手,具体结果见附录表 S5.表 S5 的数据给出了 28 个指标的得分能力和得分效率的数值.首先依据 4 分位数进行初步的观测.胜方整体占优,其中包含女接男得分能力、女发女得分效率等 18 个指标值高于负方,10 个指标低于负方.在评价标准中,王楚钦/孙颖莎组合有女发女得分效率等 5 个指标达到了优秀级别,只有 1 个女子发抢段得分效率数值为较差;负方组合王曼昱/林高远组合所有指标中,数据标准达到优秀只有女子接抢段得分能力这个指标,但有女接男得分效率等 8 个指标是较差.所以从评价标准层面也可以得出双方的胜负关系.

如前所述,用机器学习方法构建了乒乓球混双项目技战术评价模型共 8 个指标.本研究就用这 8 个指标来分析双方比赛.首先是较为重要的男子接抢段得分效率指标,王楚钦/孙颖莎组合该项指标值是 0.242,而负方林高远/王曼昱组合该项指标只有 0.065,远低于胜方的值;胜方女发抢段得分能力为 0.012,而负方该项指标的值是-0.004;胜方男相持段得分能力为-0.004,评价为良好,而负方该项指标值是-0.044,评价标准属于较差;女子相持段得分效率 0.651,处于绝对优秀,而负方这个指标的值为 0.092,属于较差标准,该项指标双方差距较大;胜方女相持段得分能力为-0.004,高于负方的-0.024;在男接女得分能力这个指标上,胜负双方评价都是良好,但该指标负方的表现高于胜方.以上 6 个重要性得分为三级的指标,胜方有 5 个指标高于负方,从这个层面也看出机器学习模型的准确性.在 2 个重要性得分为二级的指标中,男子相持段得分效率胜方的值为 0.456,负方的值为 0.117,胜方依旧高于负方;而负方女接女得分能力指标为-0.008,评价为良好,胜方该项指标值为-0.020,在该项指标上胜方低于负方.

其次,得分效率密度指标并非越大越好,因此对这场比赛 8 个重要指标中 3 个得分效率密度做了观测,其中胜负双方男子接抢段得分效率值的范围在-0.278 至 0.346 之间,女子相持段得分效率值在-0.911 至 1.237 之间,双方男子相持段得分效率值在-1 至 1.27 之间,所以 3 个得分效率指标的值都在适合的区间内,可以认为以上 3 个得分效率指标都可以被用作对比分析.

因此,在 2018 年全国锦标赛混双决赛中,王楚钦/孙颖莎获胜的原因主要在于男子相持段的得分能力和女子相持段的得分能力较强.此外,女子发抢段得分能力和男子接抢段得分效率较高也对获胜起到了一定的作用.但王楚钦/孙颖莎组合在男接女和女接女 2 个接发球轮次还需要进一步提高技战术应用水平.

接下来,分析 2 对选手在 2022 年全国锦标赛混双决赛的比赛.本场比赛王楚钦/孙颖莎以 2 : 4 输给了林高远/王楚钦组合,具体指标计算结果见附录表 S6.

首先也是从整体上看双方各指标的评价情况.在 28 个指标中胜方林高远/王曼昱有女子发抢段得分效率、女子接抢段得分效率等 14 个指标为优秀,3 个指标为较差;而负方王楚钦/孙颖莎有 6 个指标为优秀,7 个指标为较差.在胜负双方对比中,男子发抢段得分能力、男子接抢段得分能力和女子接抢段得分能力等共 21 个指标上,林高远/王曼昱的指标分值均高于王楚钦/孙颖莎组合.从以上数据来看,胜方在比赛中应占据了绝对的优势.

接下来继续应用机器学习提供的混双比赛的 8 个重要性指标来分析.在男子接抢段得分效率指标上看,负方为 0.267,低于胜方;女子相持段得分效率负方值为 0.135,胜方该指标数值为 0.415,负方低于胜方;胜方

在女子相持段得分能力指标为优秀,数值达到了 0.019,而负方在该项指标上只有一0.037,双方相差较大;胜负双方在男接女轮次的得分能力指标上的数值都较大,但胜方的值为 0.019,达到了优秀,负方组合在该指标的值则略逊一筹.在剩下 2 个三级重要的指标男子相持段得分能力和女子发抢段得分能力上,负方的指标值分别是一0.009 和 0.019,均高于胜方林高远/王曼昱组合在该项指标上的值.在 2 个二级重要的指标上,胜负双方跟 2018 年全国锦标赛的指标对比情况较为类似,王楚钦/孙颖莎组合在男子相持段得分效率指标上高于王曼昱/林高远组合,而在女接女得分能力指标上低于王曼昱/林高远组合.

同样,在对胜负双方各项指标进行评价及对比后,还需对得分效率指标进行进一步分析,以此来观测该指标是否有过大或过小的值出现.得分效率指标观测结果如表 3 所示.

表 3 2022 年全国锦标赛得分效率指标观测

Tab. 3 Observation of scoring efficiency indicators at the 2022 national championships

指标	平均值	标准差	$\mu-\sigma$	$\mu+\sigma$	林高远/王曼昱	异常	王楚钦/孙颖莎	异常
男发男得分效率	0.456	0.991	-0.534	1.446	0.649	0	0.756	0
男发女得分效率	0.539	0.875	-0.336	1.414	0.871	0	0.825	0
男接男得分效率	0.54	0.791	-0.25	1.330	0.756	0	0.649	0
男接女得分效率	0.452	0.772	-0.32	1.224	0.774	0	0.640	0
女发男得分效率	0.491	0.779	-0.288	1.270	0.515	0	0.473	0
女发女得分效率	0.525	0.648	-0.123	1.173	0.663	0	0.223	0
女接男得分效率	0.424	1.162	-0.738	1.586	0.348	0	0.172	0
女接女得分效率	0.373	0.884	-0.511	1.257	0.700	0	0.320	0
男子发抢段得分效率	0.594	0.189	0.405	0.783	0.756	0	0.646	0
女子发抢段得分效率	0.593	0.739	-0.146	1.332	1.075	0	1.172	0
男子接抢段得分效率	0.034	0.312	-0.278	0.346	0.774	1	0.267	0
女子接抢段得分效率	0.458	0.500	-0.042	0.958	0.621	0	0.820	0
男子相持段得分效率	0.135	1.135	-1.000	1.270	0.046	0	0.147	0
女子相持段得分效率	0.163	1.074	-0.911	1.237	0.415	0	0.135	0

通过 1 倍 σ 范围对得分效率指标进行观测的结果如附录表 S7 所示,仅有林高远/王曼昱组合在男子接抢段得分效率存在异常,得分远大于 $\mu+\sigma$,这说明男子接抢段得分效率的高分可能并不能对胜利造成积极的影响,反而可能存在使用率指标的不足.该项指标越大说明得分率虽然大,但使用率较小,使得整体效果较差.经具体分析得知林高远在这场比赛中接抢段的得分率为 66.67%,而使用率只有 11.21%,所以说明林高远没有把握好他在接抢段中的优势.

综上所述,在 2022 年全国锦标赛混双决赛中,从技战术评价指标整体来看胜方林高远/王曼昱组合无论在优秀指标的数量上还是双方指标对比上,都远超王楚钦/孙颖莎组合.但最后的比赛是 4 : 2 结束,比赛过程较为激烈.主要原因就是在机器学习构建出混双比赛 8 个重要性指标上,负方组合在其中 3 项的指标值都高于胜方组合,分别是三级重要指标男子相持段得分能力、女子发抢段得分能力以及二级重要指标男子相持段得分效率.此外,观测这场比赛的得分效率指标,胜方林高远/王曼昱男子接抢段得分效率虽然远大于负方王楚钦/孙颖莎组合,但是同样远超正常的 1 倍 σ 区间,且男子接抢段得分效率是极为重要的三级重要性特征,胜方在该指标上技战术使用仍有不足之处.因此,虽然在整体指标上看胜方远高于负方,但对 8 个重要指标进行比较后得知负方并没有与胜负相差太多,这也是这场比赛过程较为激烈的原因.

以王楚钦/孙颖莎组合作为技战术分析对象,该组合在 2022 年全国锦标赛输球的重要原因在于女子相持段的得分能力不够突出,且男接女得分能力不如胜方.对比 2018 年全国锦标赛混双决赛,2022 年比赛中王楚钦/孙颖莎组合中女子发抢段得分能力和男子相持段得分能力指标无较大变化,都高于林高远/王曼昱组合,而且 2022 年的比赛中男子接抢段得分效率的数值也与 2018 年比赛的数值相差不大.但较为不同的一点

是女子相持段,2022 年孙颖莎在相持段的得分效率指标和得分能力指标都比 2018 年的指标下降了不少。2018 年比赛中女子相持段得分效率指标和得分能力指标分别为优秀和良好,但 2022 年决赛中这 2 项指标的评价都变为了中等。王楚钦/孙颖莎组合男接女得分能力指标值与 2018 年相差不大,都评价为良好,但林高远/王曼昱组合在该项指标的值从 2018 年比赛的较差标准到 2022 年比赛中变成了优秀。因此在该指标上王楚钦/孙颖莎组合也并不占优。女接女得分能力与男子相持段得分效率 2 个二级指标上 2 场比赛变化不大。从以上分析中可以得出,王楚钦/孙颖莎组合在 2022 年全国锦标赛决赛中输球的主要原因是在女子相持段环节得分能力与得分效率的下降。

就这 2 场比赛而言,王楚钦/孙颖莎组合在比赛中优势环节是女子发抢段和男子相持段的得分能力,该组合在这 2 个指标上数值都较为稳定。但是孙颖莎在相持段中的表现较差,因此应在训练中应多注意强化王楚钦和孙颖莎的多拍相持中的能力,尤其是孙颖莎在多拍回合中制胜拍的终结能力。此外,王楚钦/孙颖莎组合在男接女轮次和女接女轮次的得分能力上有待提高,尤其是女接女轮次的得分能力,这就需要提高孙颖莎接发球的能力,还要加强王楚钦第 4 拍的攻防转换能力和 2 人相持段的衔接配合,这些都应是教练员在以后的训练中应关注的重点环节。

4 研究局限与展望

本研究机器学习模型的构建也有如下不足之处:1)数据量不足。2)本文使用的机器学习模型较为传统,受限于数据集较少的影响,很多更新的模型无法被应用。3)单纯基于得分率和使用率的指标并不能直观地反映运动员的技战术发挥情况。后续的研究可通过卡尔曼滤波等计算机视觉算法对球速、转速以及运动员移动速度等进行测量,并将这些指标加入机器学习模型。

附录见电子版(DOI:10.16366/j.cnki.1000-2367.2024.07.10.0002)。

参 考 文 献

[1] 乔云萍.乒乓球比赛技战术的视频快速反馈与诊断方法的研究[J].北京体育大学学报,2008,31(10):1351-1356.
QIAO Y P.On the video rapid feedback and diagnose method of techniques and tactics in table tennis[J].Journal of Beijing Sport University,2008,31(10):1351-1356.

[2] 张晓蓬.中国乒乓球队战术训练水平定量诊断方法及实践效用[D].北京:北京体育大学,2004.

[3] 李今亮,苏丕仁.对部分世界优秀男子乒乓球进攻型选手技术实力的评估-兼谈十项指标评估法的建立[J].北京体育大学学报,1998(4):71-76.

[4] 赵喜迎,徐君伟.基于熵值法和灰色关联分析法的乒乓球技战术综合评价研究[J].山东体育学院学报,2020,36(5):96-101.
ZHAO X Y,XU J W.Synthetic evaluation for table tennis technique and tactic based on the entropy method and grey correlation[J].Journal of Shandong Sport University,2020,36(5):96-101.

[5] 张晓栋,肖丹丹,周星栋,等.乒乓球技战术动态三段指标统计法的构建与应用[J].中国体育科技,2018,54(1):80-83.
ZHANG X D,XIAO D D,ZHOU X D,et al.The construction and application of dynamic three-phase method on table tennis technique and tactics[J].China Sport Science and Technology,2018,54(1):80-83.

[6] 杨青,张辉.BP 神经网络和多元回归方法在乒乓球技战术能力分析中的应用[J].成都体育学院学报,2016,42(1):78-82.
YANG Q,ZHANG H.Application of BP neural network and multiple regression in table tennis technical and tactical ability analysis[J].Journal of Chengdu Sport University,2016,42(1):78-82.

[7] 曹烨程,陈辉,郭晨.基于图论模型的乒乓球混双技战术分析与应用研究[J].广州体育学院学报,2022,42(3):110-117.
CAO Y C,CHEN H,GUO C.Research on the analysis and application of table tennis mixed doubles techniques and tactics based on graph theory model[J].Journal of Guangzhou Sport University,2022,42(3):110-117.

[8] 蒋津君,姚家新.乒乓球单打比赛技战术实力评估体系及其诊断方法的重构与应用[J].天津体育学院学报,2015,30(5):432-437.
JIANG J J,YAO J X.Evaluation on tactical skills in table tennis single match and reconstruction and applications of its diagnostic methods [J].Journal of Tianjin University of Sport,2015,30(5):432-437.

[9] 杨青,张辉.乒乓球比赛技战术“四段指标评估法”的构建与应用[J].天津体育学院学报,2014,29(5):439-442.
YANG Q,ZHANG H.Construction and application of "four phase evaluation theory" technique and tactics for table tennis[J].Journal of Tianjin University of Sport,2014,29(5):439-442.

[10] 张辉,刘伟,户进菊.对抗性项目技术效益研究[J].体育科学,2015,35(9):44-49.
ZHANG H,LIU W,HU J J.Study on technique effectiveness of antagonistic sports event[J].China Sport Science,2015,35(9):44-49.

[11] ZHANG H,LIU W,HU J J,et al.Evaluation of elite table tennis players' technique effectiveness[J].Journal of Sports Sciences,2013,31(14):1526-1534.

Construction and application of table tennis mixed dual skill tactical
evaluation model based on machine learning

Cao Yecheng^{1,2}, Zhang Qianyi³, Chen Hui⁴, Liu Min¹

(1. College of Physical Education, Shanxi University, Taiyuan 030006, China; 2. Department of Physical Education, Shijiazhuang University, Shijiazhuang 050035, China; 3. Sports Science Department, Wenzhou Medical University, Wenzhou 325035, China;
4. School of Sports and Health Management, Chongqing University of Education, Chongqing 400065, China)

Abstract: The machine learning model was innovatively introduced into the technical and tactical analysis of the mixed doubles project, and the score efficiency density and global scoring ability constructed by the scoring rate and utilization rate of each section and round were taken as the labels. The Random forest model, LightGBM model and XGBoost model in the machine learning model were used to rank the feature importance of 28 technical and tactical indicators in table tennis mixed doubles competition. In this paper, two second-level evaluation indexes and eight three-level evaluation indexes of table tennis mixed doubles competition were extracted. Finally, the constructed evaluation model was applied to the technical and tactical analysis and evaluation of the two mixed doubles finals of Wang Chuqin/Sun Yingsha and Lin Gaoyuan/Wang Manyu, which verified that the table tennis technical and tactical evaluation model constructed by machine learning method has high practicability and accuracy.

Keywords: machine learning; table tennis; technique and tactic; evaluation model

[责任编辑 杨浦 刘洋]

表 S1 评价指标的确定

Tab. S1 The identification of evaluation indicators

八轮次指标	三段指标	八轮次指标	三段指标
男发男得分效率	男子发抢段得分效率	女发男得分效率	男子相持段得分效率
男发男得分能力	男子发抢段得分能力	女发男得分能力	男子相持段得分能力
男发女得分效率	女子发抢段得分效率	女发女得分效率	女子相持段得分效率
男发女得分能力	女子发抢段得分能力	女发女得分能力	女子相持段得分能力
男接男得分效率	男子接抢段得分效率	女接男得分效率	
男接男得分能力	男子接抢段得分能力	女接男得分能力	
男接女得分效率	女子接抢段得分效率	女接女得分效率	
男接女得分能力	女子接抢段得分能力	女接女得分能力	

表 S2 三段得分效率评价标准

Tab. S2 Scoring standards for the three-phase efficiency rating

名称	选项	数量	评价标准
男子发抢段得分效率	[0.708,1.398]	56	优秀
	[0.570,0.708)	55	良好
	[0.478,0.570)	55	中等
	[−0.036,0.478)	56	较差
女子发抢段得分效率	[0.787,1.554]	56	优秀
	[0.673,0.787)	55	良好
	[0.554,0.673)	55	中等
	[−5.556,0.554)	56	较差
男子接抢段得分效率	[0.306,0.804]	56	优秀
	[−0.079,0.306)	55	良好
	[−0.230,−0.079)	55	中等
	[−0.512,−0.230)	56	较差
女子接抢段得分效率	[0.634,1.054]	56	优秀
	[0.487,0.634)	55	良好
	[0.337,0.487)	55	中等
	[−6.164,0.337)	56	较差
男子相持段得分效率	[0.519,1.079]	56	优秀
	[0.306,0.519)	55	良好
	[0.121,0.306)	55	中等
	[−6.273,0.121)	56	较差
女子相持段得分效率	[0.522,1.681]	56	优秀
	[0.330,0.522)	55	良好
	[0.135,0.330)	55	中等
	[−6.079,0.135)	56	较差

表 S3 三段和八轮次得分效率描述性统计

Tab. S2 Descriptive statistical analysis of scoring efficiency for three segments and eight rounds

变量名	样本量	中位数	平均值	标准差	偏度	峰度	S-W 检验	K-S 检验
男子发抢段得分效率	222	0.570	0.594	0.189	0.674	2.382	0.964(0.000 * * *)	0.060(0.385)
女子发抢段得分效率	222	0.673	0.593	0.739	−7.649	61.441	0.283(0.000 * * *)	0.318(0.000 * * *)
男子接抢段得分效率	222	−0.079	0.034	0.312	0.331	−1.145	0.935(0.000 * * *)	0.156(0.000 * * *)
女子接抢段得分效率	222	0.487	0.458	0.500	−10.565	139.629	0.383(0.000 * * *)	0.220(0.000 * * *)
男子相持段得分效率	222	0.306	0.135	1.135	−4.969	24.563	0.366(0.000 * * *)	0.349(0.000 * * *)
女子相持段得分效率	222	0.330	0.163	1.074	−5.139	27.326	0.394(0.000 * * *)	0.290(0.000 * * *)
男发男得分效率	222	0.612	0.456	0.990	−6.075	36.934	0.291(0.000 * * *)	0.352(0.000 * * *)
男发女得分效率	222	0.648	0.539	0.875	−6.560	44.758	0.316(0.000 * * *)	0.334(0.000 * * *)
男接男得分效率	222	0.627	0.540	0.790	−7.637	61.345	0.288(0.000 * * *)	0.314(0.000 * * *)
男接女得分效率	222	0.528	0.452	0.772	−7.597	60.830	0.291(0.000 * * *)	0.313(0.000 * * *)
女发男得分效率	222	0.577	0.491	0.779	−7.620	61.036	0.288(0.000 * * *)	0.322(0.000 * * *)
女发女得分效率	222	0.591	0.525	0.648	−9.008	88.492	0.290(0.000 * * *)	0.295(0.000 * * *)
女接男得分效率	222	0.625	0.424	1.162	−5.074	25.216	0.333(0.000 * * *)	0.377(0.000 * * *)
女接女得分效率	222	0.488	0.373	0.884	−6.575	44.837	0.312(0.000 * * *)	0.325(0.000 * * *)

注：* * *、* *、* 分别代表 1%、5%、10% 的显著性水平。

表 S4 得分效率区间

Tab. S4 Score efficiency interval

指标	平均值	标准差	$\mu-\sigma$	$\mu+\sigma$
男发男得分效率	0.456	0.990	−0.534	1.446
男发女得分效率	0.539	0.875	−0.336	1.414
男接男得分效率	0.540	0.790	−0.250	1.330
男接女得分效率	0.452	0.772	−0.320	1.224
女发男得分效率	0.491	0.779	−0.288	1.270
女发女得分效率	0.525	0.648	−0.123	1.173
女接男得分效率	0.424	1.162	−0.738	1.586
女接女得分效率	0.373	0.884	−0.511	1.257
男子发抢段得分效率	0.594	0.189	0.405	0.783
女子发抢段得分效率	0.593	0.739	−0.146	1.332
男子接抢段得分效率	0.034	0.312	−0.278	0.346
女子接抢段得分效率	0.458	0.500	−0.042	0.958
男子相持段得分效率	0.135	1.135	−1.000	1.270
女子相持段得分效率	0.163	1.074	−0.911	1.237

表 S5 2018 全国锦标赛混双决赛双方指标评价结果

Tab. S5 The 2018 national championships mixed doubles final evaluation results for both metrics

比赛场次	王楚钦/孙颖莎	评价分类	林高远/王曼昱	评价分类
男子发抢段得分效率	0.497	中等	0.581	良好
男子发抢段得分能力	0.004	中等	0.038	良好
女子发抢段得分效率	0.535	较差	0.712	良好
女子发抢段得分能力	0.012	中等	-0.004	较差
男子接抢段得分效率	0.242	良好	0.065	良好
男子接抢段得分能力	-0.008	优秀	-0.047	较差
女子接抢段得分效率	0.692	优秀	0.487	中等
女子接抢段得分能力	0.016	良好	0.024	优秀
男子相持段得分效率	0.456	良好	0.117	较差
男子相持段得分能力	-0.004	良好	-0.044	较差
女子相持段得分效率	0.651	优秀	0.092	较差
女子相持段得分能力	-0.004	良好	-0.024	中等
男发男得分效率	0.654	良好	0.447	较差
男发男得分能力	0.000	中等	-0.020	较差
男发女得分效率	0.578	中等	0.653	良好
男发女得分能力	0.016	良好	0.000	中等
男接男得分效率	0.749	良好	0.653	良好
男接男得分能力	0.020	优秀	0.000	中等
男接女得分效率	0.589	良好	0.762	优秀
男接女得分能力	0.007	良好	0.016	良好
女发男得分效率	0.507	中等	0.493	中等
女发男得分能力	-0.016	中等	-0.008	良好
女发女得分效率	0.711	优秀	0.680	良好
女发女得分能力	0.008	中等	0.020	良好
女接男得分效率	0.653	良好	0.401	较差
女接男得分能力	0.000	中等	-0.016	中等
女接女得分效率	0.418	中等	0.586	良好
女接女得分能力	-0.019	中等	-0.008	良好

表 S6 2022 全国锦标赛混双决赛双方指标评价结果

Tab.S6 The 2022 national championships mixed doubles final evaluation results for both metrics

指标	王楚钦/孙颖莎	评价分类	林高远/王曼昱	评价分类
男子发抢段得分效率	0.646	良好	0.756	优秀
男子发抢段得分能力	−0.005	较差	0.042	优秀
女子发抢段得分效率	1.172	优秀	1.075	优秀
女子发抢段得分能力	0.019	良好	0.009	中等
男子接抢段得分效率	0.267	良好	0.774	优秀
男子接抢段得分能力	−0.033	良好	0.019	优秀
女子接抢段得分效率	0.821	优秀	0.621	良好
女子接抢段得分能力	0.005	中等	0.019	良好
男子相持段得分效率	0.147	中等	0.046	较差
男子相持段得分能力	−0.001	良好	−0.047	较差
女子相持段得分效率	0.135	中等	0.415	良好
女子相持段得分能力	−0.038	中等	0.019	优秀
男发男得分效率	0.756	优秀	0.649	良好
男发男得分能力	0.023	优秀	0.000	良好
男发女得分效率	0.825	优秀	0.871	优秀
男发女得分能力	0.028	优秀	0.037	优秀
男接男得分效率	0.649	良好	0.756	良好
男接男得分能力	0.000	良好	0.023	优秀
男接女得分效率	0.640	良好	0.774	优秀
男接女得分能力	0.009	良好	0.019	优秀
女发男得分效率	0.473	中等	0.515	中等
女发男得分能力	−0.019	较差	−0.009	中等
女发女得分效率	0.223	较差	0.663	良好
女发女得分能力	−0.037	较差	0.028	优秀
女接男得分效率	0.172	较差	0.348	较差
女接男得分能力	−0.037	较差	−0.028	较差
女接女得分效率	0.320	较差	0.700	优秀
女接女得分能力	−0.028	中等	0.037	优秀

表 S7 2022 全国锦标赛混双决赛得分效率指标观测

Tab.S7 Observation of scoring efficiency indicators at the 2022 national championships mixed doubles final

指标	平均值	标准差	$\mu-\sigma$	$\mu+\sigma$	林高远/王曼昱	异常	王楚钦/孙颖莎	异常
男发男得分效率	0.456	0.991	-0.534	1.446	0.649	0	0.756	0
男发女得分效率	0.539	0.875	-0.336	1.414	0.871	0	0.825	0
男接男得分效率	0.540	0.791	-0.250	1.330	0.756	0	0.649	0
男接女得分效率	0.452	0.772	-0.320	1.224	0.774	0	0.640	0
女发男得分效率	0.491	0.779	-0.288	1.270	0.515	0	0.473	0
女发女得分效率	0.525	0.648	-0.123	1.173	0.663	0	0.223	0
女接男得分效率	0.424	1.162	-0.738	1.586	0.348	0	0.172	0
女接女得分效率	0.373	0.884	-0.511	1.257	0.700	0	0.320	0
男子发抢段得分效率	0.594	0.189	0.405	0.783	0.756	0	0.646	0
女子发抢段得分效率	0.593	0.739	-0.146	1.332	1.075	0	1.172	0
男子接抢段得分效率	0.034	0.312	-0.278	0.346	0.774	1	0.267	0
女子接抢段得分效率	0.458	0.500	-0.042	0.958	0.621	0	0.820	0
男子相持段得分效率	0.135	1.135	-1.000	1.270	0.046	0	0.147	0
女子相持段得分效率	0.163	1.074	-0.911	1.237	0.415	0	0.135	0