

智能化时代的个性化训练

——机器学习应用研究进展与数字化未来

胡海旭^{1,2}, 金成平³

(1.南京体育学院 运动训练学院,江苏 南京 210014; 2.江苏省运动与健康工程协同创新中心,江苏 南京 210014;
3.武汉理工大学 体育部,湖北 武汉 430070)

【摘要】:人工智能被视为升华精准个性化训练的新动能,尤其在整合分期训练模式引领下,亟待建立数字化为基础、网络化为载体的人工智能训练方法应用体系。数字化智能化训练方法在全球范围内方兴未艾,通过系统文献分析阐述了运动训练中机器学习算法应用研究进展及其特征,并提出运动训练中机器学习应用模式框架示例,以期搭建运动训练界与信息科学界的交流桥梁,助力国内人工智能技术生成的个性化训练研究。为更好把握新时代赋予的机遇与挑战,提出:①推进数字化转型,实现整合训练范式革新;②打造智能化体系,助力“人机协同”精准个性化训练升级;③紧跟数字化前沿,“智造”数字孪生训练黑科技;④推进智能化时代的个性化训练面临着3大问题与挑战。

【关键词】:个性化训练;整合分期;机器学习;数字教练;数字孪生

【中图分类号】: G808.1; TP18 **【文献标志码】:** A **【文章编号】:** 2096-5656(2021)04-0009-11

DOI: 10.15877/j.cnki.nsic.20210705.001

复杂性科学被誉为“21世纪的科学”^[1],同时,21世纪也是信息科学时代,信息的载体就是数据。2007年,图灵奖获得者吉姆·格雷(Jim Gray)指出大数据将成为人类触摸、理解和逼近现实复杂系统的有效途径,并认为在实验观测、理论推导和计算仿真等3种科学研究范式后,将迎来第四范式——“数据密集型的科学发现”,亦即大数据科学。因此有学者指出,“大数据科学被认为是一种新的复杂性科学,其具有特色的方法论特征也给复杂性科学解决其他问题提供了新思路”^[2]。基于数据密集型技术的大数据的“量变”引发了人类分析和思考模式的“质变”,大数据时代最大的转变是从探求因果关系转变成挖掘相关关系,即只需知道“是什么”,无需知道“为什么”。

国内于2016年开始关注数据密集型运动训练研究范式的重要意义^[3]。刘昊扬^[4-5]对大数据背景下,竞技体育、全民健身与运动社交中智能体育“黑科技”的国际现状与国内差距进行了系统阐述与展

望;袁守龙^[6]、闫琪^[7]等注意到运动训练的海量数据生成对数字化体能训练实践的迫切需求,为大数据引领的数字化智能化训练与健身模式“是什么”做了很好的诠释;而杨国庆“整合分期:当代运动训练模式变革的新思维”^[8]一文可谓数据密集型运动训练范式中国化的一个重要里程碑。大数据的核心是利用数据(数字)的价值,机器学习则是利用数据价值的关键技术,是实现人工智能的有效技术途径。为更好地理解、吸纳、推动人工智能技术(即当前的机器学习算法)生成的个性化训练新范式的研究与应用,本文拟就机器学习定义及其在运动训练中的应用典范进行归纳,提炼应用现状的要点与不足,构建一般性应用框架示例以推动其交流应用进程,展望人工智能时代的数字化训练未来与挑战。

收稿日期:2021-05-18

基金项目:国家社会科学基金一般项目(20BTY092)。

作者简介:胡海旭(1983—),男,湖北黄冈人,博士,副教授,硕士生导师,研究方向:运动训练理论与方法。

1 机器学习的定义、分类及算法特征

人工智能通常是指机器以人类的意愿执行任务的能力,而机器学习则专门表示一种用于处理数据并从中学习的计算机应用程序,可定义为:致力于研究如何通过计算的手段,利用经验来改善系统自身的性能^[9]。在计算机系统中,“经验”通常以“数据”形式存在。计算机通过训练数据得到“模型(model)”的算法,即学习算法(learning algorithm)。然后,把经验数据提供给它(重新给定一个个体数据),它就能基于这些数据产生模型。在面对新的情况时,模型会给我们提供相应的判断(根据新个体的数据进行个性化预测)。简言之,机器学习是一种能够从经验中自动获取知识,并基于知识模型对新的情况做出个性化预测的“学习算法”。

根据“训练数据”是否拥有标记(label)信息(如判断一系列生物动作能力训练是否能提高弹跳高度,弹跳高度即标记),机器学习又可分为两大类:监督学习与无监督学习。①监督学习的目标是预测未见样本的目标特征的值,此时学习获得的模型称为“预测器(predictor)”,如以50名跳高运动员的全运会比赛成绩作为数值(numerical)标记,或者是否晋级决赛作为类别(categorical)标记。采集他们赛前年度训练周期中的变量,通过机器学习算法建立一个成绩、名次标记与影响变量之间的“学习器”。如果标记变量是类别变量,如“是否晋级”,相应的学习器(learner)被称为“分类器”;如果标记是数值变量,如“比赛成绩”,相应的学习器被称为“回归模型”。②无监督学习不依赖于标记信息,其目标是发现数据的一些内在分布信息。如主成分分析、聚类分析。二者学习的目标都是为了把训练样本学习获得的知识推广到未见样本,进行个性化预测或判别。常用的学习算法有:决策树、bagging、随机森林、mboost、人工神经网络、支持向量机、K最近邻回归、朴素贝叶斯等^[10-11]。其中,人工神经网络算法经历了浅层学习和深度学习(即深度学习,机器学习的一个高效学习算法新方向)两次浪潮^[12-13]。

机器学习相比传统的数理统计的算法优势主要表现为:①传统统计是“模型驱动”,机器学习是“数据驱动”,及其与“知识驱动”相结合。以线性回归模型为例,要求所有变量都不相关等一系列模型假定,这很难做到(或至少无法验证);机器学习中诸多

方法可以从各个角度评价变量和观测值的关系,且不受数学假定的限制。②传统的统计需要减少自变量数目,以防止高维数据影响模型的简洁性和可计算性;机器学习方法不但不降维,而且希望更多相关变量纳入,以识别大型数据集中的复杂的、非线性模式,使实时动态地创建更准确地个性化预测模型成为可能。③对于特定数据,传统统计认为应选择被认为能很好解释数据的参数模型,不同模型之间是相互竞争的;机器学习方法可以把不同竞争模型组合起来,产生比单个模型更加精确的预测。④传统统计对模型的评价是基于对数据及模型形式的数学假定,且只有一个训练集本身对模型的拟合来判断模型是否合适;机器学习则是采用交叉验证的方法,用训练数据集来训练模型,然后用未见样本数据作为测试集来评价模型的优劣。⑤以机器学习为代表的大数据思维与传统统计思维有显著不同:人们不再仅仅依靠随机样本的小数据,也关注全样本的大数据;不再只追求数据的精确性,同时也容忍其混杂性;不再只热衷于数据之间难以捉摸的“因果关系”,而是更多挖掘数据的“相关关系”,让数据自己“发声”,亦即只要知道“是什么”,而不需要知道“为什么”^[10,14-15]。

2 机器学习在运动训练中的研究进展

运动训练的目的在于根据运动员当前竞技能力状况,运用科学和技术,力图精心设计日常个性化训练策略提高竞技能力、减少运动伤病,以适时出现最佳运动表现。因而,运动训练实施流程可简化为4个阶段(图1)^[16]:①计划(包括竞赛日程和训练分期的具体目标等);②实现(包括训练器械设备、热身阶段生理心理评估、监测训练课负荷强度、团体项目中的战术训练、生物动作能力训练等);③控制(包括动作技能和运动解剖的视频分析、团体项目中技战术智能计算分析,给出定性定量的运动表现反

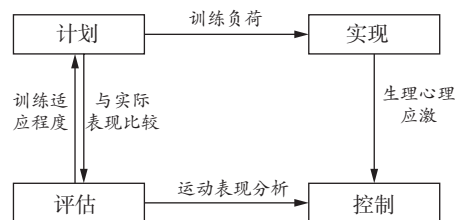


图1 运动训练实施流程^[17]

Fig.1 Sports training implementation process^[17]

馈);④评估(包括单次训练(短期)负荷的评估,一般以 Banisters 训练冲量(TRIMP)表示;总训练周期(长期)训练负荷的评估,一般以参赛目标成绩和计划的、实际完成的总训练量来测量)。

当前,以前沿科技引领运动训练,个性化高效提升运动表现的做法已成共识。事实上,传统运动训练活动往往需要专业的教练组人员开展通力合作、花费大量的人力物力财力,并因优质教练资源欠缺而采取“一刀切”的训练方法来满足需求,因训练量过度或不正确的运动模式而经受运动损伤困扰的现象甚为普遍,严重阻碍了竞技体育健康可持续发展路径。联合国教科文组织指出,人工智能技术能够促进个性化学习^[18]。鉴于此,亟须围绕“运动训练负荷”的设计与实施来打造适切不同个体的精准个性化训练模式的人工智能训练系统来提质增效:一方面,可大大减少教练员和运动员的琐碎性工作(训练)量,腾出更多时间从事创造性工作;另一方面,人“机”结合新模式有望整体提高运动训练效率、降低运动损伤风险。

近5年来,快速发展的可穿戴设备、各类传感器、物联网设备、智能计算算法等,为开展人工智能训练模式(数字教练E-Coach)提供了强有力的技术支撑,运动训练作为一个新兴且极具潜力的

应用领域,在借鉴其他领域^[19]应用经验基础上,取得了一些突破性成果^[20-24]。机器学习算法的应用正逐步成为该领域热点,研究进展归纳如下:以主题为(“sport”)AND(“training”OR“tracker”OR“logger”OR“diary”OR“trainer”)AND(“data mining”OR“computational intelligence”OR“artificial intelligence”OR“big data”OR“deep learning”OR“machine learning”),分别在“ACM Digital Library: dl.acm.org”“IEEE Xplore Digital Library: ieeeexplore.ieee.org”“Science Direct: www.sciencedirect.com”“Scopus: scopus.com”“Web of Science: web of knowledge.com”等5个数据库中进行搜索,遴选出应用于特定运动项目训练且可行性较好的文献26篇(表1),分析如下。

其一,主要解决两个维度的问题:①单一维度指仅仅关注运动训练中某一个方面的问题,数据来自反映同一问题类属的变量,但变量数量可能较多。当前,国际上对动作分析(学习)^[53]关注度较高,运动损伤评估^[54]也是一个重点研究问题。②多维度指同时关注运动训练中两个及以上的重要问题,变量的维度和数量都更高,也更接近运动训练的实践需求,通常被称为“数字教练”。

其二,主要项目分布在足球、跑步、举重和自行

表1 运动训练中的机器学习典型应用论文遴选

Tab.1 Selection of papers for typical applications of machine learning in sports training

问题	文献	年份	项目	样本	数据类别	算法	研究结果
单 维 因 素	【1】 ^[25]	2013	举重	6名年轻男性志愿者	匹配手套/臂章/腰带/哑铃传感器	RF	训练反馈可明显减少举重动作发生错误。
	【2】 ^[26]	2014	乒乓球	15名男大学生	照片打点分析	C4.5	对专家和新手区别成功率为81.2%。
	【3】 ^[27]	2015	高尔夫	275名男运动员	高速摄影的动作姿态	RF	识别高与低水平运动员的关键动作。
	【4】 ^[28]	2017	篮球	小样本	GPS/无线3D加速度计/陀螺仪/磁力针	SVM	自动识别训练动作准确率达99.5%。
	【5】 ^[29]	2018	篮球	3名业余爱好者	GPS/3D加速度计/陀螺仪/磁力针	k-NN/ RF	RF更好,动作识别准确率达87.5%。
	【6】 ^[30]	2018	射箭	50名青年男运动员	握力/纵跳等、射箭得分	SVM	对高低技能水平的区分度达到97.5%。
	【7】 ^[31]	2018	太极拳/深蹲	24名太极拳受试120个动作/50名深蹲受试90个动作	OptiTrack运动捕捉系统采集19个关节运动学特征	RF /SVM	基于运动专家知识数据,该虚拟现实训练系统可为受试者进行动作姿势诊断并实时提供口头与自动生成的视觉增强反馈。
训练 负荷 调控	【8】 ^[32]	2010	越野跑	个案	心率、环境温度、地形坡度	k-NN	基于地形、气温参数识别训练强度靶心率准确率达70%。
	【9】 ^[33]	2013	举重	15名初学者	测力传感器:力、位移、加速度等	ANN	自动识别举重器械负荷的合理与否。
	【10】 ^[34]	2016	足球	26名精英球员2080次训练	GPS/3D加速度计/陀螺仪/磁力针测度训练负荷	RF	赛季内训练周期负荷有两部分组成:开始时负荷高,临近比赛时负荷低。
	【11】 ^[35]	2017	美式足球	41名专业球员2711次训练	GPS采集训练距离/时间/高速跑占比等、3D加速度计采集运动负荷、RPE	ANN	根据GPS移动数据和运动负荷数据智能化预测RPE,以更个性化训练设计和评估。

续表 1

问题	文献	年份	项目	样本	数据类别	算法	研究结果
单维因素 疲劳损伤 营养预测	【12】 ^[36]	2019	曲棍球	弗吉尼亚大学曲棍球队	GPS/POLAR运动表:温度、湿度、对手评级百分比指数、心率变异、距离加速度节奏等训练负荷、进球数等	SVM/DT	近期训练负荷数据预测比赛结果准确率达79.8%,尤其赛前两天的中、高强度负荷对进攻表现更有益。拟提供训练方案和赛场决策。
	【13】 ^[37]	2020	自行车	小样本	TSS/TSB	LSTM	电子教练的智能生成训练方案与专职教练相近。
	【14】 ^[38]	2020	纵跳	21名青年田径运动员15周的训练数据	挺举/后蹲等复合训练次数与负重 的乘积、纵跳测试	ANN	有效模拟每周训练负荷量对弹跳高度的影响,以及每周影响的重要性程度。
	【15】 ^[39]	2014	铁人三项	未标明	主观感受/心率/天气/里程/卡路里	DT	根据感受、天气、心率和距离智能推荐进食方案。
	【16】 ^[40]	2017	手球和足球	132名职业球员	个性/心理/神经肌肉特征、动态姿势控制性、关节灵活性、核心稳定性等数据	DT	可以中等程度(ROC=0.747)预测肌肉损伤。
	【17】 ^[41]	2018	跑步	29名运动员	心率、GPS/3D加速度计/陀螺仪/磁力针、RPE	ANN/GBRT/LR	根据长达4个月的跟踪数据建模可准确预测疲劳水平。
	【18】 ^[42]	2018	足球	21名球员2年的训练数据	GPS/3D加速度计/陀螺仪/磁力针 测量训练负荷、心率、损伤事件	k-NN/SVM	提出了球员未来可能发生运动损伤风险的个性化概率图。
	【19】 ^[43]	2019	多个项目	23名女学生运动员2016赛季28次损伤	3D加速度计采集量化外部训练负荷、RPE\每天Hooper-Mackinnon量表 ^[44]	DBN	两天前的主观压力、一天前主观急性:慢性负荷比以及损伤事件当天的脑与神经兴奋性是对损伤表现最有效的监测指标。同时强调关注负荷随时间累积的动态过程。
	【20】 ^[45]	2020	足球	734名U10~U15球员的368次损伤事件	测试收集人体测量,运动协调和运动素质数据	XGBoost	基于季前赛数据,可预测运动损伤发生(准确率为85%),并且可以区别过度使用损伤与急性损伤(准确率为78%)。
	【21】 ^[46]	2013	自行车	2007—2012年世锦赛决赛选手	250 m冲刺、男30 km女20 km起跑赛、24名男女50圈淘汰赛等	BBN	预测不同赛事决赛的优胜者,为排兵布阵提供决策。
比赛决策	【22】 ^[47]	2020	足球	2个青少年足球队队员	位置、运动素质、射门助攻防守等15个足球技术	K-均值	建模遴选出了不同位置的足球技术特征,将在两个俱乐部进行验证。
	【23】 ^[48]	2015	排球	西班牙女排UCAM凌空穆尔西亚队	位置、心率等、运动学参数、温度、湿度等	MDPs/k-NN	专业排球球队实时训练反馈,提供训练间歇建议,避免损伤;分析技术的质量,如跳跃。
多维因素 数字教练 E-Coach	【24】 ^[49]	2017	中长跑	21~40岁跑步专业者2名,业余者1名	智能手机及其GPS/库珀12 min跑/Léger 20 m折返跑/营养问卷/每天摄入卡路里/BMI/身高体重性别/医疗史/跑步目标、社交媒体分享后的专家建议	NB	用于评估和系统调控运动营养和运动训练活动,并智能化生成运动营养(有效性86%)和运动训练方案(有效性82%)。总体达到专家水平的81%。
	【25】 ^[50]	2018	跑步	21名大学生	速度/步数/卡路里/饮食/800 m、4 000 m、5 000 m走和跑成绩/BMI	GBM	基于前6周营养和体能数据的建模,可用于准确预测第7周的800 m和5 000 m运动成绩。
	【26】 ^[51]	2019	举重	未标明	多种传感器采集运动学/心电/肌电数据/智能鞋采集生理信号/摄像	CNN	用于训练疲劳和动作姿势预警。

注: C4.5 Algorithm(C4.5算法); Naive Bayesian(NB,朴素贝叶斯); Artificial Neural Networks(ANN,人工神经网络); Random Forest(RF,随机森林); Bayesian Belief Networks(BBN,贝叶斯信念网络); Support Vector Machine(SVM,支持向量机); Long-Short Term Memory(LSTM,深度学习的长期记忆算法); Gradient Boosted Regression Trees(GBRT,梯度增强回归树算法); Gradient Boosting Machines(GBM,梯度提升机算法); K-nearest neighbor(k-NN,K最邻近算法); Decision Tree(DT,决策树); Markov Decision Processes(MDPs,马尔科夫决策过程); convolutional neural network(CNN,卷积神经网络); Dynamic Bayesian Network(DBN,动态贝叶斯网络); Extreme gradient boosting algorithms(XGBoost,极限梯度提升算法)。Training Stress Score®(TSS,训练负荷得分)\Training Stress Balance®(TSB,训练应激适应)^[52]; Rating of Perceived Exertion(RPE,主观努力程度); Receiver Operating Characteristic Curve score(ROC,受试者工作特征曲线得分)。

车等。且2018—2020年的文献相对更多,也预示着该领域的热度逐年递增。但是,第一作者和通讯作者有76.9%来自计算机和信息科学领域。作者国家分布前4名的是美国(4篇)、澳大利亚(3篇)、西班牙(3篇)、德国和比利时(2篇)。

其三,在样本量方面,当前的研究主要还是以小样本多变量或多维度数据为主要特点。样本样数较多的是文献【20】中有734名运动员测试数据,文献【10】【11】中分别有26名运动员2080次训练数据、41名运动员2711次训练数据,它们都是典型的以数据驱动的“机器学习”模型,而其中的小样本文献中主要是对理论建构基础上机器学习算法模型的个案验证。样本主体涵盖了专业运动员和业余运动爱好者。

其四,数据类别上,分为可穿戴、传感器、摄像机等自动采集数据和手动测试数据。其中,自动(10篇)、手动(7篇)、自动与手动混合(9篇)的数据分别占据本文文献的38.5%、26.9%、34.6%,可穿戴和传感器在运动训练中的普及计算技术应用最为广泛,尤其是GPS量化运动训练负荷特征技术,其次是运动手表、压力传感器、动作捕捉摄像技术等。此外,运动训练负荷量化需兼顾内部负荷和外部负荷,即主观客观相结合。在预测准确性上,专家经验知识驱动与数据驱动相结合的机器学习算法模型更高效实用。

其五,机器学习算法中最常用的是人工神经网络(6篇,包括深度学习2篇),随机森林(5篇),支持向量机(5篇),K最临近(4篇)。入选的运动训练论文除了【22】之外,都是采用监督学习。其“预测器”具有较好的预测准确性(70%~99%),尤其是用于动作分析的机器学习算法准确性普遍较高(81%~99.5%)。

目前,机器学习在运动训练中的应用还局限于对“小数据”的挖掘,但极大地弥补并解决了传统统计学中对多变量和多维数据的统计分析缺陷。以大数据为特性的“数据密集型”的运动训练机器学习模式尚在建构之中。因此,基于研究进展的结论,一方面,验证了机器学习算法在运动训练中应用的可行性和科学性,为数字化、个性化、智能化训练提供了支持路径和“机会窗”;另一方面,也明晰了大数据趋势下运动训练中机器学习算法的应用局限,现

阶段与算法同等重要的是建立结构性和非结构性的运动训练数据库,即数据银行。这也直接导致了距离整合多因素、多机制、多阶段数据的模型驱动、数据驱动与知识驱动相结合的智能化运动训练范式还有很长的一段距离,为更好地增进国内信息科学等领域与运动科学领域的学科交流合作,有必要探讨并构建运动训练中机器学习应用的模式框架。

3 运动训练中机器学习应用模式框架

信息革命和信息化对于人类社会发展的革命性影响是全方位的,“运动训练”更加不能“自傲地”置身事外。就代表文献的计量结果可知,开展智能化训练的多是计算科学和信息科学相关专业学者(共20篇,占比76.9%),远远多于运动科学专业的研究者(共6篇,占比23.1%),且计算科学相关专业学者们关注数字训练的热度与日俱增。然而,囿于计算机和信息科学背景,研究者们多将训练问题作为研究情景,其对技术改进的兴趣远比解决训练问题的兴趣要浓厚,某种程度上颠倒了“训练为体,计算为用”的实践逻辑。亟须运动训练科学从业者从专业理论视角建构运动训练实施模式框架,并与计算科学的技术实现模式框架相结合,相互融通促进。

模式框架是在一定训练理论(或人工智能技术)指导下建立起来的较为稳定的运动训练活动(或知识发现)结构框架和活动程序,其特征表现为:①作为理论与实践的中介,运动训练模式框架(或人工智能模式框架)是训练理论(或人工智能技术)走向运动训练实践的桥梁;②作为结构框架,反映了训练理论与方法(或人工智能实现)要素关系和整体结构;③作为活动程序,其针对性强、具体化、可操作。因而,该模式框架具有理论与实践的双重价值,既能为计算科学研究者提供运动训练科学理论支持和研究设计参考,又能为运动训练科学研究者提供计算科学的研究技术方法及路径借鉴。进而,促进跨学科合作,让不同学科背景的学者携手,将计算科学与运动训练科学中的理论、技术、方法、实践结合起来,基于计算科学解决运动训练中存在的瓶颈问题。

3.1 特定分期后运动训练负荷(应激源)与适应的整合训练模式框架示例

根据Hans selye应激理论及其发展,运动训练可视为一种典型的人体应激过程^[55-57],是训练应激源

(负荷刺激)引起运动员机体逐层级(系统、器官、组织、细胞、分子等)、多机制(物理、化学)适应的时空动态过程,与专项结合、与分期目标结合,以期实现良好的、专项的、分阶段的,生理、心理、物理等方面的适应。运动训练内容主要包括体能、技能、战术、心理等,它们与内在的人体基因、外在的环境(营养、温度、海拔等)互相作用影响训练负荷与适应的动态过程。就人体动作的外在本质是动作,内在本质是能量代谢的视角而言^[58],耐力、速度、力量等生物动作能力(体能训练的主体内容),都是能量供应下动作的不同重复方式的能力。体能训练的目的是提高动作能力(力量、速度、耐力、协调与灵敏等),其完成功能性动作的能量源于呼吸系统、心血管系统和肌肉系统协同运作,而这些系统能力的个性化、专项化

训练适应能力与基因信息密切相关。根据分期划分主要基于运动员自身状况和训练任务、目标安排,具有明显的阶段化、专项化特点,图2是在特定分期中训练应激源与人体适应过程的一个理论框架。

基因常常被视作运动选材的先天因素。近期,得益于精准医学取得的研究进展,遗传学、基因组学、代谢组学、免疫组学等技术在运动员训练个性化适应的实践中逐渐得以重视,如训练适应性(R等位基因与速度力量、X等位基因与耐力表现等)^[59],肥厚型心肌病(如MYH7基因)^[60],受伤风险(如COL5A1基因)^[61],营养个性化(CYP1A2、BCMO1、HFE、SLC30A8等基因)^[62-63]。对于独特的个体而言,携带的基因信息特点导致其训练、营养、健康等适应能力因人而异。基因信息不同,训练负

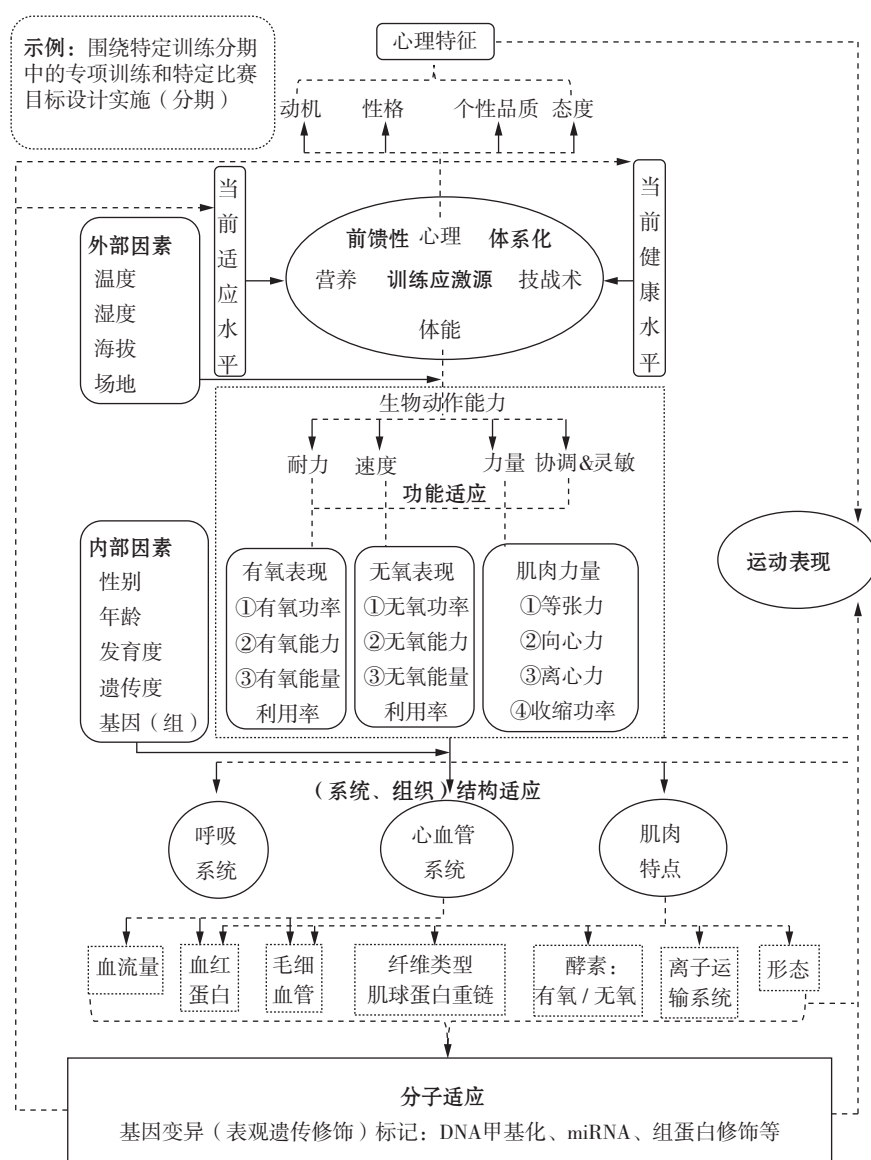


图2 特定分期后运动训练负荷(应激源)与适应模式的整合训练框架

Fig.2 The integrated training framework of sports training load (stressor) and adaptation mode after specific periodization

荷方式甚至相反,如ACTN3的R等位基因个体对大负荷抗阻训练的反应较大,应该优先进行高负荷阻力训练,着重于离心负荷以及高强度间歇训练。相反,携带X等位基因型的个体更适合进行低负荷、高运动量的阻力训练,使离心负荷最小化,以及进行更长、更低强度的有氧运动训练^[64-65]。因此,关注分子水平个性化是打造数字化训练适应的一个重要“尺度”。简言之,个性化训练是基因、训练与环境交融的过程,而基因变异性涉及训练、营养、心理等因素个性化调控的深层次机制。在精准医疗和运动训练中推广性较高的3个表观遗传调控机制和标记有:DNA甲基化,组蛋白修饰和microRNA(miRNA)^[66]。它们可作为识别运动适应、健康水平、疲劳等的新标志,且这种因运动训练等引起的表观遗传修饰效应可能是可遗传的、可塑造的和短暂的^[67],有明显的个体异质性,已被当作运动训练适应的潜在重要修

饰因子^[68-69]。因而,就分子角度而言,个人的运动表现阈值由基因组决定,运动训练是一个挖掘基因潜能的过程。基因以内在的、相对固定的遗传效应和可变异的表观遗传修饰共同影响运动训练效果。

3.2 数字化智能辅助训练系统实现的模式框架示例

运动训练是一个多成因、多阶段、多结局的全周期、全身系统性应激过程。智能化训练任务是自动从运动训练复杂数据中发现知识,根据Han J等^[70]界定的知识发现过程,即数据预处理、数据挖掘、模式评估、知识呈现(解释与可视化),可将基于机器学习的运动知识发现过程划分为3个流程:数据采集、数据挖掘(机器学习)、智能训练系统(模式评估、知识呈现)(图3)。本框架以机器学习算法实现数据挖掘,智能训练系统的任务即进行模式评估和发现训练中的知识,并有效反馈到运动训练计划中。

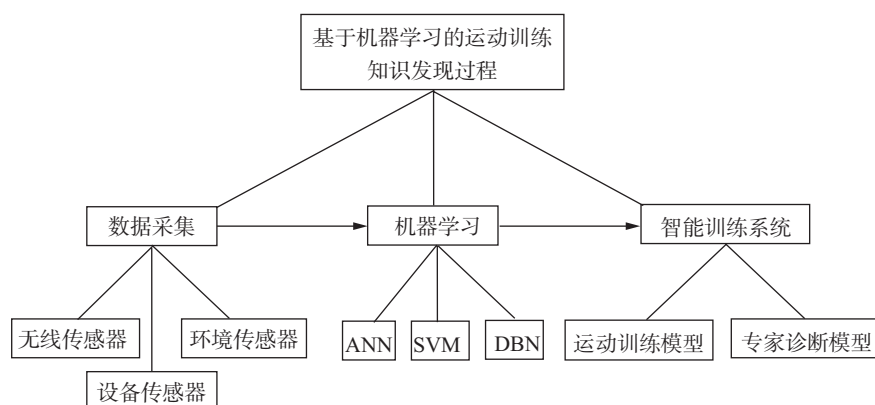


图3 普及计算的运动训练中知识发现过程

Fig.3 The process of knowledge discovery in sports training based on pervasive computing

结合图1中运动训练实施流程,借鉴电子工程与计算机科学专家Fister I将智能计算应用于运动训练研制出的理论框架,制定出数字化智能辅助训练系统实现的模式框架示例(图4)。该框架的基本程序源自知识发现的基本步骤和运动训练过程的4个流程:①计划,即划分具体的运动训练周期。分期计划根据当前运动员个体的训练或比赛目标,传统训练主要由教练员和运动员来设计。计划有战略(规划)和战术(方案)两部分组成。战略(规划)根据运动员内部因素(年龄、性别、发育状况、基因等)和专项化特征制订;战术(方案)则根据训练课程结果调整训练计划以实现目标,需兼顾训练课程环境(外部因素),包括温度、湿度、海拔高度、营养和场地条

件等。②实现。可穿戴移动设备的传感器和普及计算等,可实时采集位置、心理、生理等多维度、多层次、多尺度(图2)的训练负荷、反应、适应等数据信息。智能辅助系统的数据来源有两类:可穿戴移动设备数据、文件资料等其他非结构性数据(Excel格式、PDF格式、视频等)。③控制。是基于知识发现技术的数据管理和短期表现分析。首先,实现阶段获取数据,经过预处理保存至结构化数据库中;随后,采用计算智能(机器学习)算法从数据中丰富、学习知识,进行长期表现分析,智能化训练辅助系统有赖于新数据挖掘算法的发展。④评估。通过比较分析,将长期表现分析的结果与该训练周期内取得的结果进行比较,为下一步训练周期提供决策判据,评

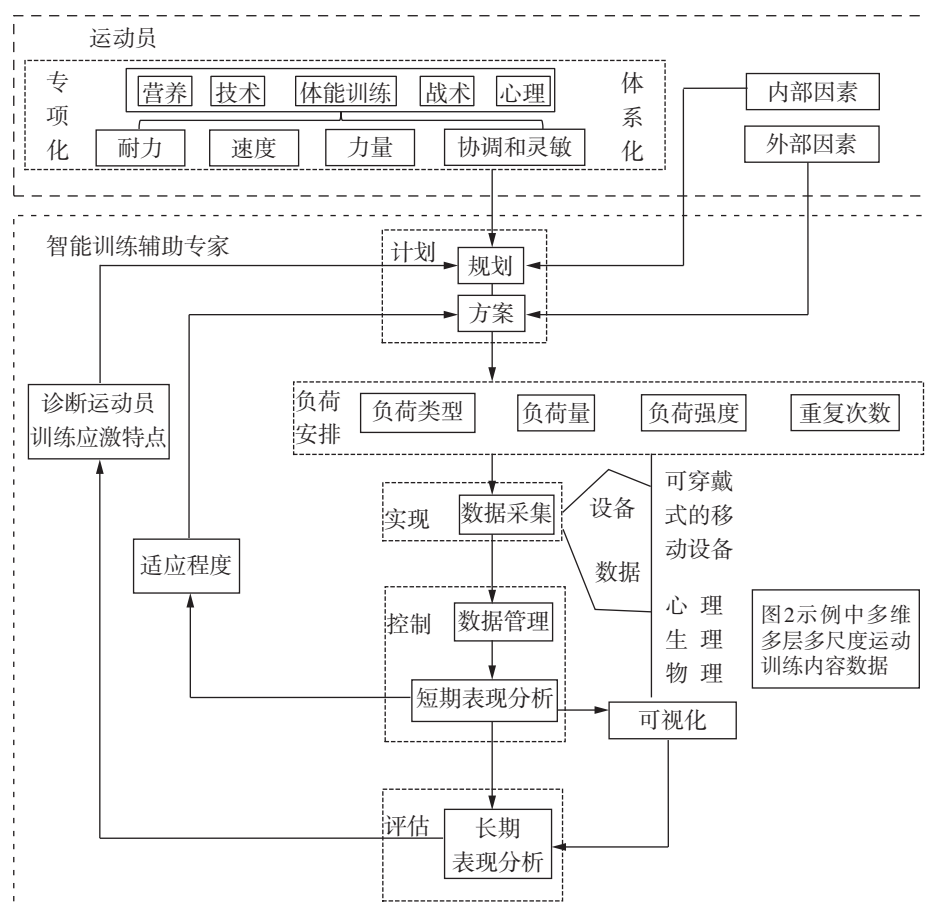


图4 数字化智能辅助运动训练系统实现的模式框架示例(依Fister, I.等改制,2019)

Fig.4 An example of the model framework implemented by the digital intelligent assisted sports training system (restructured according to Fister, I., etc., 2019)

估阶段的作用是确定运动员的训练适应水平与适应特点,为规划下一个周期的运动训练提供参考。

4 运动训练数字化未来及挑战

以数字化转型、网络化重构、智能化升级为特征的运动训练新研究范式、新训练方式正在全球范围内兴起。数字化是进军新一代人工智能体育的基石,其趋势是数据收集、聚合、分析与应用的运动训练全面数据化。尽管我国体育智能科技和体育大数据的供给侧上还存在诸多不足,但随着2022年北京冬奥会的临近,在竞技体育“训练—科技助力—保障”新模式的裂变赋能中,中国体育正在迈入智能化时代。为紧随国际前沿、坚持融合创新、推动智能化进程,提出如下数字化训练未来建议。

4.1 推进数字化转型,实现整合训练范式革新

人类社会、物理世界、信息空间构成了当今世界的三元^[71],也构成了运动训练活动的三元。运动员所处的社会环境、运动训练、人体这3大复杂系统融合并交互作用于特定的运动训练过程,致使运动训

练呈现为一个多因素(基因、生理、心理、物理、营养等)影响,多阶段(青少年过渡成年的启动期、成长期、夺冠期,训练分期的积累、转化、实现、过渡,参赛备战的调整、启动、兴奋、高峰等)特征,多结局(无效、超量恢复与适应、疲劳、过度训练、运动损伤等)效应的实践活动。

为有效控制运动训练过程,必须采取整合的训练理论与方法调控时空、过程与结果等因素。杨国庆于2020年提出的“整合分期”模式是破解该问题的一个引领性范例^[8]。随着数字化技术研究的深入和应用的推广,数字化愈发成为感知运动训练三元构成,实现运动训练信息关联与交互的整合分期训练模式走向实践的最佳技术路径。

打造运动训练学科数字化、训练实践数字化、科研方法数字化是当前运动训练从业者亟待形成的共识。借鉴2019年美国高等教育信息化协会(EDUCAUSE)的3个阶段论述^[72],也可将运动训练数字化转型逐步分为:信息数字化,指信息从模拟形式向数字形式过渡,如将纸质记录数字化;流程

数字化,指用数字技术和信息来改变单个机构或人员的运作流程,如工资发放、研究管理、训练计划、伤病情况、参赛安排等;业务数字化涉及文化、运动训练从业者和技术的深层次和相互协调一致的转变。从而促成新的训练模式,并改变训练团队的训练思维、战略方向和价值主张。

4.2 打造智能化体系,赋能“人机协同”精准个性化训练升级

运动训练即刻反应与长期适应、重大赛事的运动表现调控等问题在现有框架下都未能较好解决。认知缺乏与理论限制、建模与分析方法存在局限、计算与试验能力不足,确认和验证难以实现等问题愈发凸显,导致这一困境的原因是运动训练愈发呈现出高维度、动态性、非线性、多尺度等复杂系统的特征。超出了现有运动训练理论与实践的认知范围,亟待借助人工智能(如机器学习算法),搭建“人机协同”的智能训练系统。

人工智能利用智能方法和机器集合达到一种类人的能力,从而取代人类的部分事务性、重复性、程序性、机械性工作或者超出教练团队的专业知识范围等工作。协助教练团队从现有的繁重、繁琐的运动训练事务中解放出来,做到事半功倍,便于腾出更多的时间和精力来完成创造性、创新性训练工作。

此外,大数据背景下的人工智能技术得益于体量庞大的数据,相较于精度,更追求宏观层面对数据的洞察,发现经验难以察觉的相关关系;研究的起点是“基于数据”而非“基于假设”,强调让数据“开口说话”;在数据收集上倾向于获取“即时数据”。因此,基于瞬态生理、心理等多模态参数的个性化预测是数字化智能化训练的典型特征^[73]。随着更大规模和深度地整合训练、环境、人体系统及组学、训练历史、个体习惯偏好、整体健康水平、个体生活方式等高清晰度的数字化参数,以及指标参数量化技术的提高,未来将产生更有价值的预测模型;随着机器学习算法优化发展,加上输入到模型中的数据量的增加,必将驱动人机协同精准个性化训练升级。

4.3 紧跟数字化前沿,“智造”数字孪生(digital twin,DT)训练黑科技

数字孪生在人工智能基础上发展而来^[74],可界定为:利用数字技术对物体、系统、流程的信息进行实时映射,完成虚拟仿真的过程。目前,工业数字孪

生(industrial DTs)应用较多。

近期,数字孪生技术由工业领域逐渐向健康医学^[75-80]、营养学^[81]等人体研究拓展,即“人类数字孪生”(human DTs)。运动训练中数字孪生技术也受到信息科学、计算机科学等专家学者青睐。英国迈凯伦的Peter博士通过数字孪生成功改进一级方程式赛车比赛策略^[82];信息科学教授Barricelli等通过专业运动的SmartFit软件和手动记录的饮食、训练和睡眠等数据建立数字孪生体,并实时更新数据成功预测了训练期间的运动表现,以及根据运动表现情况,修改训练日常的生活行为方式^[83]。

在现有专业训练系统(如SmartFit)基础上,丰富和实时动态更新数据建立的数字孪生体将成为现实。因此,学者们也将“数字孪生教练系统”界定为电子教练或智能教练与数字孪生的组合,即数字教练系统通过数据技术为训练者量身定制个性化训练方案。其实现方式是通过整合分析心理、生理、物理等变量,并在物联网的协助下完成实体孪生与数字孪生之间的快速、无缝传输与计算。其核心技术是机器学习与人机交互的结合^[84]。未来,借鉴医学中的结合临床、组学功能和生活方式等多层次、多尺度数据,构建从器官到人体的数字孪生,做到整合人体深层表型(生物学、基因、组学、临床、电子健康记录等生物医学数据)、外部数字表型(连接医学设备,可穿戴设备与社交媒体反映的生活方式、心理、情绪、社会人口统计学等,数字脚印,数字生物学标记等现实世界数据)^[85]与运动训练变量(训练计划、训练方法、恢复与适应、比赛安排等)的历史与实时数据,“智造”数字孪生训练系统,将会是人工智能时代的大势所趋。

4.4 推进智能化时代的个性化训练面临的问题与挑战

智能化训练在全球范围内方兴未艾,但目前还没有一条完全成功的模式可以遵循。率先解锁并利用者将成为明天的冠军。运动训练中的数字化、网络化、智能化的软硬件升级是其基本框架前提,实现数字化转型、网络化重构,才能逐步实现智能化升级。因此,当前面临的主要问题与挑战是在达成数字化训练共识的前提下,保障运动训练中“线上有数据”,团队中“线下有人才”。具体而言:一是,团队数字化文化建设。引导运动训练从业者积极拥抱

数字文化,提升运动训练从业者的“数字技能”是关键。数字技能是适应智能化时代运动训练科学范式转型的必备素养,包括“对信息和数据的基本理解、沟通与合作、数字内容的创新与创造、数字安全与可持续发展、问题解决能力”^[86],直接体现在通过数据来改变传统的训练思路 and 模式,习惯用数据说话、用数据决策、用数据管理、用数据创新。二是,运动训练场所的物联网搭建。物联网设备划定了数字世界和物理世界之间的边界。亟待围绕“传统训练+数字技术”(如传统训练设备中添加传感器,场馆中布置高速摄像头等)和“数字技术+训练方法”(如虚拟现实VR、增强现实AR、混合现实MR等数字训练方法)进行分层逐步升级改造。进而,将“运动训练实况”连成网络,并捕获这些“运动训练实况”的运行及关联数据,实现控制与优化训练方法手段、训练过程。需要利用边缘计算技术使信息处理更加靠近数据源,连接并集成运动训练活动与智能分析技术,实现准实时反馈与改进功能。三是,团队中人工智能数据分析人才供给。运动训练物联网中包含了海量数据,需要人工智能算法将这些数据加以分解和分析,从而帮助教练员等运动训练从业者进行合理决策,因此,运动训练团队的数据分析人才配备尤为必要,坚持开放性原则,跨行业跨学科结成生态伙伴关系,让信息技术与训练方法手段领域专家相互学习彼此领域的知识、了解彼此的限制与困难,解决同一运动训练问题与挑战。

5 结语

运动训练影响因素涉及深层表型与外部数字表型等,还需要训练之前的识别诊断、之中的训练方法实施、之后的运动适应与恢复调理等过程调控,其复杂性不言而喻,整合分期训练思维及模式应用亦必不可少。智能化时代为解决复杂性问题提供了新观念、新技术、新方法,该方法具有典型的非消耗性、非竞争性、非排他性和“正外部”性(经济学术语,指生产和消费给他人带来收益而受益者不必为此支付的现象,本文特指未被训练所体现的额外训练收益)。机器学习作为人工智能的主要技术为赋能个性化、瞬时动态反馈等精准训练方法提供了强力支撑。基于数据驱动、模型驱动、知识驱动的智能训练系统已经吸引了国际上大批专家学者的验证追捧,机器学

习和人机交互相结合的数字孪生训练系统的研制也开始被青睐。它们都为运动训练从业者提供了详实的素材和宝贵经验。同时,当前运动训练大都是“外行”学者们在投石问路,以人工智能技术验证性应用为主流,绝大多数研究的样本数量、数据质量和训练理论与实践深度都存在客观缺陷,这为直接接触运动训练实践活动运动训练从业者们预留了巨大空间。就国内而言,完全可以充分发挥我国举国体制优势,借鉴英国生物银行、美国全民研究计划等成功模式建立国家体育行政系统主导的训练数据银行,并与计算科学、数据科学、行为科学、生物科学、神经科学等领域精诚合作,达成数字化训练与科研共赢的美好未来。

参考文献:

- [1] 黄欣荣. 复杂性科学的方法论研究[D]. 北京: 清华大学, 2005.
- [2] 齐磊磊, 贾玮晗. 从复杂性科学哲学看大数据的方法论价值[J]. 系统科学学报, 2020, 28(1): 24-28.
- [3] 胡海旭. 竞技能力增长理论模型及其演进[J]. 体育科学, 2016, 36(2): 14-24.
- [4] 刘昊扬. 基于人工智能的运动教练系统分析与展望[J]. 北京体育大学学报, 2018, 41(4): 55-60.
- [5] 刘昊扬, 崔一雄, 陶宽. 体育中的“黑科技”[N]. 光明日报, 2020-10-29(16).
- [6] 袁守龙. 体能训练发展趋势和数字化智能化转型[J]. 体育学研究, 2018, 1(2): 77-85.
- [7] 闫琪, 廖婷, 张雨佳. 数字化体能训练的理念、进展与实践[J]. 体育科学, 2018, 38(11): 3-16.
- [8] 杨国庆. 整合分期: 当代运动训练模式变革的新思维[J]. 体育科学, 2020, 40(4): 3-14.
- [9] 周志华. 机器学习: Machine learning[M]. 北京: 清华大学出版社, 2016.
- [10] 吴喜之. 应用回归及分类[M]. 北京: 中国人民大学出版社, 2016.
- [11] 周志华. 集成学习: 基础与算法[M]. 李楠, 译. 北京: 电子工业出版社, 2020.
- [12] 余凯, 贾磊, 陈雨强, 等. 深度学习的昨天、今天和明天[J]. 计算机研究与发展, 2013, 50(9): 1799-1804.
- [13] 孙志远, 鲁成祥, 史忠植, 等. 深度学习研究与进展[J]. 计算机科学, 2016, 43(2): 7-14.
- [14] 郭义, 王江, 陈波, 等. 论计算针灸学[J]. 世界中医药, 2020, 35(11): 5394-5398.
- [15] 杨强, 孟松鹤, 仲政, 等. 力学研究中“大数据”的启示、应用与挑战[J]. 力学进展, 2020, 50: 406-449.
- [16] FISTER I, FISTER J I, FISTER D. Computational intelligence in sports[M]. Berlin: Springer, 2019.
- [17] RAJP A, FISTER I F. A systematic literature review of

- intelligent data analysis methods for smart sport training[J]. *Applied Sciences*, 2020, 10(9): 3013.
- [18] 顾明远.《中国智能教育发展报告》序.刘邦奇,吴晓如.中国智能教育发展报告[M].北京:人民教育出版社,2019.
- [19] RAJŠP A, FISTER I. A systematic literature review of intelligent data analysis methods for smart sport training[J]. *Applied Sciences*, 2020, 10(9): 3013.
- [20] FISTER I., LJUBIČ K., SUGANTHAN P N, et al. Computational intelligence in sports: challenges and opportunities within a new research domain[J]. *Appl. Math. Comput.* 2015,262(31): 178-186.
- [21] RAUTER S. New approach for planning the mountain bike training with virtual coach[J]. *Trends Sport Sci.* 2018,3(11): 69-74.
- [22] ORESTI B, HERMIE H, CHRISTOPHER N, et al. Smart sensing technologies for personalised e-coaching[J]. *Sensors*, 2018, 18(6): 1751.
- [23] SILACCIA A, ABOU K O, MUGELLINI E, et al. Designing an e-coach to tailor training plans for road cyclists[C]//International Conference on Human Systems Engineering and Design: Future Trends and Applications. Springer, Cham, 2019: 671-677.
- [24] WOJCIK K, PIEKARCZYK M. Machine learning methodology in a system applying the adaptive strategy for teaching human motions[J]. *Sensors*, 2020, 20(1): 105-112.
- [25] VELLOSO E, BULLING A, GELLERSEN H, et al. Qualitative activity recognition of weight lifting exercises[C]//augmented human international conference, 2013: 116-123.
- [26] MAEDA T, FUJII M, HAYASHI I. Sport motion picture analysis as time series data[C]//2014 IEEE 13th International Workshop on Advanced Motion Control (AMC). IEEE, 2014: 530-535.
- [27] JOHANSSON U, KÖNIG R, BRATTBERG P, et al. Mining trackman golf data[C]//2015 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, 2015: 380-385.
- (注:由于版面限制,参考文献[28-86]略,如需查询,请与作者联系。)
- 作者贡献声明:**
- 胡海旭:选题撰写、修改定稿;金成平:选题讨论、修改完善。

Personalized Training in the Intelligent Era: The Application Research Progress of Machine Learning and the Digital Future of Sports Training

HU Haixu^{1,2}, JIN Chengping³

(1. School of Sports Training, Nanjing Sport Institute, Nanjing 210014, China; 2. Jiangsu Sports and Health Engineering Collaborative Innovation Center, Nanjing 210014, China; 3. Department of Physical Education, Wuhan Technical University, Wuhan 430070, China)

Abstract: Artificial intelligence is regarded as a new kinetic energy for refinement of targeted and personalized training. Under the guidance of Integrated Periodization Training mode in particular, it is urgent to establish an application system of artificial intelligence training methods based on digitalization and networking. However, digital and intelligent training methods are in the ascendant all over the world. Through systematic literature analysis, the research progress and characteristics of the application of machine learning algorithms in sports training are explained, and an example of the application model framework of machine learning in sports training is proposed to bridge the communication between the community of sports training and the information science, and help domestic artificial intelligence technology, and assist the research on personalized training generated by domestic artificial intelligence technology. Finally, in order to better grasp the opportunities and challenges given by the new era, it is proposed: ① Promote digital transformation and realize the innovation of integrated training paradigm; ② Create an intelligent system to help upgrade “human-machine collaboration” and targeted and personalized training; ③ Follow up with the digital frontier, take advantage of black technology for sports training that intelligently creates digital twins; ④ Realization of the personalized training in the intelligent era faces 3 major problems and challenges.

Key words: sports training; integrated staging; machine learning; digital coaching; digital twin