

学校代码： 10394

图书分类号： G842

学 号： QSZ20190815

密 级：



专业学位硕士学位论文

基于 BP 神经网络的网球比赛成绩预测模型研究

论 文 作 者： 黄志颖

指 导 教 师： 林向阳教授

学 科 专 业： 体育教学

研 究 方 向： 网球教学训练与理论

学 习 形 式： 全日制

培 养 单 位： 福建师范大学

提 交 时 间： 2022.5.18

中文摘要

成绩预测对职业体育至关重要的赛事转播、体育竞猜、新闻报道等都具有重要的价值和意义。在如今的体育环境中,预测对运动员取得良好成绩和提升训练效率都有显著的效果。在大数据背景下,运动的各项指标都被分解和细化,体育赛事结束后将有技术数据统计的结果,伴随着网球项目的发展和在世界范围内的日益普及,目前其技术指标的分解和统计已经相对成熟。本文针对网球运动的特点,引入 BP 神经网络方法并应用于网球赛事的成绩预测。运用了逻辑分析法、对比分析法、录像观察法、文献资料法等研究方法为本文的研究提供支持。以 2019-2021 的四大满贯比赛作为典型算例,建立网球项目的成绩预测模型,在 Matlab 中详细研究了针对网球项目的指标体系建立、BP 网络预测模型构建、样本库建立、网络训练评价和预测的全过程。主要研究内容包括:

1. 介绍 BP 神经网络算法的特点和计算方法,在结合网球项目技术指标特点的基础上分析了其应用于网球项目成绩预测的可行性。
2. 基于 BP 网络的网球赛事成绩预测模型建立过程中赛事数据的收集和处理,指标体系的分解、量化方法及样本集的建立,利用处理后的数据在不同的特征值和训练样本集情况下分别搭建预测模型并进行训练。
3. 在不同的特征值集群、输入样本量、场地类型下的 BP 神经网络预测模型的预测准确率进行分析和探讨,对网络预测有效性验证的同时深入分析了影响网络预测准确率的关键性参数,并对技术参数取值进行了优化。

通过研究表明,基于 BP 神经网络原理,运用 Matlab 软件进行编程,构建的网球四大满贯的成绩预测模型是可行的。计算结果也表明,收集的有效训练数据样本数需要达到一定程度才可以保证模型的预测精度。在特征值为 10、训练样本量为 7 个大满贯赛事数据的条件下,建立的 BP 神经网络预测模型的整体预测精度可达 78.125%。结合预测结果进一步分析发现,在大满贯赛事中,ACE 球、双误、主动得分和非受迫性失误这几个关键参数对网络预测结果的变化较为敏感。论文结果和建立的预测方法将为网球项目的预测和数据分析提供了新的方向。

关键词: 网球; 四大满贯; 技战术; 预测模型; BP 神经网络;

Abstract

Performance prediction is of great value and significance to professional sports, such as event broadcasting, sports guessing, news reporting and so on. In today's sports environment, prediction has a significant effect on athletes to achieve good results and improve training efficiency. In the context of big data, all sports indicators have been decomposed and refined. After the sports event, there will be the results of technical data statistics. With the development of tennis and its increasing popularity all over the world, the decomposition and statistics of its technical indicators have been relatively mature. According to the characteristics of tennis, this paper introduces BP neural network method and applies it to the performance prediction of tennis events. Taking the four Grand Slam competitions from 2019 to 2021 as a typical example, the performance prediction model of tennis project is established. The whole process of index system establishment, BP network prediction model construction, sample database establishment, network training evaluation and prediction is studied in detail in Matlab. The main research contents include:

1. This paper introduces the characteristics and calculation method of BP neural network algorithm, and analyzes the feasibility of its application in tennis performance prediction based on the characteristics of tennis technical indexes.
2. During the establishment of tennis match performance prediction model based on BP network, the collection and processing of match data, the decomposition of index system, quantitative methods and the establishment of sample set. Using the processed data, the prediction models are built and trained under different eigenvalues and training sample sets.
3. The prediction accuracy of BP neural network prediction model under different eigenvalue clusters, input sample size and site type is analyzed and discussed. While verifying the effectiveness of network prediction, the key parameters affecting the accuracy of network prediction are deeply analyzed, and the value of technical parameters is optimized.

The research shows that the prediction model of tennis four Grand Slams based on BP neural network principle and Matlab software is feasible. The calculation results also show that the number of effective training data samples collected needs to reach a certain degree to ensure the prediction accuracy of the model.. Under the condition of comparing the characteristic value of 10 and the training sample size of 7 Grand Slam events, the overall prediction accuracy of the BP neural network prediction model can reach 78.125%. Combined with the prediction results, it is found that the key parameters of ACE ball, double error, active score and unforced error are more sensitive to the changes of network prediction results in Grand Slam events.. The results of this paper and the prediction method established will provide a new direction for the prediction and data analysis of tennis events.

Key words: tennis; Four Grand Slams; Technical tactics; Prediction model; BP neural network;

目 录

绪论.....	1
1.1 选题背景和目的.....	1
1.1.1 选题背景.....	1
1.1.2 选题目的.....	2
1.2 研究对象与方法.....	2
1.2.1 研究对象.....	2
1.2.2 研究方法.....	2
1.3 文献综述.....	3
1.3.1 网球技战术指标相关研究.....	3
1.3.2 国内外关于体育预测的研究现状.....	5
1.4 相关概念及界定.....	9
1.4.1 BP 神经网络.....	9
1.4.2 体育成绩预测.....	9
1.4.3 网球四大满贯.....	10
1.4.4 预测模型.....	10
2. BP 神经网络模型构建的理论基础.....	11
2.1 体育预测方法研究.....	11
2.1.1 灰色预测法.....	11
2.1.2 德尔菲法.....	11
2.1.3 时间序列预测法.....	11
2.1.4 回归预测法.....	12
2.1.5 预测方法评价.....	12
2.2 BP 神经网络的基本理论.....	13
2.3 本章小结.....	16
3. 基于 BP 神经网络的网球赛事成绩预测模型构建.....	17
3.1 BP 神经网络的特征选取.....	17
3.1.1 影响网球比赛成绩的相关因素.....	17
3.1.2 BP 神经网络的数据来源.....	18
3.1.3 BP 神经网络的量化指标体系建立.....	19
3.2 BP 神经网络输入数据的预处理.....	22
3.3 BP 神经网络参数设定和模型的实现.....	23
3.3.1 BP 神经网络参数的设定.....	24
3.3.2 BP 神经网络的 Matlab 实现.....	27
3.4 BP 神经网络模型的训练及评价.....	28

3.4.1 BP 神经网络模型的训练.....	28
3.4.2 模型过拟合及交叉验证.....	29
3.5 本章小结.....	31
4.网球比赛成绩预测模型的实验结果与分析	32
4.1 选取不同轮次对模型的准确率影响对比分析	32
4.2 不同特征值对模型的准确率影响对比分析	34
4.3 不同的训练样本量对模型预测准确率的影响与分析	40
4.4 不同场地类型的模型准确率对比分析	43
4.5 取技术平均值后模型的准确率对比分析	44
4.6 2021 年四大满贯冠亚军预测结果对比	49
4.7 本章小结.....	52
5.结论与展望.....	53
5.1 结论.....	53
5.2 未来展望.....	53
附录.....	55
参考文献.....	64

绪论

1.1 选题背景和目的

1.1.1 选题背景

随着科技的日新月异以及体育行业的快速发展，数据和信息呈现爆炸式增长，互联网高科技技术风起云涌，大数据浪潮汹涌来袭。当代体育的发展、体育赛事成绩和训练水平的提升越来越依靠大数据，我们可以针对这些数据来对训练方法和训练内容进行适当的调整，以期在有限时间内最有效率的提高运动成绩。

科学的预测方法已经广泛应用于各个领域，而体育赛事在近几年也不断的引入各种各样的预测方法，如灰色预测法，统计回归分析法等^[1-2]。预测重要的体育赛事成绩越来越受到每个国家体育研究者的重视，利用科学的预测方法研究体育比赛能够帮助为运动员赢得比赛提供帮助。因此，通过预测方法和相对技术指标的关系，指导运动员的科学训练、提升体育成绩，是大数据时代下提高竞技水平的重要途径之一。

网球运动被称为世界第二大球类运动，观赏性强，比赛过程紧张刺激，受到了各国人民的喜爱^[3]。近年来，随着网球运动在我国的普及，参与到网球项目中的人也日渐增加，其赛事指标也在逐步细化。尤其是在大规模赛事中，我们将可获取各场次比赛双方运动员的各项技术水平，这些指标的数据量已经非常庞大且详细。基于 BP 神经网络方法，有效的赛事预测方法将可以得到更为精准的预测结果。

网球项目的预测即是利用以往的比赛技术数据和一些数学模型来实现的。过去网球项目的预测往往集中于统计分析和线性回归模型^[4-5]。通过对比发现，这些预测方法存在一定的局限性：（1）在收集数据时随意性高、模糊性大，靠自己的主观判断来选取指标，缺乏一定的科学依据。（2）预测精度不高、预测耗时长，泛化能力差。这些预测方法可以在特定条件下实现对网球比赛的预测，但是应用前景可能是有限的。针对网球运动的特点，量化指标是由多参数构成的，并且我们可以得到这些参数，我们想要知道参数之间的非线性关系以及每个参数的权重是如何分配的，因此，在这些预测方法中，神经网络可能是一个应用于网球项目当中的相对

较好的方法。

神经网络是一门多个学科领域的边缘交叉学科，BP 神经网络是应用程度最高的机器学习算法之一，同时也是人工神经网络的重要分支，在计算机、农学、医学等实际应用领域中取得了丰硕的成果，且运用 BP 神经网络通过收集数据并且建立数学模型进行预测已经非常成熟了。BP 神经网络可以存储海量的输入-输出的映射关系，更好的解决非线性问题和一些不确定性问题，对不确定的因素进行权重的分配。

基上分析，本研究试图通过搜集数据并挖掘网球比赛相关的技术指标，进而用这些数据改善训练水平和训练手段，将会得到比较有效的预测理论体系。引入神经网络技术来构建有效的预测模型，可以有效地减少人工计算、节省大量的时间，并且得到客观而又准确的预测结果。

1.1.2 选题目的

1. 寻找适合网球赛事成绩预测的有效方法，推动 BP 神经网络理论在网球项目的发展，扩大该理论在网球项目的延展性。
2. 将 BP 神经网络尝试应用于网球赛事的相关预测，建立适合于网球赛事的神经网络预测模型，并具有一定的精度。
3. 验证网球成绩预测模型的可行性和实效性，为网球项目的成绩预测提供一种新的方法和思路。
4. 通过预测的结果找到影响网球赛事成绩的关键性指标，有的放矢的对运动员能力进行高效地提升。

1.2 研究对象与方法

1.2.1 研究对象

本文的研究对象为基于 BP 神经网络的网球比赛成绩预测模型，以四大满贯赛事为调查对象，以 2019-2020 年四大满贯赛事作为训练集，2021 年四大满贯赛事作为测试集。

1.2.2 研究方法

1.2.2.1 文献资料法

通过以“体育”、“体育成绩预测”、“预测模型”和“BP 神经网络”等关键词在超星数据库、万方数据库和中国知网等获取国内外报刊正式发表的有关体育成绩预测方面和国内外各大型赛事成绩等获取有价值的文献资料，查阅有关 BP 神经网络的应用资料和其他预测方法相关资料，基本掌握了 BP 神经网络的基础知识和 Matlab 软件的应用，为后续的研究打下了坚实的基础。

1.2.2.2 数理统计法

对所需要的 2019-2021 年四大满贯数据进行搜集，保存在计算机中，并利用 EXCEL 软件对所搜集的四大满贯相关数据进行整理和排序，之后输入到 Matlab 软件中进行模型的训练和评价，并验证模型的时效性，为本研究提供可靠的数据支撑。

1.2.2.3 录像观察法

对 2019-2021 年大满贯赛事的录像进行收集整理，在录像中，比赛间歇和结束时所给出的官方统计数据也是重要数据，作为本文的参考数据。

1.2.2.4 实证分析法

以 2019-2021 年网球四大满贯公开赛男子单打选手为实验对象，运用网球竞赛成绩神经网络预测模型对一发得分率、主动得分和总得分率等数据进行建模预测并验证。

1.2.2.5 逻辑分析法

本文将从不同技术方面进行分析，主要是通过正反手、发球、高压、截击等四个技术方面进行理论分析，将统计分析和整理后的数据，按照写作的顺序调理清晰，符合逻辑思维的方法。

1.2.2.6 对比分析法

通过反复观看视频和慢速回放，对 2019-2021 年四大满贯比赛进行技术统计，把对阵选手的相关数据进行对比分析，找出不同选手技战术使用的差异，找出不同选手技术的特点和数值变化，以此来充实本文研究。

1.3 文献综述

1.3.1 网球技战术指标相关研究

任波^[6]通过李娜在 2013 年中的 11 场比赛作为研究对象，运用数理统计的方法找出李娜获胜以及失利的技术数据。发现李娜取胜的主要优势指标是 Aces 球、破

发成功率和制胜分；失利的劣势指标有一发成功率、非受迫性失误、一发得分率和破发成功率。

刑晓圆^[7]针对 2006 年我国 7 名女单选手的 9 场比赛，毛瑞斯莫 7 场比赛以及 32 强选手的 62 场比赛的各项技术指标进行归纳整理并统计处理得出：在国内女子网球选手中，中国网球女子单打选手在技术上与世界顶尖选手仍存在一定的差距，尤其是在一发得分率、二发得分率和非受迫性失误等方面。

刑晓圆^[8]对我国女单选手在 2006-2007 年度参加的大满贯比赛的技战术数据分析，再对李娜的技战术数据进行分析，将我国其他选手和李娜对比。结果表明：我国女子网球选手在几项指标中（一发得分率、制胜分、二发得分率、破发点把握率）都有所突破，并且非受迫性失误和双发失误率和有所下降。单打比赛底线技术非常重要，我国女子在底线技术需要多打磨，不断调整细节发现自己的不足，大胆进攻和调动对手为自己的得分创造机会。

罗溢杰^[9]选取 2011 年 10 场重要的赛事 8 名男子顶尖选手作为研究对象分析在硬地赛事“关键分”的处理情况得出，关键分领先越多胜率会越高，而其中三位优秀选手穆雷、纳达尔和德约科维奇在逆风情况下翻盘的概率很高；在比赛过程中前 3 拍球结束率之和为 51.4%，不拖泥带水，说明发球和接发球环节是很关键的。得分技术运用最多的顺序是正手、侧身攻、发球和网前，以后场居多，中场其次，前场最少。

李国立^[10]选取了 2013 年的 59 场职业网球比赛，运用统计分析方法分析制胜能力因素，建立多元回归方程对球员的实力进行评价。影响球员实力的几个重要因素分别为：发球得分率、挽救破发点成功率、ACE 球数和破发点把握率。

张潮^[11]以我国女子网球运动员参加的四大满贯比赛为研究对象，运用因子分析法得出影响我国女子网球单打技战术因素的 4 个共性因子为得失分因子、接发因子、发球因子和攻击因子。

郭立亚^[12]对 2008 年的 ATP 年度积分排名与技术指标进行统计分析并建立回归方程，发现与 ATP 年度积分排名的关联程度较高的技术指标有三个，分别为挽救破发点成功率、接发局胜率和二发得分率。

杨海兵^[13]采用统计方法对大满贯男单比赛的主要技术进行了分析，筛选出了三个对比赛胜负影响较大的指标，分别为接发得分率、一发赢球率和二发赢球率。利用三项指标建立了比赛胜负判别方程。

付饶^[14]对 2005-2006 年的澳大利亚网球公开赛的 239 场比赛进行分析,得到了 3 个对比赛胜负影响程度较高的 3 个指标,依次排序为:接发球得分率、一发得分率和一发成功率。利用这 3 项指标对运动员竞技水平进行评估,并制定出评估标准。谢相和^[15]对 2009-2011 年温布尔顿网球公开赛 250 场女子单打比赛的 20 项数据进行统计分析,影响胜负最大的 3 个指标是:一发得分率、二发得分率和接发球得分率。利用三项指标建立起的实力评估针对草地女子单打,具有较高的准确性。

高徐^[16]对比分析了 2010-2012 年的法国网球公开赛中 371 场女子单打比赛的技术指标,发现胜负双方中有 14 项指标存在显著差异;判别分析得出其中 3 项指标对比赛胜负影响程度较高,进一步对国内外女子网球运动员的实力评估后发现:国内球员普遍不适应红土地,发球和接发球的能力较弱。

陈国豪^[17]对 1476 场 WTA 选手比赛技术数据进行统计分析发现,在 2015 年整体指标显示得出女子比赛风格偏向于进攻,从整体分布情况得出,我国女子网球运动员已达到世界级水平。

综上所述,对网球技战术相关的文献进行整理和分析发现,由比赛的技战术分析、技术统计指标及各种类型的比赛场地特征等方面进行分析的研究已不少。对于一些微小的指标(1-3 拍、3 拍-5 拍、5-9 拍的数量)也进行了较深入的研究。总得分率与比赛的胜负具有最强的相关性,因此发球得分率显得尤为关键。总结来说,网球方面的专家们对技战术的研究越来越深入、越来越细化,呈现出越来越丰富的状态。同时上述文献关于网球技战术的研究虽然是从不同角度不同方面来进行,但是始终存在着一些不足,相关研究成果亦存在着一定的局限性:1. 从一个赛事或者一个运动员的几场比赛中提取的研究样本,数据来源不够详细,略显单一,样本容量小,只针对其中几场比赛进行研究,缺乏大局观,没有整体大数据的概念。2. 最终提炼出的关键指标只有 3-4 个,而影响网球比赛的数据指标较多,无法全面的反映出多个指标对网球比赛的影响。3. 对大量的数据进行了一些的统计和分析,但是缺少一种能找到所有数据的非线性关系并且得出结论的有效手段。因此,本文将探索 BP 神经网络在网球项目中对多样本数据和多个指标的非线性关系的处理。

1.3.2 国内外关于体育预测的研究现状

1.3.2.1 国外关于体育预测的相关研究

国外关于体育预测方面的研究有着较早开展历史和开展经验,早在二战结束后,

美苏两国进入冷战阶段，更是将奥林匹克运动会和世界级比赛作为第二战场。追溯到上个世纪 60 年代，随着战争带来的科技进步，科学、系统的预测方法就逐渐开始运用到了竞技体育比赛中。随着时间的推移，欧美国家从最早的单纯的对比赛成绩进行预测，逐渐演变到对运动员个人甚至运动项目未来发展演变的预测。

圣保罗大学的 João Gustavo Claudino^[18]等人通过运用人工智能和人工神经网络等方法对运动员的个人能力进行分析及预测，在对运动损伤风险的预测中应用了人工神经网络、决策树分类器和支持向量机。并把这些技术运用到了足球、篮球等项目中进行比赛成绩预测，结果都取得了不错的成果；使用 AI 智能对团队运动项目进行了进一步提升。

Adam Maszczyk, Artur Golas^[19]等对标枪项目运动员成绩进行预测，对比了非线性回归模型，神经网络模型得出神经网络模型的预测精度更高。

西班牙皇家橡树大学的 Rocio de la Iglesia^[20]率领的团队则通过生物学角度对运动员遗传因素这一方面进行分析预测，通过设计遗传硬化预测评分 (GES)，有助于寻找遗传天赋比较出众的运动员，然后改进他们的训练、饮食以及居住环境，有利于进一步提高运动员成绩，并提供参考和依据。

神经网络算法在 NBA 比赛中也得到了成功的应用：收集球队和球员的历史数据，根据数据分析的结果预测该队伍是否能够胜出，避免了被弱旅战胜的可能，及时调整球员的出场顺序，以此增加获胜的可能^[21]。

澳大利亚足球联赛也成功将神经网络算法应用于赛事球员的选择，通过模拟比赛得分数据，预测出参与者的运动素质（包括体能、心理和技能），挑选出其中最优秀的球员，提高了获胜的成功率^[22]。

越来越多的数据挖掘技术和相关算法应用于体育赛事数据的分析和预测中，学者们尝试添加更多的数据特征并减少冗余特征、将科学的模型进行结合，从而大幅提高预测准确率。整体来说，各国都十分重视机器学习算法在体育领域的应用，取得了不同程度的突破。而神经网络算法作为机器学习算法的一种，为体育项目的成绩预测提供了广泛的发展空间，在预测和分析数据方面有着它得天独厚的优势，是值得我们关注的。

1.3.2.2 国内关于体育预测的相关研究

体育科学预测是体育发展战略研究中的一项极其重要的内容。体育科学预测是国家体育事业稳定且协调发展的基石，它让体育事业能够快速走上正轨，并且往有

计划、有目标的方向发展。综上所述,体育预测工作在体育发展和体育科学化管理中占据着极其重要的地位。随着科学发展的日新月异,引起了体育各界工作人员的高度重视,其中包括体育领导决策层、科研以及管理人员等等。大型体育赛事成绩的预测掀起了当今世界体育研究的热潮,对各国的体育战略发展目标的构建和体育决策管理发挥着越来越重要的作用。在体育预测方面,中国的起步稍晚。为了改变这一现状,中国的体育研究者们纷纷投身于体育预测工作当中去,争取通过有效的预测研究,给中国体育事业和体育科学打上一针强心剂。

蒋启飞^[23]等人运用文献资料法、数理统计法,对现代男子网球单打技术与综合实力关系进行了因子与回归分析,得出了如下结论:发球局得分、接发球局得分效率因子、发球成功率因子构成了现代男子网球单打技术综合实力评价的特征因子;以综合因子得分为因变量,高度相关性的10项技术指标为自变量构建了最优回归预测方程模型,并对模型进行实证检验。

蒋勇^[24]等人首次将灰色预测法应用于体育项目中,对洛杉矶奥运会的12项田径运动成绩进行预测,结果表明,预测平均精度达到了99%以上。

杨宗友^[25]等人收集了6届奥运会男子400米自由泳冠军的成绩,利用GM(1,1)模型对下一届奥运会男子400米自由泳冠军成绩进行预测,预测成绩为220.36秒,大胆地预测了该成绩将打破奥运会记录,并验证了该模型的可信度。

陈亮^[26]在研究中利用灰色预测模型中的发展系数,通过对第三到第十四届田径世界锦标赛男子各项前三名成绩,其中13项径赛进行发展趋势的研究。并利用三数据建模预测法对第十届田径世界锦标赛11项男子径赛项目成绩预测。根据预测结果得出:男子短距离径赛、中长距离径赛、田赛和十项全能项目都呈下降趋势,而长距离径赛、竞走和跨栏项目则呈增长态势。

针对网球四大满贯赛事,文献^[27]对2008年美国网球公开赛和2009年澳大利亚网球公开赛的单打比赛数据进行相关性分析,得出了在男子网球比赛中一发成功率、一发得分率、二发得分率和接发球得分率为比赛获胜的关键制胜因子。基于这4个因子建立回归方程来判别大满贯比赛的胜负,准确率达到了94%。

综上所述,各类型的成绩预测法在体育领域中已经有了相关研究,同时对于一些大型赛事的预测较为准确。国内研究主要集中在通过数据对运动员在比赛中的成绩、名次预测,以及对现阶段大学生体育健康水平的预测。在体育成绩预测方面则是对田径、游泳和体操等项目的预测,上述的几个预测方法适用于短期预测和数据

指标较少的预测,而网球比赛的特点具有更多的不确定性因素,网球比赛中各个指标对比赛的影响并不呈线性的关系,上述几种基于线性回归式的方法可能在预测网球比赛方面带来较大的误差,而之前提到的神经网络可以有效的改进和解决这类问题,因此,要想进行更加准确的预测,神经网络是更合适的方法。

1.3.2.3 BP 神经网络算法在体育预测方面的相关研究

BP 神经网络能够解决传统预测方法无法解决的问题。BP 神经网络相当于人的大脑,能够对灵活运用所学习的数据,进行推理和记忆。目前, BP 神经网络日益受到人们的重视,同时其他学科的发展,为其提供了更大的机会^[28]。神经网络模型有许多种类,而 BP 神经网络是适用性最强的一种,尤其是在模式识别及分类、系统仿真、故障智能诊断、图像处理、函数拟合和最优预测等方面^[29]。

王杰^[30]在研究中得出:在乒乓球项目中,目标误差在合理范围内,神经网络方法能够拟合测试集的数据,该评价方法可实用性强,评价结果公正,能够正确区分比赛双方的技战术实力水准。

龚剑^[31]通过收集中国男篮在 2004 年奥运会、第 14、15 届男篮世锦赛和 2006 年洲际篮球赛的 4 项技术数据(包括 3 分球命中率、投篮命中率、助攻和后场篮板球),建立了人工神经网络预测模型,预测了 2008 年北京奥运会中国代表团的 4 项技术数据,凭借这些数据大胆预测中国男篮将取得佳绩。

宋爱玲等人通过研究得出:运动员的专项成绩和身体素质训练程度之间有明显的函数关系。为了实现运动员专项成绩的精准预测,建立了运动员专项成绩与身体素质关系的 BP 神经网络模型,该模型为指导运动员身体素质训练提供了科学的理论依据。

张宇^[33]等人以 23-29 届奥运会男子 100 米等 5 项田径比赛的成绩为训练集,在 Matlab 工具箱的支持下,构建 BP 神经网络模型,对第 30 届奥运会男子 100 米等 5 项田径比赛进行预测。研究结果显示, BP 神经网络模型可以实现对田径项目成绩的预测,模型不会过拟合,且预测精度高。

夏祥伟^[34]等人通过选取是否运动、锻炼形式、科学运动、重视程度、运动量(运动强度、运动时间、运动频率)等敏感指标,设计建立 BP 神经网络,定量地求出体育锻炼对研究生的生理、心理、社会 and 全面健康的影响系数,定性地研究体育锻炼对研究生全面健康的促进作用,并分析和提出促进研究生身心健康的相关建议,这对于研究生全面健康干预措施的决议有一定的指导性作用。

综上所述，BP 神经网络技术已经广泛的应用于体育方面，尤其是在预测方面。本研究通过对 BP 神经网络的研究学习，认为运用一定的软件和正确的指标体系，可以有效地建立 BP 神经网络预测模型来进行网球竞赛的预测。以上结果表明，BP 神经网络在体育预测方面来说，预测精度高。为了解决体育运动成绩中存在的不确定因素导致预测失准的问题，BP 神经网络具有以下两个特点。第一是 BP 网络算法能够快速学习和存贮大量的输入/输出模式映射关系，无需事先揭示描述这种映射关系的数学方程；第二是 BP 网络算法具有较强的自适应与自组织能力^[35]，各个指标之间的权重关系可以通过训练来得到，而不用人为的赋予权重值，它可以通过网络的学习来进行更新，并且通过新的数据将这个网络进行不断的更新和改进。而 BP 神经网络是近年来的研究热点，有着与传统方法不同的处理和表达信息的方式，也是目前应用最广泛的人工智能方法的一种。在网球项目中，技术指标近几年也提炼的越来越精细，数据化的水平也越来越高，因此，笔者认为 BP 神经网络可以有效地引入网球项目的预测当中去，对于网球项目成绩预测有很好的帮助。由于网球的最终比赛结果与其影响因素之间存在非线性关系，符合 BP 神经网络的特性，因此将 BP 神经网络尝试应用于网球项目的预测，是有可能得到很好的结果的。

1.4 相关概念及界定

1.4.1 BP 神经网络

1986 年，BP 神经网络第一次出现在人们的视野当中，它属于误差反向传播算法的多层前馈神经网络。BP 神经网络由三层结构组成，分别是输入层、输出层和隐含层。每一层含有若干个神经元，每一个神经元都代表着一个节点，每个节点之间都通过权连接，每个层之间的节点互相连接，这些连接形成了一个庞大且复杂的网络。

1.4.2 体育成绩预测

体育成绩预测指的是通过科学的预测方法，根据以往的成绩预测出某个项目集体或个人的体育成绩。这些预测往往应用于大型运动会的成绩预测，如：奥林匹克运动会、中华人民共和国全国运动会等。

1.4.3 网球四大满贯

网球四大满贯是澳大利亚网球公开赛、法国网球公开赛、温布尔登网球公开赛和美国网球公开赛的简称。每年依次各举办一次。

按比赛场地的不同来划分，其中，澳大利亚网球公开赛与美国网球公开赛为硬地赛，法国网球公开赛为红土赛，温布尔登网球公开赛为草地赛。

1.4.4 预测模型

预测模型是指运用特定的数学算法公式或者某些语言来描述存在于事物间的数量关系。它揭露了事物之间的内在关联，它一定程度上能作为计算预测值的直接依据。综上所述，预测模型与预测准确度息息相关。

2.BP 神经网络模型构建的理论基础

2.1 体育预测方法研究

随着近几十年的发展,体育预测方法形成了定量分析为主的预测体系。预测方法的选择对于不同项目有不同的预测精度,预测结果的准确性和所选取的预测方法之间具有很大的关系,体育预测也越来越依靠计算机和数学学科的参与。

2.1.1 灰色预测法

灰色预测法的提出是建立在灰色系统理论的基础上。灰色系统理论通过挖掘原始数据的变动规律,建立微分方程,预测未来事物的发展。该方法是通过少量、不完整的数据和信息,建立预测模型之后并进行预测的一种方法。GM(1,1)模型是常见的灰色预测模型,是一阶一变量的灰色方程模型,是采用线性方法建立的指数预测模型。杜娜娜^[36]利用该模型探讨了田径速度类项目的变化和发展趋势,曲淑华^[37]等人运用该预测模型对第16届世界田径锦标赛男子和女子的标枪成绩进行了预测,并给出了预测结果。

当预测数据呈指数变化时,灰色预测模型精度较高。而网球比赛成绩的变化是非线性的动态过程,并且数据并不是呈现递增的趋势,灰色预测模型不足以有效预测网球比赛成绩的变化。

2.1.2 德尔菲法

德尔菲法本质上是一种质性分析方法。德尔菲法在目前以定量分析的方法为主的体系下,能够对一些定量分析法进行一些必要的补充,获得有价值的观点和意见。但是它的局限性在于:(1)不能正确反映出体育项目各因素之间的关系。(2)专家的意见不完整或不切合实际。在网球比赛中,需要对各项数据进行量化分析,而这些数据对运动成绩来说是重要的因素,通过定量的分析方法才能够把握运动员的数据和成绩变化,否则就无法合理的进行预测。

2.1.3 时间序列预测法

时间序列预测法认为变量的现有值只与该变量的过去值以及过去的随机扰动项有关^[38]。

时间序列预测法属于定量预测,具有以下两方面基本原理:一方面承认事物发展的延续性,运用过去的时间序列数据进行统计计算和分析,推测出事物的发展趋

势；另一方面充分考虑到由于偶然因素影响而产生的随机性，为了消除随机波动产生的影响，利用历史数据进行统计分析，并对数据进行一定的处理，进行趋势预测^[39]。张小龙探讨了 ARIMA（时间序列）模型在体育中的运用并分析其可行性，结果表明 ARIMA 模型可用于体育统计、体质健康以及体育环境等多方面的预测^[40]。张锐等人也运用了时间序列法对中小学体质健康发展趋势进行了预测^[41]。时间序列预测法大部分适用于预测变化趋势明显、简单或周期性的数据，对于一些较复杂和非线性的数据而言，预测值会与实际值偏差较大，所以时间序列方法目前还无法满足预测网球项目成绩的要求。

2.1.4 回归预测法

回归预测法主要有一元线性回归和多元线性回归两种。通过回归分析来进行预测，也是体育预测领域常用的方法，国内外的研究较为普遍。魏春玲等人运用因子分析法构建了运动成绩的多元线性回归预测模型，具有较高的预测精度^[42]。王国凡等人建立了奥运会成绩预测模型，结果表明该方法预测精度高，稳定性好^[43]。采用回归预测法，需要提前准备大量的统计数据，对各项数据的特点较为熟悉，找出隐藏在数据背后的规律，快速地进行预测。回归预测法的计算速度快，操作简单，同时也存在着 2 个不足：（1）当样本数据不足或者数据存在冗余时，预测结果的准确率会下降，出现不同程度的错误结果。（2）回归模型变量不能太复杂，算法低级。回归分析法无法承载网球项目成绩较为庞大的数据，所以也并不适用于此。

2.1.5 预测方法评价

长久以来，国内对体育预测的研究侧重于时间序列预测法、灰色预测法和德尔菲法等传统方法，在对上述的几种方法的预测准确度做了大量的研究后，得出结论如下：在几种方法中，均能够进行预测并得到较好的结果，应针对不同的目的和不同的运动项目选择不同的预测方法。几种预测方法的特点见表 1：

表 1 预测方法特点比较

	灰色预测 法	德尔菲法	时间序列预测 法	回归预测法
数据类型	具有不确定性因素的数据	陈述事实、观点的数据	简单、趋势明显、周期性数据	连续性较强、自变量与因变量相关程度较高的数据

所需样本量	不需要大量的样本	不需要大量的样本	需要的样本量较多	需要的样本量少
所需样本的要求	样本不需要有规律性分布	样本有一定的规律性	个别呈随机变动，整体呈统计规律	有规律性的数据
适用范围	适用于短期、近期，中长期预测	定性预测方法，可作为定量方法的补充	多用于短期预测	中短期预测

大数据时代来临，随着计算机技术的高速发展，人工智能逐渐与各个行业深度融合，国内的研究学者们开始利用人工智能的学习方法进行建模和预测，并取得了较好的结果。

除了神经网络算法以外，其他的预测方法都很难准确的识别复杂数据之间的非线性关系，应用上述的几种方法对网球项目进行预测和分析，结果往往差强人意。BP 神经网络具有以下几个特点：（1）BP 神经网络模型是一种非线性的计算模型，能够有效的处理各种信息，依据数据本身的内在联系建模。（2）容错性高，对冗余数据和缺失数据的应用较多。（3）BP 神经网络能够自主学习，真实的反映出网球项目成绩的数据之间的非线性关系。（4）并行处理功能，提高计算速度和预测效率。因此，本文将 BP 神经网络技术应用于网球项目比赛成绩的预测，在预测模型的构建的适用性与可操作性都有其独特的优势，为网球项目比赛成绩的预测提供了无限可能和发展。接下来将介绍 BP 神经网络的相关内容和理论基础。

2.2 BP 神经网络的基本理论

BP 神经网络是误差反向传播神经网络（Error Back-propagation Neural Network），是一种有导师指导的学习算法。基本思想是：（1）对具有一定数量的样本进行训练和学习。（2）利用梯度下降法不断反馈误差且修正权值，减小误差。BP 神经网络的实质是不断调整权值使网络计算结构足够精确的过程。

BP 神经网络所用的基本方法为梯度下降法，本节以一个三层神经网络为例，对 BP 神经网络做一个说明，三层 BP 网络结构如图 1 所示。BP 网络模型主要包括输入

层、隐含层和输出层^[44]。输入层包含 n 个神经元，记为 $X = (x_1, x_2, x_3 \cdots, x_n)$ ，隐含层包含 m 个神经元，记为 $Y = (y_1, y_2, y_3 \cdots y_m)$ ，输出层包含 k 个神经元，是实际输出值记为 $Z = (z_1, z_2, z_3 \cdots z_k)$ ，同时期望输出值也包含 k 个神经元，记为 $D = (d_1, d_2, d_3 \cdots d_k)$ ，权值 $W = (W_1, W_2, W_3 \cdots W_q)$ ，阈值 $\vartheta = (\vartheta_1, \vartheta_2, \vartheta_3, \dots, \vartheta_p)$ ，权值是输入输出向量维数有关的矩阵向量，阈值是与隐含层相关的向量。

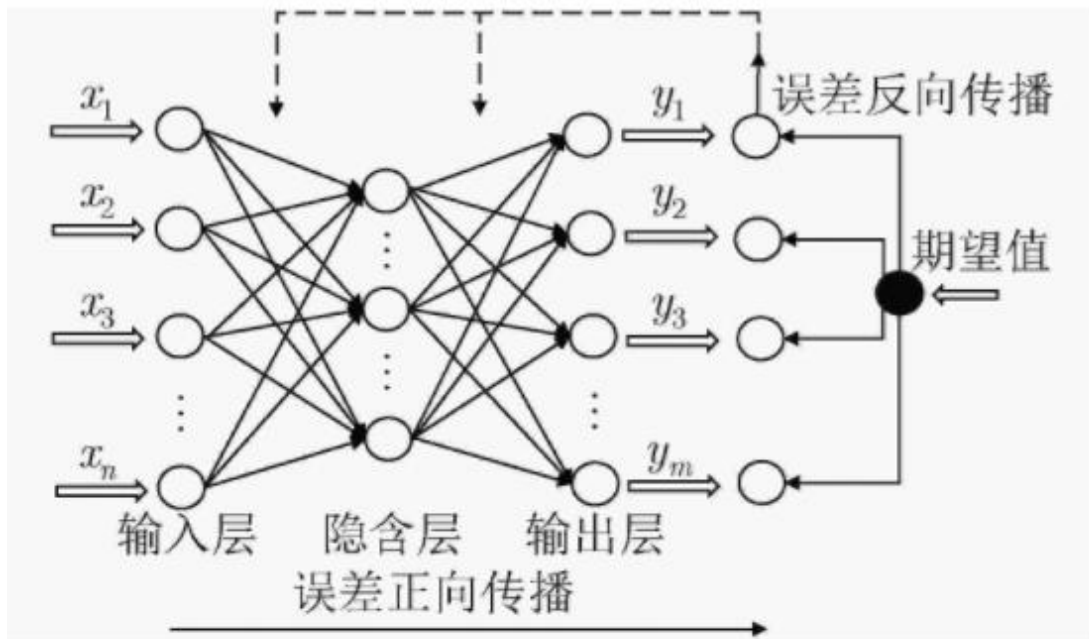


图 1 BP 神经网络结构图

假设 k 为迭代次数对于每一层的权值和阈值简单来说用下面公式进行修正：

$$x(k+1) = x(k) - \eta g(k) \quad (2-1)$$

其中 $x(k)$ 表示第 k 次迭代各层之间的权向量或者阈值向量， η 代表学习效率。

$g(k) = \frac{\partial E(k)}{\partial x(k)}$ ，表示的是第 k 次迭代神经网络输出的权值和阈值的梯度向量。

对于神经网络中的重要权值 W 做随机赋值。定义激活函数 f ，根据定义的输入向量和权值向量，结合激活函数计算得到隐含层的数据向量和输出层实际输出向量。

隐含层 $y = (y_1, y_2, y_3 \cdots y_n)$:

$$y_i = f(net_i), net_i = \sum_{j=1}^m \omega_{ij} x_j - \vartheta_i, j = 1, 2, 3 \dots m \quad (2-2)$$

实际输出层 $Z = (z_1, z_2, z_3 \dots z_k)$:

$$z_i = f(\text{net}_i), \text{net}_i = \sum_{j=1}^m \omega_{ij} y_j - \vartheta_i, j = 1, 2, 3 \dots n \quad (2-3)$$

计算实际输出和期望输出之间的平方误差:

$$E(\omega) = \frac{1}{2} (D - Z)^2 = \frac{1}{2} \sum_{i=1}^k (d_i - z_i)^2 \quad (2-4)$$

根据梯度下降法的基本定义, 选择误差减小最快的方向就是对平方误差函数求对 ω 的偏导数, 可以得到权值向量和阈值向量的调整公式。

$$\Delta W_j = \omega_j(N+1) - \omega_j(N) = -\eta \frac{\partial E(\omega)}{\partial \omega} \quad (2-5)$$

$$\Delta \vartheta_j = \vartheta_j(N+1) - \vartheta_j(N) = \eta \delta \quad (2-6)$$

输出层权值与阈值调整:

$$\frac{\partial E(\omega)}{\partial \omega_j} = -(d_j - z_j) f'(W_j^T Y) Y, j = 1, 2, 3 \dots n \quad (2-7)$$

$$\Delta W = \omega_j(N+1) - \omega_j(N) = \eta (d_j - z_j) f'(W_j^T Y) Y, j = 1, 2, 3 \dots n \quad (2-8)$$

$$\Delta \vartheta_i = \vartheta_j(N+1) - \vartheta_j(N) = \eta (d_j - z_j) f'(W_j^T Y), j = 1, 2, 3 \dots n \quad (2-9)$$

隐含层的权值与阈值的调整:

$$\frac{\partial E(\omega)}{\partial \omega_j} = -[\sum_{i=1}^k (d_j - z_j) f'(W_j^T Y) \omega_i] f'(W_j^T X) X, j = 1, 2, 3 \dots m \quad (2-10)$$

$$\Delta W = \omega_j(N+1) - \omega_j(N) = \eta [\sum_{i=1}^k (d_j - z_j) f'(W_j^T Y) \omega_i] f'(W_j^T X) X, j = 1, 2, 3 \dots m \quad (2-11)$$

$$\Delta \vartheta_j = \vartheta_j(N+1) - \vartheta_j(N) = \eta [\sum_{i=1}^k (d_j - z_j) f'(W_j^T Y) \omega_i] f'(W_j^T X), j = 1, 2, 3 \dots m \quad (2-12)$$

通过不断重复和修正使得误差最小, 直至期望值与实际值达到使用要求为止。

上述内容叙述的是最为经典的梯度下降法, 还有动量 BP 算法和 LM 算法, 以上算法都属于误差修正的算法, 并广泛地应用在 Matlab 中。

2.3 本章小结

本章节首先对体育预测的相关方法进行了探讨并作了特点的比较，选择了 BP 神经网络作为本文的预测方法。其次介绍了 BP 神经网络的基本方法（梯度下降法），为后续 BP 神经网络的模型构建奠定了坚实的理论基础。

3.基于 BP 神经网络的网球赛事成绩预测模型构建

模型的整体框架如图 2 所示，主要由特征选取、数据处理、模型参数设定和实现、模型训练测试及评价和预测结果对比与分析这 5 个部分组成：

(1) 特征选取部分通过对影响网球比赛成绩的相关因素进行分析，高效、快速地从数据来源处获取有价值的数据，利用这些数据建立网球比赛的指标体系。(2) 数据处理部分主要解决如何将互联网平台上收集到的数据转换成可以直接输入 bp 神经网络算法的问题，通常有标准化、归一化等方法，根据网球比赛的特点来对数据进行处理。(3) 模型参数设定和实现是核心部分，通过设定函数和相关算法对模型进行设计，并用 Matlab 软件将模型具体实现。(4) 利用 Matlab 神经网络工具箱所提供的训练函数对处理过的数据样本进行训练和测试，最后将训练好的模型进行评价。(5) 将模型训练后的结果与真实结果进行对比和分析，这部分的内容将放在第四章重点呈现。模型的整体框架如图所示：

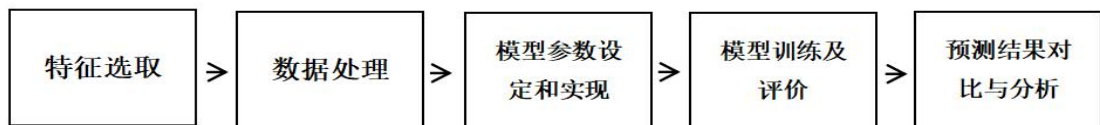


图 2 BP 神经网络模型的整体框架

3.1 BP 神经网络的特征选取

3.1.1 影响网球比赛成绩的相关因素

网球项目竞赛成绩在一定程度上受到一些技术指标和相关因素的影响，如天气因素，技战术因素等。数学公式或是某种语言并不能正确表达出它们相互作用的关系。这种不可抗力的关系和因素，为精准预测比赛制造了许多障碍。结合网球比赛的特点，归纳出影响网球竞赛成绩的 4 点因素，并进行以下分析：

(1) 体能因素：网球运动的特点是比赛时间长、运动负荷大、需要进行大范围的间歇性移动并挥拍击球，这就对运动员的体能方面提出了很高的要求。体能是基础，好的体能会让你的动作不易变形，帮助运动员成绩的提高。如一发得分率、二发得分率、主动得分和非受迫性失误等指标都跟体能因素相关。

(2) 心理因素：一位优秀的网球运动员除了拥有好的体能之外，还要拥有良好的心理素质，才能冷静的观察对手的球路和弱点，对自己技战术的运用充满信心，对夺取比赛胜利充满渴望。如破发成功率、接发球得分率等指标都与心理因素有关，在关键时刻能够临危不惧，敢打敢拼。

(3) 技战术因素：技战术是网球运动员比赛的基础，现代网球运动员的技战术和打法更加的全面，运动员击出的球力量更大，旋转更足、落点更刁钻。随着技术的丰富，战术也得到了相应的变化，球拍的科技含量越来越高，场地的弹性和粗糙程度改变，网球运动的主流打法也在慢慢变化，过去网球运动员多数采用发球上网战术，现代网球运动员可供选择的战术更多，如发球上网、底线相持、底线进攻并伺机上网等，对于网球比赛起着至关重要的作用。

(4) 其他因素：在网球比赛中，客观因素对比赛也有直接影响，例如天气、场地类型、球员对场地的适应性及身体状况，球员的赛前准备是否充分等一些因素着实影响着比赛，这些无法提前预知的因素，也是影响比赛预测结果准确率的关键因素。

影响网球赛事成绩的因素有很多，主要表现在技战术、体能和心理等方面，这些因素当中有部分因素是无法量化成具体的指标，本文仅从技术数据的角度来分析网球赛事成绩，部分因素没有考虑在内，但是这些因素或多或少都有在技术指标中有所体现。例如，在面对破发点的时候，球员的心理因素也会对技术数据有所影响；在长时间的拉锯战中，球员的体能也会令技术数据产生变化。因此，选择合适的预测算法和模型成为解决此问题的主要环节。综上所述，结合 ATP 官方网站和相关数据网站的支持，对 BP 神经网络模型的指标进行建立。

3.1.2 BP 神经网络的数据来源

大满贯赛事是职业网球比赛中最具有含金量的比赛，数据丰富且具有代表性，常规技术统计已在四大满贯和全球各大赛事中广泛使用，可作为科学学术研究、检验教学训练工作和提升训练质量的客观依据。运用数理统计法统计四大满贯赛事各项技术统计数据，统计完成并对数据进行处理后，为模型的构建提供数据基础。

网球赛事成绩预测模型构建是以一定的数据资料为基础的。本文数据的来源主要有两个：（1）通过 ATP 官方网站及四大满贯官方网站公布的赛后技术统计（2）爱奇艺体育 APP 发布的大满贯赛事视频中的赛后技术统计作为研究指标，收集了

2019-2021 年四大满贯的前十六强的选手及其对阵选手的对阵数据共计 11 个大满贯赛事, 693 场比赛, 共计 13676 个数据 (由于每个赛事都有少数选手退赛, 退赛场次共计 18 场, 这 18 场比赛的数据不作记录), 由于受到新冠肺炎疫情的影响, 2020 年温布尔登网球公开赛暂停举办。

3.1.3 BP 神经网络的量化指标体系建立

一般认为, 网球项目的基本技术是指正手击球、单手反手击球、双手反手击球、发球和截击 5 项。上一节详细剖析了影响网球比赛成绩的因素, 所以我们在量化技术指标时, 既要全面, 又要在量化指标体系时选取代表性较强的典型指标。

通过查阅 ATP 官方网站和相关体育赛事网站, 专家们普遍认为这几项技术数据: 一发成功率、一发得分率、二发得分率、接发球得分率、破发率、总得分率、ACE 球、双误、非受迫性失误和主动得分能反映出比赛双方的技战术运用情况, 且统计中又易于判断和记录。文献^[45]主要探索了上述 10 个指标对于职业男子网球单打比赛胜负结果的作用, 很好的诠释了男子职业网坛的发展, 对运动员训练起到了帮助的同时准确预测了 94% 的比赛胜负。文献^[46]利用上述十项指标进行多元回归分析, 胜负结果对实际情况对比, 准确率达到了 84.5%。综上所述, 所选的 10 项指标已经能够反映出一场职业网球比赛的过程, 满足指标体系构建的要求。

因此, 本文将以一发成功率、一发得分率、二发得分率、接发球得分率、破发率、总得分率、ACE 球、双误、非受迫性失误和主动得分共计 10 项技术统计数据来进行 BP 神经网络模型的构建, 将 2019-2020 年的 7 个大满贯赛事的数据作为训练集, 2021 年的 4 个大满贯赛事作为测试集, 输入层为所收集到的大满贯 16 强选手的技术统计数据, 输出层为大满贯 16 强选手的最终比赛成绩。

所选的特征可以用来表示一场比赛中一位选手表现的总体情况, 例如: 一发成功率高且一发得分率和二发得分率高、ACE 球数量多表示该选手在当前的比赛中发球效率高, 在发球局中能够掌握更多的主动权, 之后的进攻比较流畅, 接发球得分率和破发率高, 说明选手在接发球端做的较好, 且防守强悍, 相反来看, 选手的双误和非受迫性失误多表明选手在场上的状态不好, 主动得分多说明选手在进攻端手感火热, 能够打出对手难以招架的球路, 在场上战术意图和思路清晰。

考虑地理环境、赛事场地的差异性, 我们以 2019 年澳大利亚网球公开赛为例, 所收集的选手的最终成绩与部分选手原始对阵数据展示如下:

表 2 2019 年澳大利亚网球公开赛前 16 强比赛结果

16 强	8 强	4 强	亚军	冠军
巴斯塔	锦织圭	普依	纳达尔	德约科维奇
梅德韦杰夫	拉奥尼奇	西西帕斯		
兹维列夫	巴蒂斯塔阿古特			
西里奇	蒂亚福			
费德勒				
迪米托夫				
伯蒂奇				
丘里奇				

表 3 德约科维奇 2019 澳网男单前三轮技术统计数据

	德约科 维奇	克鲁格		德约科 维奇	特松加		德约科 维奇	丹尼斯 沙特鲁
一发成 功率	73%	66%	一发成 功率	73%	65%	一发成 功率	61%	56%
一发得 分率	77%	48%	一发得 分率	80%	69%	一发得 分率	75%	70%
二发得 分率	67%	24%	二发得 分率	60%	45%	二发得 分率	63%	32%
接发球 得分率	46%	20%	接发球 得分率	55%	32%	接发球 得分率	46%	30%
破发率	67%	17%	破发率	25%	20%	破发率	62%	60%
总得分 率	62%	38%	总得分 率	60%	40%	总得分 率	57%	43%
ACES 球	2	1	ACES 球	4	15	ACES 球	5	6

双误	1	1	双误	1	6	双误	4	5
非受迫性失误	30	65	非受迫性失误	15	32	非受迫性失误	33	57
主动得分	28	10	主动得分	40	48	主动得分	16	21

表 4 德约科维奇 2019 澳网男单后三轮技术统计数据

	德约科维奇	梅德韦杰夫		德约科维奇	卢卡斯普勒		德约科维奇	纳达尔
一发成功率	67%	65%	一发成功率	61%	56%	一发成功率	74%	75%
一发得分率	73%	70%	一发得分率	75%	70%	一发得分率	84%	70%
二发得分率	50%	35%	二发得分率	63%	32%	二发得分率	60%	33%
接发球得分率	42%	35%	接发球得分率	46%	30%	接发球得分率	39%	22%
破发率	39%	22%	破发率	62%	60%	破发率	50%	0%
总得分率	53%	47%	总得分率	57%	43%	总得分率	59%	41%
ACES 球	6	18	ACES 球	5	6	ACES 球	11	13
双误	5	6	双误	4	5	双误	0	0
非受迫性失误	50	49	非受迫性失误	30	65	非受迫性失误	23	34

主动得 分	43	38	主动得 分	25	18	主动得 分	29	24
----------	----	----	----------	----	----	----------	----	----

以上数据为德约科维奇在 2019 年澳大利亚网球公开赛中的对阵数据，其中 8 进 4 时，对手退赛。因此，有共计 6 个数据样本。在观察前 3 个数据（前三个轮次）可以发现，德约科维奇在这 10 项数据中几乎都领先于对手，可以说明在技术层面，德约科维奇的自身实力较强，战胜对手较为容易。但在进入更深的轮次后，所遇到的对手也实力也更加强大，数据差异相对小。通过观察后 3 轮的数据，在各项数据差异较小的情况下，在场上的实际情况往往比较胶着，比赛常常趋于白热化。

3.2 BP 神经网络输入数据的预处理

在 BP 神经网络的训练中，要保证数据在同一个量纲，需要对网络的输入数据和输出数据进行处理，处理的方法有许多种，这里我们采用归一化（标准化）的方法对数据进行处理，目的是让这两个特征的值压缩至 $[0, 1]$ 之间，减少由于和其他特征尺度不同造成的影响，如果不对数据进行归一化处理，就会造成一些大数值的输出误差加大，数值小的输出误差变小。

在本研究中，我们都采用标准化的方法对原始数据进行处理，常用的公式为：

$$y = \frac{x - \min}{\max - \min} \quad (3-1)$$

其中， x 为当前需要处理的数据， \min 为数据中的最小值， \max 为数据中的最大值， y 为最后计算出的结果。我们通过该公式的计算后将数据归一化到 $[0, 1]$ 范围内。（在下文中，Minmaxscaler 函数可以满足数据处理的要求）

根据 BP 神经网络原理，通过大量的数据集训练得出神经网络模型，因为一项竞技赛事中可进行预测的训练集的数据量有限，如果每个特征取值过多，导致训练样本集中样本的个数少于特征组合的个数，那么模型将会过拟合，即训练出来的模型在训练集上的表现优秀，但是在测试集的表现较差，对未知样本的预测表现一般。我们需要对每个特征的取值进行归一化处理，从而降低均值方差（MSE），避免模型过度拟合。

下面的表为 2019 年澳大利亚网球公开赛数据进行归一化处理后的 10 项数据统计，该表也是用于 BP 神经网络模型训练的实验数据集（训练集的一部分，2019 年澳

大利亚网球公开赛的部分比赛对阵数据)：

表 5 2019 年澳大利亚网球公开赛比赛数据

一发 成功 率	一发 得分 率	二发 得分 率	接发 得分 率	破发 率	总得 分率	ACE 球	双误	非受 迫性 失误	主动 得分
0.61	0.75	0.63	0.46	0.62	0.57	0	0.36	0.4	0
0.56	0.7	0.32	0.3	0.6	0.43	0.07	0.45	0.85	0.18
0.67	0.73	0.5	0.42	0.39	0.53	0.07	0.45	0.72	1
0.65	0.7	0.35	0.35	0.2	0.47	1	0.54	0.7	0.81
0.61	0.75	0.63	0.46	0.62	0.57	0	0.36	0.35	0.33
0.56	0.7	0.32	0.3	0.6	0.43	0.07	0.45	1	0.07
0.74	0.84	0.6	0.39	0.5	0.59	0.46	0	0.22	0.48
0.75	0.7	0.33	0.22	0	0.41	0.61	0	0.42	0.29
0.56	0.65	0.41	0.31	0.67	0.44	0.53	1	0.53	0.88
0.67	0.74	0.59	0.46	0.5	0.56	0.07	0.18	0	0.81
0.67	0.66	0.32	0.27	0	0.4	0.15	0.27	0.35	0.25
0.65	0.81	0.57	0.45	0.56	0.6	0.07	0.36	0.07	0.62
0.47	0.7	0.44	0.25	0	0.4	0	0.27	0.38	0.29
0.69	0.8	0.64	0.44	0.38	0.6	0	0.09	0.11	0.59
0.75	0.7	0.33	0.22	0	0.41	0.61	0	0.42	0.29
0.74	0.84	0.6	0.39	0.5	0.59	0.46	0	0.22	0.48
0.69	0.64	0.3	0.2	0	0.37	0	0.18	0.2	0.03

如表 5 所示，为 2019 年澳大利亚网球公开赛经过归一化处理后的数据，也是实验数据集中的一部分。从表中可以看出，前 10 列数据是对阵双方选手在预处理后的特征数据，每 2 行为一场比赛中的对阵选手。利用这些处理后的数据构造特征函数，运用 BP 神经网络原理构造符合训练样本集的模型。

3.3 BP 神经网络参数设定和模型的实现

3.3.1 BP 神经网络参数的设定

3.3.1.1 BP 神经网络设计

BP 网络设计是一个复杂的问题，它应该需要满足不同的要求。举个例子，期望所设计的网络有良好的泛化能力，训练效率高等。在普遍的情况之下，在体育比赛成绩预测研究中，数据量是固定的，所以可以归纳为在样本量固定的情况下，如何选择网络规模的问题。进行 BP 网络预测模型设计时，笔者所考虑的因素如下：初始化参数的设定、激活函数的确定、网络的层数、每层中的神经元个数、学习率、迭代次数和目标误差等。

3.3.1.2 输入层、输出层的确定

输入层是进行网络设计的第一步，在本文中，输入层即为四大满贯的技术指标，输入层的节点数为输入层的参数来决定的，选取技术指标为 BP 网络的输入层节点数（例如：输入层的参数为 10 项指标，即节点数为 10）。输出层是选手在某项赛事的最终成绩，节点数为 1。本文根据预测结果的不同来调整网络的输出指标，例如：当预测结果设定为胜负时，1 代表选手获胜，0 代表选手告负；当预测结果设定为 16 进 8 乃至后面更深轮次的预测时，（1, 0, 0, 0, 0）代表选手获得了冠军，（0, 1, 0, 0, 0）代表获得了亚军的成绩，以此类推。

3.3.1.3 初始化参数的设定

本文的 BP 神经网络采用梯度下降法（具体在 2.2 有介绍）作为网络的训练方法，其训练函数为 `traindm` 函数，动量项系数 $\eta = 0.7$ ，最大训练次数为 1000，训练目标的误差设置为 0.1。初始的阈值和权值均为系统默认的值，其他参数均为缺省值。

3.3.1.4 学习率

在 BP 神经网络中，大部分学习过程都少不了函数的设置，学习率也是如此，学习率过低会导致模型训练时间加长，学习率过高，训练的网络不收敛甚至发散。经过对学习率的调整，得到图 9 所示的误差折线图，可以看出，当学习率设置为 0.002-0.014 之间时，模型的误差均不超过训练目标误差 0.1，且学习率设置为 0.01 时训练误差最小。

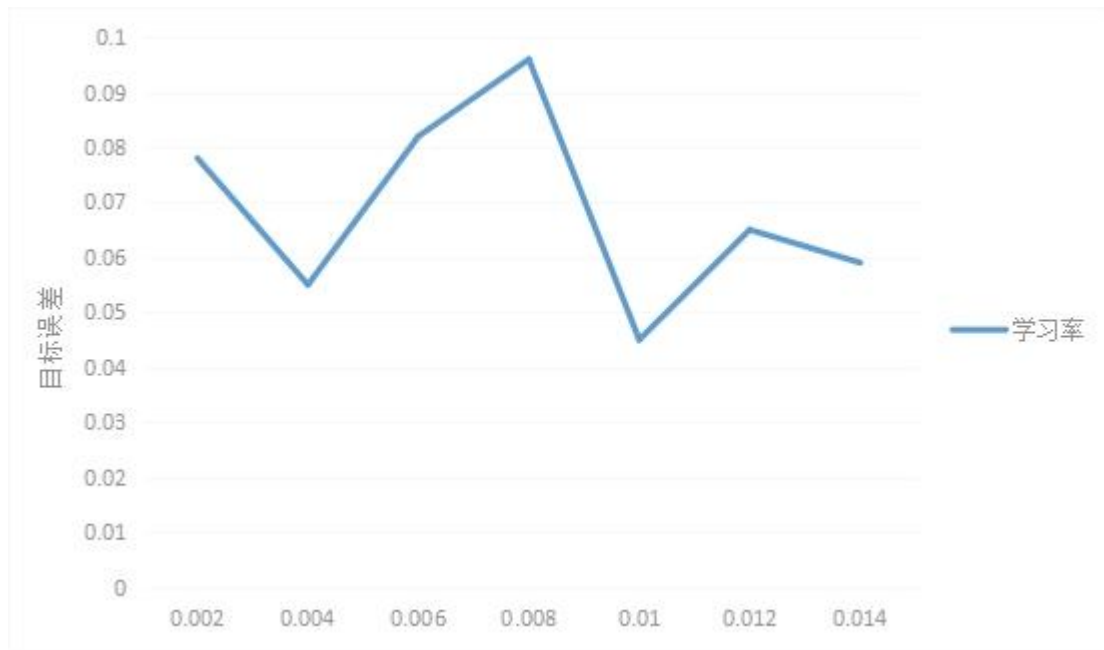


图 3 学习率的设置对目标误差的影响

3.3.1.5 隐含层节点数的确定

本章的数值模拟选取了 3 层结构的 BP 神经网络来进行预测模型的建立。BP 神经网络的隐含层节点通常用试凑法^[47]确定,预先设置较少的隐含层节点来训练网络,之后逐渐增加隐含层节点数,并使用同一个样本集进行训练,最后选取网络误差最小时,所对应的隐含层节点数。在数值模拟试验中,一般根据经验公式得出一个粗略的估计值作为初始值,然后使用试凑法确定最佳节点数。常用的经验公式为:

$$m = \sqrt{n+1} + \alpha \quad (3-2)$$

其中, m 为隐含层节点数, n 为输入层节点数, 1 为输出层节点数, α 为 1 到 10 之间的整数。本文设计的 BP 神经网络的输入层节点个数为 10, 输出层节点个数是为 1, 根据经验公式得出该神经网络隐含层的神经元个数应该为 5-14, 接着在 5-14 之间进行筛选, 确定隐含层节点数对目标误差的影响。如图 10 所示为某工况训练下在选取不同隐含层节点数时目标误差处于起伏的状态。在该工况下当隐含层节点数为 8 时, 目标误差最小。

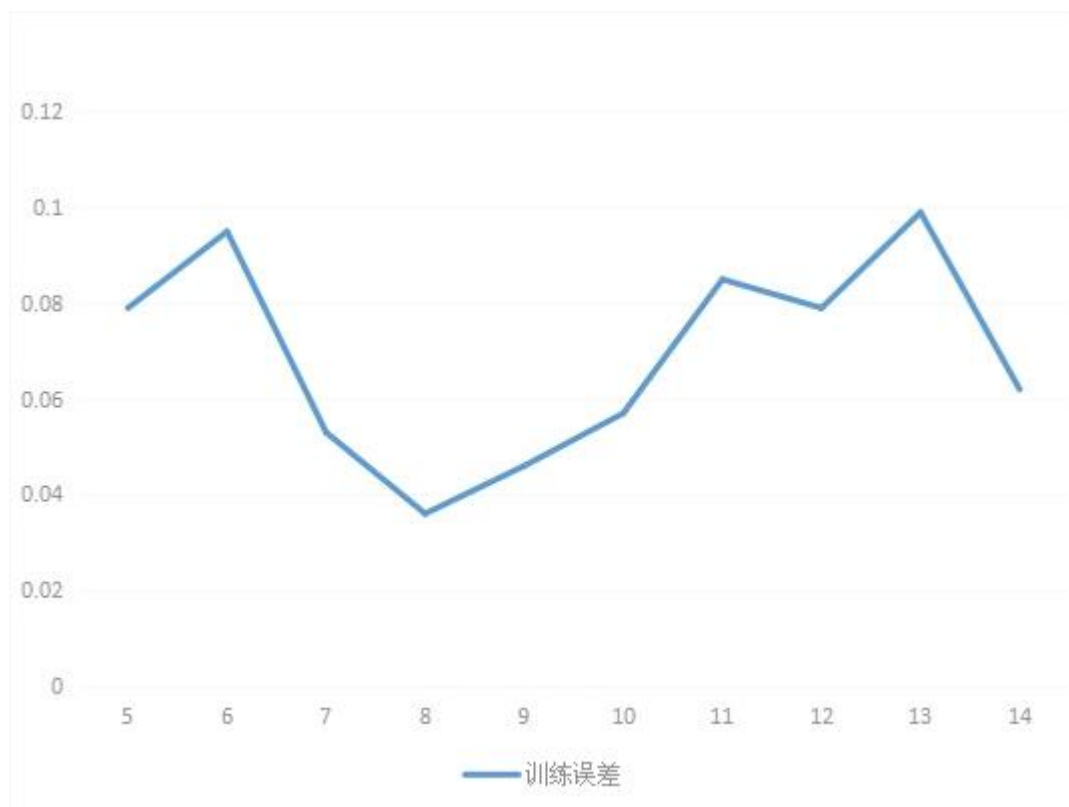


图 4 隐含层节点数对目标误差的影响

3.3.1.6 激活函数的确定

每个神经元的数学模型取决于不同的激活函数，这让神经网络能够采取不同的信息处理模型来应对各种类型的数据。其中，信息模式是神经网络整体性能的其中一大要素，合适的激活函数能加快神经网络的运行时间和提高预测的精确度。本文是针对网球竞赛的预测问题，对于三层结构的 BP 神经网络选取不同的激活函数。

输入层：输入层接收到的训练样本是离散数据（数值只能用自然数或整数为单位的数据），不需要经过函数转换，所以输入层使用一般的线性激活函数即可。

$$f(x)=x \quad (3-3)$$

隐含层：为了增强神经网络处理非线性关系的能力，本文在隐含层选择的激活函数是 Sigmoid 函数。

$$S(x)=\frac{1}{1+e^{-x}} \quad (3-4)$$

Sigmoid 函数为非线性函数，在信号特征空间的映射效果较好，能够将神经元的输入变量映射到[0, 1]区间，数据在转换过程中不容易发散。

输出层：隐含层的激活函数 Sigmoid 函数的值域是 $[0, 1]$ ，而网球于此问题的输出结果是一个 $[0, 1]$ 区间内的概率值，所以需要将输入变量映射到相应区间内，本文在输出层采用的激活函数是 relu 函数。公式为：

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3-5)$$

Sigmoid 函数在输入值处于 $[-1, 1]$ 之间时，函数值变化较为敏锐，超出区间就失去敏感性，处于饱和状态，影响预测精度。Relu 的输入值和输出值能够保持非线性单调上升和下降关系，契合 BP 神经网络的梯度求解法则，收敛速度较快，所以本章的隐含层选择使用 Sigmoid 函数，在输出层使用 relu 函数。

3.3.2 BP 神经网络的 Matlab 实现

Matlab 是有着出色编程效率的软件。Matlab 能帮助用户解决复杂繁琐的运算，拥有极高的图形处理能力，将数据快速转换成图像，实现可视化。除此之外，Matlab 还拥有功能齐全的用户工具箱，其中包含通信工具箱和信号处理工具箱等等。综上所述，Matlab 应用广泛，它的计算程序是由资深的专家编写，具有稳定可靠的性能。

3.3.2.1 BP 神经网络模型的程序设计

构建预测模型及预测主要利用的 Matlab 程序语言包括：

① newff () 函数：前馈网络创建函数 newff 函数可以表示为：

`net = newff (P, T, S, TF, BTF, BLF, PF, IPF, OPF, DDF)。`

如上所示，P 值代表一定样本量的输入数据；T 值表示前馈网络的输出数据；S 值表示每个神经元的结果；TF 代表前馈网络的传递函数，BLF 代表该模型的训练函数，OPF 则为神经网络模型的输出函数；在神经网络使用过程中，一般设置以上六个参数，其余参数在普遍情况下为默认数值。

②train：训练一个神经网络。

train 函数的语法代表`[net, tr] = train(NET, X, T, Pi, Ai)`

其中，NET 表示待训练网络；X 表示输入数据；T 表示输出数据；Pi 表示输入层条件；Ai 表示输出层条件；net 表示训练好的网络；tr 表示训练过程记录。

在日常使用中，一般仅需设置前三个参数，Pi 和 Ai 通常为默认数值。

③sim：使用网络进行仿真。

Sim 函数的语法代表`Y=sim(net, X)`

上述函数中，net 表示网络；X 表示输入给网络的 $K \times N$ 矩阵。 $K \times N$ 矩阵中，K 表示网络输入个数，N 表示数据样本数；Y 表示输出矩阵 $Q \times N$ ，Q 表示网络输出个数。

3.3.2.2 网球成绩预测模型编程设计

将训练样本数据（之前收集的 7 个赛事的数据和结果）归一化后输入网络，BP 神经网络参数的设定如 3.3.1 章节所示。本文借助 Matlab 神经网络工具箱中的相关函数，实现对网球成绩预测模型的构建、训练与测试。BP 神经网络的训练程序主要代码见附录。

3.4 BP 神经网络模型的训练及评价

3.4.1 BP 神经网络模型的训练

我们将 2019-2020 的四大满贯赛事技术统计数据指标量化后作为输入层，2019-2020 的四大满贯赛事的最终比赛结果作为输出层输入到 BP 神经网络模型中进行训练，训练的具体步骤如下所述：

（1）通过收集大满贯赛事的数据和比赛结果，单个赛事的单个技术数据相当于 BP 神经网络中的一个神经元，输入层为网络训练的多个指标，输出层为大满贯赛事的 16 强参赛选手的最终成绩。

（2）将 BP 神经网络模型初始化，权重和阈值设置为 $[0, 1]$ 的数值，设定模型参数（迭代次数、学习率等）。

（3）判断训练次数是否小于设置的最大学习次数，成立则结束网络训练，不成立则进行第 4 步。

（4）正向计算每个网球技术数据在网络中的输出。

（5）计算 BP 神经网络中的权值误差。

（6）更新网络中的权值和阈值。

（7）计算全局误差。

（8）判断所有技术数据输出完成时，与输出层（原赛事的比赛名次）的结果是否符合，若符合则结束学习，反之则回到第三步重新开始计算。

模型训练完成后,需要对模型进行评价,本文采用 MSE 性能函数和 5 折交叉验证法对模型进行评价。

3.4.2 模型过拟合及交叉验证

过拟合指的是模型在训练集上表现优秀,但在交叉验证集合测试集上表现一般,可以得出模型对未知样本的预测表现较差,泛化能力一般。出现过拟合情况,表明模型尚未学习到数据的本质规律,从而导致模型的预测能力不理想。在实际情况中,我们所设计的模型要能够进行使用,需要反复对模型进行调整,防止模型出现过拟合的情况,只有保证模型的泛化能力,模型的构建才有意义。因此,交叉验证在模型构建的过程中显得极其重要。

在 sk-learn 中,可以使用训练集/测试集拆分和交叉验证的方法来避免过拟合现象的,如下图所示,将数据集进行拆分,在经过交叉验证后得到模型最优参数,最后得到稳定可靠的模型。

本章以 sk-learn 工具包和笔者收集的数据集进行说明:

```
# import necessary packages
import numpy as np
from sklearn.neural_network import MLPClassifier,MLPRegressor
from sklearn.model_selection import train_test_split
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import KFold
```

图 5 设置的部分函数和交叉验证代码图

将数据集拆分为 7 个赛事数据集和 4 个赛事数据集,并且进行交叉验证的代码如下:

```
# 4. 5-Fold cross-validation
kf = KFold(n_splits=5)
for train,test in kf.split(feature):
    x_train = feature[train]
    y_train = rank[train]
    x_test = feature[test]
    y_test = rank[test]
    mlp.fit(x_train,y_train)
    pred = mlp.predict(x_test)
```

```
print('each folds MSE=', np.mean((y_test-pred)**2))
```

为了避免出现过拟合的情况，sklearn 提供了交叉验证（cross-validation），需要注意的是，交叉验证中：k 值越大，越能减少偏差而导致的模型误差，但是训练集越大，会由于方差的增大而导致模型误差增大，并且导致模型训练的时间成本变高，因此 k 值的选取很关键，根据经验，k 值一般取 5 或者 10，本文选取 k=5，即本章采用 5 折交叉验证法，具体步骤如下：

（1）将所有训练数据集分成 5 份。

（2）不重复的每次取其中一份做测试集，用其他四份做训练集训练模型，之后利用设计好的代码计算该模型在测试集上的 MSE 值。

（3）将 5 次的 MSE 计算出的值取平均值得到最后的 MSE（均方误差）。

按照上述步骤对模型进行交叉验证，最后计算出 6 特征值的 MSE 结果如下：

Each folds MSE=0.02200272715372025

Each folds MSE=0.02610513891030476

Each folds MSE=0.05469206719249236

Each folds MSE=0.026911016509758242

Each folds MSE=0.035472038151442474

MSE=0.0330365975835436172

8 特征值的 MSE 结果如下：

Each folds MSE=0.03407843633539809

Each folds MSE=0.030986686853877995

Each folds MSE=0.03471291272357248

Each folds MSE=0.029893132630058383

Each folds MSE=0.02020189749480778

MSE=0.0299746132075429456

10 特征值的 MSE 结果如下：

Each folds MSE=0.028302799409375454

Each folds MSE=0.02351002858689573

Each folds MSE=0.026197814339067114

Each folds MSE=0.028480183447760925

Each folds MSE=0.028425579486082292

MSE=0.026983281053836303



图6 6、8、10特征值时的均方误差图(MSE)

图13折线图可以看出，均方误差的值介于0.02-0.06之间，趋近于0，随着输入的特征值的增大，均方误差减少，当特征值分别为6、8、10时，模型的均方误差都满足要求，模型的构建完成。

MSE（均方误差）函数用来检验模型预测值和真实值之间的偏差，可以用于评价数据的变化程度，MSE的值越小，表示模型描述实验数据的精确度高，与训练样本集的特征值及训练的次数有关，BP神经网络预测就是一个不断调试和学习过程，最后形成精确和稳定的预测模型。

3.5 本章小结

本章是基于BP神经网络的网球成绩预测模型构建部分，通过分析影响网球比赛的因素并且利用所能搜集到的数据进行特征选取，再对所收集到的数据进行预处理，设置好模型的参数后输入2019-2020年的大满贯赛事数据进行训练，对训练的结果进行MSE评价，模型构建完成。

4. 网球比赛成绩预测模型的实验结果与分析

本文共收集了 2019-2021 的 11 个赛事的数据及结果，这一章分别讨论了不同的比赛轮次、训练样本量和不同的特征值（6、8、10 个技术指标）对预测准确率的影响，并且利用模型对赛事的 16 强、8 强、4 强、冠亚军的结果进行了预测并对比分析。

在 BP 神经网络的预测中，成绩的预测只与选手本身有关，与对手在赛事中的发挥无关，后续的实验数据仅选取选手本人在赛事中的数据。以下预测均不考虑对阵表，也没有限定样本的竞争输出个数，因此在预测结果中可能会出现预测冠军、亚军和四强人数与实际不符的情况。

在 BP 神经网络训练中，为了量化训练目标，设定 1×5 的数组以对应最终战绩作为输出参数。具体表示为：

[1, 0, 0, 0, 0] 冠军
 [0, 1, 0, 0, 0] 亚军
 [0, 0, 1, 0, 0] 4 强
 [0, 0, 0, 1, 0] 8 强
 [0, 0, 0, 0, 1] 16 强

4.1 选取不同轮次对模型的准确率影响对比分析

本小节以 7 个赛事的数据作为模型的训练样本集，2021 年法国网球公开赛作为模型的测试集，分别用：（1）前三轮的 10 项技术指标，预测 16 强到冠军的比赛结果。（2）前四轮的 10 项技术指标，预测 8 强到冠军的比赛结果。将该项赛事预测的最终结果分别展示如下（输入指标和输出训练目标将具体在 4.2 中介绍）：

表 6 前三轮技术指标预测 16 强-冠军的最终结果

						预测排名	真实排名	是否正确
德约科维奇	1	0	0	0	0	1	1	是
西西帕斯	0	0	1	0	0	4	2	否
纳达尔	0	1	0	0	0	2	4	否
兹维列夫	1	0	0	0	0	1	4	否
贝雷蒂尼	0	0	0	1	0	8	8	是

施瓦茨曼	0	0	1	0	0	4	8	否
福基纳	0	0	0	0	1	16	8	否
梅德韦杰夫	0	0	0	1	0	8	8	是
穆塞蒂	0	0	0	0	1	16	16	是
费德勒	0	0	1	0	0	4	16	否
辛纳	0	0	0	0	1	16	16	是
勒纳德	0	0	0	0	1	16	16	是
锦织圭	0	0	0	0	1	16	16	是
德尔波尼斯	0	0	0	0	1	16	16	是
巴斯塔	0	0	1	0	0	4	16	否
C. 加林	0	0	0	0	1	16	16	是

表 7 前四轮技术指标预测 8 强-冠军的最终结果

	预测排名					真实排名	是否正确
德约科维奇	1	0	0	0	1	1	是
西西帕斯	0	1	0	0	2	2	是
纳达尔	0	0	0	1	8	4	是
兹维列夫	0	0	0	1	8	4	是
贝雷蒂尼	0	0	1	0	4	8	否
施瓦茨曼	0	0	0	1	8	8	是
福基纳	0	0	0	1	8	8	是
梅德韦杰夫	0	0	1	0	4	8	否

观察表格可以发现,采用前三轮数据预测全部结果时,最终成绩在 9-16 名的选手预测结果较为准确,8 位选手准确预测了 6 位,同时 8 强之后的结果不尽人意,错误地预测了 5 位选手,随着竞赛轮次的加深,选手的统计数据也会随着变动,用前三轮的技术指标预测较深轮次的结果不可靠,在增加了一轮的数据对 8 强到冠军的最终结果进行预测较为可靠,8 位选手的成绩准确预测了 6 位,也说明了 BP 神经网络预测具有一致性。

下文的预测也将采用同样的方法来进行,即采用前三轮的技术指标来预测第四轮的结果,采用前四轮的技术指标来预测 8 强到冠军的最终结果。将两次预测的结果合并到一个表格,表格展示如下:

表 8 2021 法国网球公开赛男子单打 16 强预测结果

						预测排名	真实排名	是否正确
德约科维奇	1	0	0	0	0	1	1	是
西西帕斯	0	1	0	0	0	2	2	是
纳达尔	0	0	0	1	0	8	4	是

兹维列夫	0	0	0	1	0	8	4	是
贝雷蒂尼	0	0	1	0	0	4	8	否
施瓦茨曼	0	0	0	1	0	8	8	是
福基纳	0	0	0	1	0	8	8	是
梅德韦杰夫	0	0	1	0	0	4	8	否
穆塞蒂	0	0	0	0	1	16	16	是
费德勒	0	0	1	0	0	4	16	否
辛纳	0	0	0	0	1	16	16	是
勒纳德	0	0	0	0	1	16	16	是
锦织圭	0	0	0	0	1	16	16	是
德尔波尼斯	0	0	0	0	1	16	16	是
巴斯塔	0	0	1	0	0	4	16	否
C. 加林	0	0	0	0	1	16	16	是

4.2 不同特征值对模型的准确率影响对比分析

以 7 个赛事的数据作为模型的训练样本集, 2021 年澳大利亚网球公开赛作为模型的测试集, 将该赛事的冠军德约科维奇和亚军梅德韦杰夫在本次赛事中选取的 1-4 轮的技术统计展示如下:

表 9 德约科维奇 1-4 轮技术统计

德约科维奇	第一轮	第二轮	第三轮	第四轮
一发成功率	80%	62%	66%	68%
一发得分率	83%	84%	75%	78%
二发得分率	83%	59%	50%	59%
接发球得分率	50%	38%	33%	35%
破发率	46%	36%	44%	27%
总得分率	64%	54%	50%	53%
ACES 球	7	26	15	10
双误	1	5	5	3
非受迫性失误	11	37	53	25
主动得分	41	56	52	41

表 10 梅德韦杰夫 1-4 轮技术统计

梅德韦杰夫	第一轮	第二轮	第三轮	第四轮
一发成功率	69%	67%	54%	65%

一发得分率	78%	84%	77%	78%
二发得分率	65%	48%	52%	60%
接发球得分率	44%	56%	41%	43%
破发率	56%	88%	39%	86%
总得分率	59%	64%	53%	57%
ACES 球	8	13	15	7
双误	2	3	4	2
非受迫性失误	13	29	44	15
主动得分	20	31	52	29

在 6 特征值的神经网络训练中,选取前三轮比赛的技术统计数据,预测第四轮的比赛成绩,则训练样本的输入指标和输出训练目标为:

德约科维奇 输入[0.80, 0.83, 0.83, 0.50, 0.46, 0.64]

[0.62, 0.84, 0.59, 0.38, 0.36, 0.54]

[0.66, 0.75, 0.50, 0.33, 0.44, 0.50]

输出[0, 0, 0, 1, 0]

梅德韦德夫 输入[0.69, 0.78, 0.65, 0.44, 0.56, 0.59]

[0.67, 0.84, 0.48, 0.56, 0.88, 0.64]

[0.54, 0.77, 0.52, 0.41, 0.39, 0.53]

输出[0, 0, 0, 1, 0]

在 8 特征值的神经网络训练中,选取前三轮比赛的技术统计数据,预测第四轮的比赛成绩,则训练样本的输入指标和输出训练目标为:

德约科维奇 输入[0.80, 0.83, 0.83, 0.50, 0.46, 0.64, 7, 1]

[0.62, 0.84, 0.59, 0.38, 0.36, 0.54, 26, 5]

[0.66, 0.75, 0.50, 0.33, 0.44, 0.50, 15, 5]

输出[0, 0, 0, 1, 0]

梅德韦德夫 输入[0.69, 0.78, 0.65, 0.44, 0.56, 0.59, 8, 2]

[0.67, 0.84, 0.48, 0.56, 0.88, 0.64, 13, 3]

[0.54, 0.77, 0.52, 0.41, 0.39, 0.53, 15, 4]

输出[0, 0, 0, 1, 0]

在 10 特征值的神经网络训练中, 选取前三轮比赛的技术统计数据, 预测第四轮的比赛成绩, 则训练样本的输入指标和输出训练目标为:

德约科维奇 输入[0.80, 0.83, 0.83, 0.50, 0.46, 0.64, 7, 1, 11, 41]
[0.62, 0.84, 0.59, 0.38, 0.36, 0.54, 26, 5, 37, 56]
[0.66, 0.75, 0.50, 0.33, 0.44, 0.50, 15, 5, 53, 52]
输出[0, 0, 0, 1, 0]

梅德韦德夫 输入[0.69, 0.78, 0.65, 0.44, 0.56, 0.59, 8, 2, 13, 20]
[0.67, 0.84, 0.48, 0.56, 0.88, 0.64, 13, 3, 29, 31]
[0.54, 0.77, 0.52, 0.41, 0.39, 0.53, 15, 4, 44, 52]
输出[0, 0, 0, 1, 0]

其他选手的数据依次输入。

同理, 我们以 10 特征值的神经网络训练为例, 选取前四轮比赛的技术统计数据, 预测 8 强到冠军的最终比赛结果, 则训练样本的输入指标和输出训练目标为:

德约科维奇 输入[0.80, 0.83, 0.83, 0.50, 0.46, 0.64, 7, 1, 11, 41]
[0.62, 0.84, 0.59, 0.38, 0.36, 0.54, 26, 5, 37, 56]
[0.66, 0.75, 0.50, 0.33, 0.44, 0.50, 15, 5, 53, 52]
[0.68, 0.78, 0.59, 0.35, 0.27, 0.53, 10, 3, 25, 41]
输出[1, 0, 0, 0, 0]

梅德韦德夫 输入[0.69, 0.78, 0.65, 0.44, 0.56, 0.59, 8, 2, 13, 20]
[0.67, 0.84, 0.48, 0.56, 0.88, 0.64, 13, 3, 29, 31]
[0.54, 0.77, 0.52, 0.41, 0.39, 0.53, 15, 4, 44, 52]
[0.65, 0.78, 0.60, 0.43, 0.86, 0.57, 7, 2, 15, 29]
输出[0, 1, 0, 0, 0]

其他选手的数据依次输入。

将测试集的数据输入模型后, 最后的结果为:

表 11 6 特征值预测结果

						预测排名	真实排名	是否正确
德约科维奇	0	1	0	0	0	2	1	否
梅德韦杰夫	0	0	0	1	0	8	2	否

卡拉特瑟	0	0	0	0	1	16	4	否
西西帕斯	0	0	1	0	0	4	4	是
A. 兹维列夫	0	1	0	0	0	2	8	否
迪米托夫	0	0	0	1	0	8	8	是
A. 卢布列夫	0	0	1	0	0	4	8	否
纳达尔	0	0	0	1	0	8	8	是
拉奥尼奇	0	0	0	0	1	16	16	是
拉约维卡	0	0	0	0	1	16	16	是
D. 蒂姆	0	0	1	0	0	4	16	否
卡斯珀·鲁德	0	0	0	0	1	16	16	是
巴里蒂尼	0	0	0	0	1	16	16	是
F. 奥格	0	0	0	0	1	16	16	是
麦当纳德	0	0	0	1	0	8	16	否
F. 弗格尼尼	0	0	0	0	1	16	16	是

在 6 特征值的预测结果中, 整体预测的准确率较低, 仅为 56.25%, 其中, 16 进 8 的预测结果中, 错了 2 人, 1 人错误的预测为 8 强, 另 1 人错误的预测为 4 强, 在 8 强之后的预测中, 仅仅预测正确了 3 个。说明 6 特征值的预测结果不理想, 所选的特征值数量不足以满足模型训练的要求, 该模型不具有实际的参考价值。

表 12 8 特征值预测结果

						预测排名	真实排名	是否正确
德约科维奇	1	0	0	0	0	1	1	是
梅德韦杰夫	0	0	1	0	0	4	2	否
卡拉特瑟	0	0	1	0	0	4	4	是
西西帕斯	0	1	0	0	0	2	4	否
A. 兹维列夫	0	0	0	1	0	8	8	是
迪米托夫	0	0	0	1	0	8	8	是
A. 卢布列夫	0	0	1	0	0	4	8	否
纳达尔	0	0	0	1	0	8	8	是

拉奥尼奇	0	0	0	0	1	16	16	是
拉约维卡	0	0	0	0	1	16	16	是
D. 蒂姆	0	0	1	0	0	4	16	否
卡斯珀·鲁德	0	0	0	0	1	16	16	是
巴里蒂尼	0	0	0	0	1	16	16	是
F. 奥格	0	0	0	0	1	16	16	是
麦当纳德	0	0	0	1	0	8	16	否
F. 弗格尼尼	0	0	0	0	1	16	16	是

8 特征值中, 添加了 ACE 球和双误这 2 个特征值, 这 2 个特征值都是属于发球技术中的特征。从预测结果来看, 整体预测准确率为 68.75%, 准确的预测了 11 个排名, 与 6 特征值预测结果相比, 准确率得到了大幅度的提升, 总共提升了 12.50%, 说明 ACE 球和双误这 2 项特征值在成绩预测方面的相关性较强。也说明在职业男子网球比赛中, 发球这一项技术在比赛中起到了至关重要的作用。

在所选的 8 个指标中, 与发球有关的指标为 5 个, 发球在网球比赛中占据十分重要的地位。网球发球是惟一由自己掌握而不受对方控制的击球方法^[48]。有了好的发球技术, 能够让球员在发球局中更好的实施战术, 在发球之后的进攻和衔接更加得心应手, 在发球的同时构思清楚自己的目的和意图, 将速度、力量和旋转的变化相结合, 并且能够根据场上的具体情况来发不同类型的球, 给对手的接发球制造困难, 给自己创造优势, 进一步致胜对手。

在大满贯这种高级别的赛事中, 每位运动员的技战术水平都是非常全面的, 如果发球不够稳定或者攻击性不高时, 很容易就被对手抓住机会进行压制, 给自己的发球局制造很大的压力, 比赛结果也将不尽人意。在当今的男子职业网坛中, ACE 球的出现较为频繁, 原因是 ACE 球带来分数的同时也给自己带来状态上的提升, 给对手的心理造成压力, 令对手处于下风, ACE 球往往能够成为打败对手的杀手锏。在比赛的过程焦灼的时候, 发球一方的心理会出现波动, 领先时想要一击制胜, 落后时着急赶上, 容易出现双误的情况。

所以通过上述分析, 在以后的网球比赛中, 运动员和教练可增加对发球技战术的研究, 来更大程度地获取比赛的胜利。

表 13 10 特征值预测结果

						预测排名	真实排名	是否正确
德约科维奇	1	0	0	0	0	1	1	是
梅德韦杰夫	0	1	0	0	0	2	2	是
阿斯兰-卡拉特瑟	0	0	0	1	0	8	4	否
西西帕斯	0	0	1	0	0	4	4	是
A. 兹维列夫	0	0	1	0	0	4	8	否
迪米托夫	0	0	0	1	0	8	8	是
A. 卢布列夫	0	0	0	1	0	8	8	是
纳达尔	0	0	0	1	0	8	8	是
拉奥尼奇	0	0	0	0	1	16	16	是
拉约维卡	0	0	0	0	1	16	16	是
D. 蒂姆	0	0	1	0	0	4	16	否
卡斯珀·鲁德	0	0	0	0	1	16	16	是
巴里蒂尼	0	0	0	0	1	16	16	是
F. 奥格	0	0	0	0	1	16	16	是
麦当纳德	0	0	0	0	1	16	16	是
F. 弗格尼尼	0	0	0	0	1	16	16	是

在 10 特征值的预测中, 添加了主动得分和非受迫性失误这 2 个特征, 从结果来看, 这 2 项特征对比赛成绩的预测结果也有一定的影响, 与 8 特征值预测结果相比, 多预测正确 2 个, 整体预测准确率为 81.25%, 提高了 12.50%。主动得分和非受迫性失误在对预测结果产生影响的同时, 也从侧面说明了这 2 项技术数据在比赛中的重要性, 在当今职业网球比赛中, 主动得分是反映一名网球运动员攻击性强弱的重要指标, 而非受迫性失误是反映运动员技术稳定性的一个重要指标, 非受迫性失误的多少是决定比赛胜负的关键因素之一^[49]。努力减少非受迫性失误是运动员在比赛中要时刻记住的宗旨^[50]。在提高主动得分能力的同时, 降低非受迫性失误, 是世界顶级选手应该具备的素质, 才能够高水平的博弈中取得胜利。

上述 3 个实验结果对比可以看出，所选 10 个特征都对比赛成绩有一定程度的影响，随着特征值的增加，成绩预测模型的准确率也越来越高，模型具有一定的参考价值，ACE 球、双误、主动得分和非受迫性失误这 4 个关键参数对网络预测结果的变化较为敏感。

4.3 不同的训练样本量对模型预测准确率的影响与分析

网络的可靠性与选取的训练样本数和样本的代表性均有很大关系，因此本小节选取不同的训练样本量对模型预测准确率的影响进行分析。

通过实验可以发现，我们选取特征值为 10 时，模型的准确率最高。因此，在特征值一定时（特征值为 10），我们分别选取了以下 3 种情况探讨不同样本量对模型预测准确率的影响：

（1）2019 年的 4 个大满贯赛事作为训练样本集，2021 年美国网球公开赛作为测试集。

（2）2019-2020 年的其中 6 个大满贯赛事作为训练样本集，2021 年美国网球公开赛作为测试集。

（3）2019-2020 年的 7 个大满贯赛事作为训练样本集（2020 温布尔登网球锦标赛因疫情停止举办），2021 年美国网球公开赛作为测试集。

表 14 4 个赛事数据为训练样本量输入时，预测 2021 年美国网球公开赛结果

						预测排名	真实排名	是否正确
德约科维奇	1	0	0	0	0	1	1	是
梅德韦德夫	0	0	0	1	0	8	2	否
A. 兹维列夫	0	0	1	0	0	4	4	是
F. 奥格	0	0	1	0	0	4	4	是
贝雷蒂尼	0	0	0	1	0	8	8	是
L. 哈里斯	0	0	1	0	0	4	8	否
阿尔卡拉斯	0	0	1	0	0	4	8	否
博迪赞德	0	0	0	1	0	8	8	是
布鲁克斯比	0	0	0	0	1	16	16	是
O. 奥特	0	0	0	0	1	16	16	是
辛纳	0	0	0	1	0	8	16	否

奥佩尔卡	0	0	0	1	0	8	16	否
蒂亚福	0	0	0	0	1	16	16	是
高若尼兹克	0	0	0	0	1	16	16	是
施瓦茨曼	0	0	1	0	0	4	16	否
D. 埃文斯	0	0	0	0	1	16	16	是

在 16 进 8 的预测结果中, 8 人止步, 预测对了 5 人, 2 人预测结果为 8 强, 一人预测结果为 4 强, 预测准确率为 62.50%。在 8 进 4 的比赛中, 4 人止步, 预测对了 2 人, 另外 2 人预测结果皆为 4 强, 预测准确率为 50.00%。而冠亚军和半决赛错误预测了 1 人。从整体结果来看, 效果并不理想, 整体预测准确率仅为 62.50%。

当样本量为 4 个赛事时, 所预测出的结果并不理想, 我们选取了 2019 和 2020 年的 6 个大满贯赛事的数据作为样本输入, 预测 2021 年美国网球公开赛的比赛结果:

表 15 6 个赛事数据为训练样本量输入时, 2021 年美国网球公开赛的预测结果

						预测排名	真实排名	是否正确
德约科维奇	1	0	0	0	0	1	1	是
梅德韦德夫	0	1	0	0	0	2	2	是
A. 兹维列夫	0	0	0	1	0	8	4	否
F. 奥格	0	0	0	1	0	8	4	否
贝雷蒂尼	0	0	0	1	0	8	8	是
L. 哈里斯	0	0	0	1	0	8	8	是
阿尔卡拉斯	0	0	1	0	0	4	8	否
博迪赞德	0	0	0	1	0	8	8	是
布鲁克斯比	0	0	0	0	1	16	16	是
O. 奥特	0	0	0	0	1	16	16	是
辛纳	0	0	0	0	1	8	16	是
奥佩尔卡	0	0	0	0	1	16	16	是
蒂亚福	0	0	0	0	1	16	16	是
高若尼兹克	0	0	0	0	1	16	16	是
施瓦茨曼	0	0	0	0	1	16	16	是
D. 埃文斯	0	0	0	0	1	16	16	是

在 16 进 8 的预测结果中, 预测对了 7 人, 1 人预测结果为 8 强, 预测准确率为 87.50%。在 8 进 4 的比赛中, 预测对了 3 人, 预测准确率较高, 达到了 75.00%。而冠亚军和半决赛则错误预测 2 人。从整体来看, 预测准确率达到 81.25%, 与上个模型的结果对比, 预测准确率提高了 18.75%, 提升幅度较大, 预测的准确率较理想。

继续增加样本量, 选取 2019-2020 的 7 个大满贯赛事的数据作为样本输入, 预测 2021 年美国网球公开赛的比赛结果。

表 16 7 个赛事数据为训练样本量输入时, 2021 年美国网球公开赛的预测结果

						预测排名	真实排名	是否正确
德约科维奇	1	0	0	0	0	1	1	是
梅德韦德夫	0	0	0	1	0	8	2	否
A. 兹维列夫	0	0	1	0	0	4	4	是
F. 奥格	0	1	0	0	0	2	4	否
贝雷蒂尼	0	0	0	1	0	8	8	是
L. 哈里斯	0	0	0	1	0	8	8	是
阿尔卡拉斯	0	0	1	0	0	4	8	否
博迪赞德	0	0	0	1	0	8	8	是
布鲁克斯比	0	0	0	0	1	16	16	是
O. 奥特	0	0	0	0	1	16	16	是
辛纳	0	0	0	0	1	16	16	是
奥佩尔卡	0	0	0	0	1	16	16	是
蒂亚福	0	0	0	0	1	16	16	是
高若尼兹克	0	0	0	0	1	16	16	是
施瓦茨曼	0	0	0	0	1	16	16	是
D. 埃文斯	0	0	0	0	1	16	16	是

在 16 进 8 的预测结果中, 预测对了 7 人, 1 人错误的预测为 4 强, 预测准确率为 87.50%。在 8 进 4 的比赛中, 预测对了 3 人, 预测准确率为 75.00%。在冠亚军和半决赛的预测中, 错误预测了 2 人, 将梅德韦德夫预测为了 8 强。在 16 个成绩预测结果中, 正确了 13 个, 正确率为 81.25%, 与上个模型相比, 准确率相同, 模

型的正确率较为理想。

通过上面的模型对比可以得出：当特征值为 10（固定值）时，训练模型的样本量越大，模型的准确率越高。当样本量为 4 个赛事数据时，准确率仅为 62.50%，当样本量分别为 6 个和 7 个赛事的数据时，准确率达到 81.25%。说明神经网络的预测结果还是需要充足的训练样本量来保障其正确率，但是当数据量达到一定程度时，模型已经完成训练并且权重系数调整完成，继续添加数据并不会对模型的准确率有提高。观察准确率较高的 2 个模型可以发现预测结果错误的地方并不一致，6 赛事数据为训练样本集时，模型是依据 2019–2020 年所选的 6 个大满贯赛事数据所建立，模型之间的权重依据这 6 个赛事的训练数据所建立。7 赛事数据为训练样本集时，模型是依据 2019–2020 年的 7 个大满贯赛事数据所建立，模型之间的权重依据 2019–2020 所有的大满贯赛事数据，两个模型的权重有所变化，出现了不同的错误结果。

观察模型可以发现，模型对 16 进 8 的预测准确率较高，而对更深轮次的比赛结果预测效果相对差一些，16 进 8 时，对战双方的实力还是有一定差距，一方面是 BP 神经网络对数据的敏感性较强，模型调整出的权重比例准确，另一方面，进入更深轮次的选手在数据方面要明显优于排名靠后的选手，而随着轮次的加深，选手的实力和水平大相径庭，BP 神经网络要判断出选手的排名也会变得相对困难，出现错误的情况也无可厚非。

4.4 不同场地类型的模型准确率对比分析

上述实验已经对特征值、训练样本量进行探讨，且对 2021 年澳大利亚网球公开赛、2021 年美国网球公开赛的结果进行预测，为了验证 BP 神经网络模型在任意的场地类型中均适用，将对 2021 年法国网球公开赛和 2021 年温布尔登网球公开赛的结果进行预测并对比分析，为了保证模型的特征值和训练样本量一致，特征值取 10，模型训练样本量为 7（除了要进行测试的 4 个 2021 年的大满贯赛事样本，其余样本均进行模型的训练），结果如下：

表 17 2021 年温布尔登网球公开赛预测结果

	预测排名					真实排名	是否正确预测
德约科维奇	1	0	0	0	0	1	是

贝雷蒂尼	0	1	0	0	0	2	2	是
沙波瓦洛夫	0	0	0	0	1	4	4	是
胡尔卡奇	0	0	0	1	0	8	4	否
富切维奇	0	0	0	1	0	8	8	是
卡恰诺夫	0	0	0	0	1	16	8	否
F. 奥格	0	0	0	1	0	8	8	是
费德勒	0	0	0	1	0	8	8	是
C. 加林	0	0	0	0	1	16	16	是
A. 卢布列夫	0	0	0	0	1	16	16	是
科达	0	0	0	0	1	16	16	是
巴蒂斯塔	0	0	0	0	1	16	16	是
伊瓦什卡	0	0	0	0	1	16	16	是
A. 兹维列夫	0	0	1	0	0	4	16	否
L. 索内阁	0	0	0	0	0	16	16	是
D. 梅德韦德夫	0	0	0	1	0	8	16	否

从表 15 的结果可以看出, 预测准确率为 75.00%, 综合其他 3 个赛事的预测结果, 整体预测准确率为 78.13%, 说明 BP 神经网络模型在任意的场地类型上均适用。因此, 使用 BP 神经网络模型进行网球赛事成绩的预测是可行的。神经网络需要不断的调试, 了解自己的预测结果是否准确, 才能够快速进行调整, 并使结果越来越准确, 以上是神经网络学习的过程。在不同的场地类型中, 产生变化的部分是球速和旋转, 球速快旋转强烈得分就越多, 例如: 发球的球速快, 旋转强烈, 能够带来的主动得分和 ACE 球数量就越多, 也能够为球员的战术布置打下基础, 球速和旋转产生的结果可以在技战术数据中体现出来, 且都能够在不同程度上影响着比赛的最终结果, 因此, 神经网络可以对不同场地类型的网球赛事进行预测。

4.5 取技术平均值后模型的准确率对比分析

笔者在对模型不断进行训练的同时, 观察数据并且思考后发现: 在男子网坛中, 能够取得好成绩的往往都是发挥较为稳定的, 技术全面的选手, 在较深轮次的比赛中, 往往很少有黑马能够异军突起, 从数据的角度来看, 不同的轮次中, 黑马选手

的技术统计数据往往起伏较大,而德约科维奇、纳达尔和西西帕斯等选手的技术统计数据往往较为稳定。在模型训练完成后,笔者通过对测试集数据取技术平均值的方式进行实验:

以7个赛事的数据作为模型的训练样本集,2021年法国网球公开赛作为模型的测试集,将该赛事的冠军德约科维奇和亚军西西帕斯在本次赛事中选取的1-4轮的技术统计如下:

表 18 德约科维奇 1-4 轮技术统计

德约科维奇	第一轮	第二轮	第三轮	第四轮
一发成功率	67%	68%	80%	70%
一发得分率	86%	75%	83%	77%
二发得分率	54%	57%	83%	65%
接发球得分率	47%	45%	50%	39%
破发率	45%	63%	46%	44%
总得分率	59%	58%	64%	55%
ACES 球	4	10	7	10
双误	2	3	1	1
非受迫性失误	21	22	11	19
主动得分	33	32	41	44

表 19 西西帕斯 1-4 轮技术统计

西西帕斯	第一轮	第二轮	第三轮	第四轮
一发成功率	64%	70%	60%	66%
一发得分率	75%	64%	86%	75%
二发得分率	61%	54%	60%	59%
接发球得分率	46%	51%	32%	43%
破发率	83%	44%	43%	42%

总得分率	58%	55%	55%	57%
ACES 球	5	8	8	6
双误	3	3	5	0
非受迫性失误	25	30	19	17
主动得分	31	45	40	41

在 10 特征值的神经网络训练中,选取前三轮比赛的技术统计平均值预测第四轮比赛成绩,则训练样本的输入指标和输出训练目标为:

德约科维奇 输入[0.72, 0.81, 0.65, 0.47, 0.51, 0.60, 7, 2, 18, 24]

输出[0, 0, 0, 1, 0]

西西帕斯 输入[0.65, 0.75, 0.58, 0.43, 0.57, 0.56, 7, 3.6, 24, 38]

输出[0, 0, 0, 1, 0]

其他选手的数据依次输入。

同样选取前四轮比赛的技术统计平均值,则训练样本的输入指标和输出训练目标为:

德约科维奇 输入[0.71, 0.80, 0.64, 0.45, 0.50, 0.59, 7.7, 1.7, 18.3, 37.5]

输出[1, 0, 0, 0, 0]

西西帕斯 输入[0.65, 0.75, 0.58, 0.43, 0.53, 0.56, 6.7, 2.8, 22.8, 39.3]

输出[0, 1, 0, 0, 0]

其他选手的数据依次输入。

分别对 2021 年 4 个大满贯赛事进行实验,结果为:

表 20 2021 澳大利亚网球公开赛预测结果

						预测排名	真实排名	是否正确
德约科维奇	1	0	0	0	0	1	1	是
梅德韦杰夫	0	1	0	0	0	2	2	是
卡拉特瑟	0	1	0	0	0	2	4	否
西西帕斯	0	0	1	0	0	4	4	是

A. 兹维列夫	0	0	0	1	0	8	8	是
迪米托夫	0	0	0	1	0	8	8	是
A. 卢布列夫	0	0	0	1	0	8	8	是
纳达尔	0	0	0	1	0	8	8	是
拉奥尼奇	0	0	0	0	1	16	16	是
拉约维卡	0	0	0	0	1	16	16	是
D. 蒂姆	0	0	1	0	0	4	16	否
卡斯珀·鲁德	0	0	0	0	1	16	16	是
巴里蒂尼	0	0	1	0	0	4	16	否
F. 奥格	0	0	0	0	1	16	16	是
麦当纳德	0	0	0	0	1	16	16	是
F. 弗格尼尼	0	0	0	0	1	16	16	是

表 21 2021 法国网球公开赛预测结果

						预测排名	真实排名	是否正确
德约科维奇	1	0	0	0	0	1	1	是
西西帕斯	0	1	0	0	0	2	2	是
纳达尔	1	0	0	0	0	1	4	否
A. 兹维列夫	0	0	0	1	0	8	4	否
贝雷蒂尼	0	0	0	1	0	8	8	是
施瓦茨曼	0	0	0	1	0	8	8	是
福基纳	0	0	0	1	0	8	8	是
梅德韦杰夫	0	0	0	1	0	8	8	是
穆塞蒂	0	0	0	0	1	16	16	是
费德勒	0	0	1	0	0	4	16	否
辛纳	0	0	0	0	1	16	16	是
勒纳德	0	0	0	0	1	16	16	是
锦织圭	0	0	0	0	1	16	16	是

德尔波尼斯	0	0	0	0	1	16	16	是
巴斯塔	0	0	0	0	1	16	16	是
C. 加林	0	0	0	0	1	16	16	是

表 22 2021 温布尔登网球公开赛预测结果

						预测排名	真实排名	是否正确
德约科维奇	1	0	0	0	0	1	1	是
贝雷蒂尼	0	0	1	0	0	4	2	否
沙波瓦洛夫	0	0	0	0	1	16	4	否
胡尔卡奇	0	0	1	0	0	4	4	是
富切维奇	0	0	0	1	0	8	8	是
卡恰诺夫	0	0	0	1	0	8	8	是
F. 奥格	0	0	0	1	0	8	8	是
费德勒	0	0	0	1	0	8	8	是
C. 加林	0	0	0	0	1	16	16	是
A. 卢布列夫	0	0	0	0	1	16	16	是
科达	0	0	0	0	1	16	16	是
巴蒂斯塔	0	0	0	0	1	16	16	是
伊瓦什卡	0	0	0	0	1	16	16	是
A. 兹维列夫	0	0	1	0	0	4	16	否
L. 索内阁	0	0	0	0	1	16	16	是
D. 梅德韦德夫	0	0	0	0	1	16	16	是

表 23 2021 美国网球公开赛预测结果

						预测排名	真实排名	是否正确
梅德韦杰夫	1	0	0	0	0	1	1	是
德约科维奇	0	1	0	0	0	2	2	是

A. 兹维列夫	0	0	1	0	0	4	4	是
F. 奥格	0	0	1	0	0	4	4	是
贝雷蒂尼	0	1	0	0	0	2	8	否
L. 哈里斯	0	0	0	1	0	8	8	是
阿尔卡拉斯	0	0	0	0	1	16	8	否
博迪赞德	0	0	0	1	0	8	8	是
布鲁克斯比	0	0	0	0	1	16	16	是
O. 奥特	0	0	0	0	1	16	16	是
辛纳	0	0	0	0	1	16	16	是
奥佩尔卡	0	0	0	0	1	16	16	是
蒂亚福	0	0	0	0	1	16	16	是
高若尼兹克	0	0	0	0	1	16	16	是
施瓦茨曼	0	0	0	1	0	8	16	否
D. 埃文斯	0	0	0	0	1	16	16	是

随着比赛轮次的加深,一些无法量化的指标则逐渐潜移默化,体能的深度,训练的强度还有一些心理因素无法忽略。预测精度的提升,想要多预测正确一个选手的排名都不容易。在对选手们技术统计数据取平均值后,模型的预测准确率为81.25%,相较于未取平均值的模型,准确率提升了3.125%,这也说明了取平均值后的技术数据更能代表该选手在该项赛事中状态和水平的发挥。

4.6 2021 年四大满贯冠亚军预测结果对比

在体育运动项目中,最受观众和球迷关注的都是决赛的精彩对决,网球项目也不例外。我们将双方的技术参数优化后进行结果的对比。

试算时首先采用2019-2020年大满贯决赛冠亚军对阵双方前几轮的技术参数平均值作为输入参数,输出结果为冠军及亚军。在实际网络训练中,由于样本数过少(仅7次决赛样本),神经网络训练不收敛,无法得到理想的预测结果。

考虑冠亚军比赛仅仅在于预测对阵选手双方的胜负关系,故增加网络的训练样本数,将16强对阵、8强对阵以及4强对阵胜负均作为输入样本。输入样本设置为:

A vs B

输入：对阵双方前几轮的技术参数差值

输出：1——代表 A 获胜

0——代表 B 获胜

如 2019 年澳网决赛对阵双方，德约科维奇对阵纳达尔，表中所列为两位选手前几轮技术参数平均值及相减后得到的技术参数差值

表 24 澳网决赛对阵双方技术参数差值

德约科维奇	纳达尔	技术参数差值
0.73	0.67	0.06
0.77	0.74	0.03
0.67	0.59	0.08
0.60	0.46	0.14
0.67	0.50	0.17
1.00	2.00	-1.00
2.80	3.46	-0.66
2.00	6.00	-4.00
30.65	28.12	2.53
40.30	42.55	-2.25

输入样本为 [0.06 , 0.03 , 0.08, 0.14 , 0.17 , -1.00 , -0.66 , -4.00, 2.53, -2.25]

输出结果为 1 （德约科维奇获胜）

建立的总样本数共 63 场比赛的胜负关系用于网络训练，实际预测的 2021 年四个大满贯赛事的结果如下：

表 25 2021 年四个大满贯赛事预测结果

赛事	对阵双方	预测胜方	预测结果
2021 澳网	德约科维奇 vs 梅德韦杰夫	德约科维奇	正确
2021 美网	德约科维奇 vs 梅德韦杰夫	梅德韦杰夫	错误
2021 温网	德约科维奇 vs 贝雷蒂尼	德约科维奇	正确
2021 法网	德约科维奇 vs 西西帕斯	德约科维奇	正确

观察上表可以得出，在优化技术参数后对 4 场 2021 年大满贯决赛进行胜负的预测，4 个对了 3 个，我们对神经网络判别错误的这一场（2021 年法网）进行具体的分析，在这场男单决赛之前，德约科维奇已经取得了 27 胜 0 负的骄人战绩，观察录像可以赛后技术数据可以发现，我们能够找到德约科维奇被梅德韦杰夫击败的答案。赛后技术数据统计如下：

表 26 2021 年美国网球公开赛决赛技术统计数据

	德约科维奇	梅德韦杰夫
一发成功率	54%	57%
一发得分率	80%	80%
二发得分率	40%	58%
接发球得分率	29%	38%
破发率	17%	50%
总得分率	46%	54%
ACES 球	6	16
双误	3	9
非受迫性失误	38	31
主动得分	27	38

梅德韦杰夫以 3 个 6-4 成功首夺大满贯，狙击德约科维奇夺取 21 个大满贯的进程，德约科维奇年度全满贯梦碎。纵观整场比赛，德约科维奇在比赛中的具体表现是：体力耗竭、跑动慢、击球力量小，也没有发挥出世界顶尖的接发球技术，所以无法通过改变比赛节奏来打乱梅德韦杰夫的战术意图。从数据上就可以看出：梅德韦杰夫“拼命三郎”式的发球，取得了奇效，双误数量比德约科维奇要多出 6 个，同时也在 ACE 球方面领先了德约科维奇 10 个；更关键的是，底线相持中，德约的正反手频频失误，在主动得分低于梅德韦杰夫的情况下，非受迫性失误还高于梅德韦杰夫，处于下风。从技术角度来看，德约科维奇几项突出的技术环节都受到了压制，有力使不出。在这场比赛之前，德约科维奇的表现较好，外界也都看好德约科维奇，将本场比赛预测错误也无可厚非，也有一些黑马和意外情况的发生，这也是竞技体育的魅力之一。

综合以上实验结果，说明了神经网络模型对网球比赛成绩预测具有较高的预测精度，该模型准确有效，能够为运动员和教练员在实际的训练和比赛中提供一定的支持。随着特征值和训练样本量的增加，构建的 BP 神经网络模型可以很好的反映

出所选择的技术指标与四大满贯成绩之间的关系，所搜集的数据都对模型的训练及成绩预测模型准确率的提升有着较大作用。通过调整技术参数差值，能够使 BP 神经网络做出正确的预测，所以本文研究的 BP 神经网络模型的优势就体现出来了，BP 神经网络模型能利用大量的数据来对模型进行构建和拟合。

4.7 本章小结

本章主要是网球成绩预测模型的实验验证与对比分析，分别对特征值、样本量、场地类型和技术平均值等因素进行了实验与探讨，最后，对 2021 年大满贯决赛进行胜负预测，评价本文的成绩预测模型。实验结果说明将 BP 神经网络理论运用到网球比赛的成绩预测中是可行的。

5.结论与展望

5.1 结论

1. 基于 BP 神经网络原理, 运用 Matlab 软件进行编程, 构建的网球四大满贯的成绩预测模型是可行的。

2. 在 BP 神经网络模型中, 收集的有效数据样本量越少, 模型准确率较低, 随着样本量的增加, 模型准确率会上升, 当训练样本量达到一定程度时, 模型训练完成, 预测准确率会维持在较高的水平。

3. 在大满贯赛事中, ACE 球、双误、主动得分和非受迫性失误这 4 个关键参数对网络预测结果的变化较为敏感。

4. 在特征值为 10, 训练样本量为 7 的条件下, 整体预测精度可达 78.125%。

5. 在 BP 神经网络模型中, 不同的特征值对模型产生的准确率不同, 采用的特征值越多, 模型预测的正确率就越高。

5.2 未来展望

目前, 网球赛事的相关预测研究远远落后于足球、排球等项目的研究。未来可以增加本研究中没有涉及的技术指标, 如发球速度、网前得分率等, 探讨这些指标对模型准确率的影响。在此基础上, 结合不同的球员技术特点给不同的技术指标赋予不同的权重值, 再做进一步的分析。

经过不断的研究和再往后发展, 逐步实现了能够对运动员的跑动, 发球的落点旋转等进行一个预测, 为未来的网球赛事甚至其他体育赛事做一个展望, 甚至有完全预测比赛成绩的可能。

笔者仅用了 2019-2021 三年的四大满贯赛事数据作为支撑, 随着数据越来越详尽, 之后的研究可沿用 5-10 年的数据来进行预测并进行分析, 观察是否有更好的效果。

经验证此预测原理和方法具体可行。合理的推论同样可以应用在棒球、羽毛球等赛事的成绩预测, 但是需要根据该项赛事的对战数据进行合理分析, 决定 BP 神经网络的输入特征及后续的验证。另外, 随着人民群众对竞技体育的参与度提高,

以及体育行业的迅速发展，网球比赛的预测有广阔的发展前景和商业价值。

附录

成绩预测代码

```
% BP Éñ¾ÍøÂçÓÃÓÚÍøÇòÔ²â
```

```
% 8²ÎÊýÊäËë
```

```
% Ô²â8Ç¿
```

```
clc
```

```
clear
```

```
close all
```

```
%Ô²â8Ç¿
```

```
TrainTarget1=load('target8.txt');
```

```
nnn=8; %8²ÎÊýÑµÁ·
```

```
num_sample=8;
```

```
num_round=3;
```

```
%% ÓÃËÄÂÖÆ½¾üÖµ
```

```
% rawData = load('2019winopen_8.txt');
```

```
% rawData=rawData';
```

```
% rawData1=rawData(1,:)+rawData(2,:)+rawData(3,:)+rawData(4,:);
```

```
% rawData1=rawData1'/4;
```

```
% rawData1=rawData1(1:nnn*8);
```

```
%
```

```
%
```

```
% rawData = load('2019auopen_8.txt');
```

```
% rawData=rawData';
```

```
% rawData2=rawData(1,:)+rawData(2,:)+rawData(3,:)+rawData(4,:);
```

```
% rawData2=rawData2'/4;
```

```

% rawData2=rawData2(1:nnn*8);

%

% rawData = load('2019usopen_8.txt');

% rawData=rawData';

% rawData3=rawData(1,:)+rawData(2,:)+rawData(3,:)+rawData(4,:);

% rawData3=rawData3'/4;

% rawData3=rawData3(1:nnn*8);

%

% rawData = load('2019fropen_8.txt');

% rawData=rawData';

% rawData4=rawData(1,:)+rawData(2,:)+rawData(3,:)+rawData(4,:);

% rawData4=rawData4'/4;

% rawData4=rawData4(1:nnn*8);

%

% rawData=[rawData1;rawData2;rawData4];

% TrainTarget=[TrainTarget1;TrainTarget1;TrainTarget1];

%% ÓÃ×î½üÒ»ÄÖ¼¼ÊðÖ,±ê

% rawData = load('2019winopen_8.txt');

% rawData=rawData';

% rawData1=rawData(4,:);

% rawData1=rawData1(1:nnn*8);

%

% rawData = load('2019auopen_8.txt');

% rawData=rawData';

% rawData2=rawData(4,:);

```

```

% rawData2=rawData2(1:nnn*8);

%

% rawData = load('2019usopen_8.txt');

% rawData=rawData';

% rawData3=rawData(4,:);

% rawData3=rawData3(1:nnn*8);

%

% rawData = load('2019fropen_8.txt');

% rawData=rawData';

% rawData4=rawData(4,:);

% rawData4=rawData4(1:nnn*8);

%

% rawData=[rawData1,rawData2,rawData4];

% TrainTarget=[TrainTarget1;TrainTarget1;TrainTarget1];

%% 19 20Ï²â21 Í¬,öÈÛÊÂ

rawData = load('2019auopen_8.txt');

rawData=rawData';

rawData_temp=zeros(1,16*nnn);

for i=1:num_round

    rawData_temp=rawData_temp+rawData(i,:);

end

rawData2=rawData_temp'/num_round;

rawData2=rawData2(1:nnn*8);

rawData = load('2020auopen_8.txt');

rawData=rawData';

```

```
rawData_temp=zeros(1,16*nnn);  
for i=1:num_round  
    rawData_temp=rawData_temp+rawData(i,:);  
end  
rawData3=rawData_temp'/num_round;  
rawData3=rawData3(1:nnn*8);  
  
rawData = load('2021auopen_8.txt');  
rawData=rawData';  
  
rawData_temp=zeros(1,16*nnn);  
for i=1:num_round  
    rawData_temp=rawData_temp+rawData(i,:);  
end  
rawData4=rawData_temp'/num_round;  
rawData4=rawData4(1:nnn*8);  
  
rawData=[rawData2;rawData3];  
TrainTarget=[TrainTarget1;TrainTarget1];  
  
% rawData=[rawData1;rawData2;rawData3;rawData4];  
% TrainTarget=[TrainTarget1;TrainTarget1;TrainTarget1;TrainTarget1];  
  
nTrainData=length(rawData(:,1));  
  
nn=length(rawData);
```



```

nTrainData=nn/nnn;

TrainInput=zeros(nTrainData,nnn);
% TrainTarget=zeros(16,5);
% TrainTarget=load('target.txt');

for i=1:nTrainData
    TrainInput(i,:)=rawData(1+(i-1)*nnn:i*nnn);
end

xn_train=TrainInput';
dn_train=TrainTarget';

% xn_train = n1;          %  $\tilde{N}_\mu \cdot \tilde{N}_\mu^{\pm 3/4}$ 
% dn_train = x1;          %  $\tilde{N}_\mu \cdot \tilde{A}_\mu^{\pm \epsilon}$ 

xn_test = xn_train;        %  $2\hat{a}\hat{E}\hat{O}\tilde{N}_\mu^{\pm 3/4}$ 
dn_test = dn_train;        %  $2\hat{a}\hat{E}\hat{O}\tilde{A}_\mu^{\pm \epsilon}$ 

%-----
%  $\sigma^- \hat{E}_\mu^{1/2} \hat{O}_\mu \hat{U}, {}^3\hat{O}_\mu$ 

NodeNum = 16;              %  $\hat{O}p^2\hat{a}^{1/2}\hat{U}_\mu\hat{a}\hat{E}_\mu$ 

```

```

TypeNum = num_sample/2;          % 输入输出层神经元个数
p1 = xn_train;                    % 训练输入数据
t1 = dn_train;                    % 训练目标数据
Epochs = 1000;                   % 训练次数

P = xn_test;                      % 测试输入数据
T = dn_test;                      % 测试目标数据

%-----
% 网络初始化

%TF1 = 'tansig';TF2 = 'purelin'; % 网络初始化为tansig和purelin
%TF1 = 'tansig';TF2 = 'logsig';
TF1 = 'logsig';TF2 = 'purelin';
%TF1 = 'tansig';TF2 = 'tansig';
%TF1 = 'logsig';TF2 = 'logsig';
%TF1 = 'purelin';TF2 = 'purelin';

net = newff(minmax(p1),[NodeNum TypeNum],{TF1 TF2},'trainlm');

% 网络训练
%net.trainFcn = 'trainlm'; % 网络训练函数为trainlm
%net.trainFcn = 'trainbfg';
%net.trainFcn = 'trainrp'; % 网络训练函数为trainrp
%net.trainFcn = 'traingda'; % 网络训练函数为traingda
%net.trainFcn = 'traingdx';

```

```

net.trainParam.epochs = Epochs;      % ×î'óÑμÁ·'ÎÊý
net.trainParam.goal = 1e-8;           % ×îĐı¼ù·½ÎÓı
net.trainParam.min_grad = 1e-20;      % ×îĐıİÝ¶Ê
net.trainParam.show = 200;            % ÑμÁ·İÖÊ¾¼¼ä,ô
net.trainParam.time = inf;             % ×î'óÑμÁ·Ê±¼ä

%-----

% ÑμÁ·Óë²âÊÖ

net = train(net,p1,t1);                % ÑμÁ·
X = sim(net,P);                        % ²âÊÖ - Êä³õÎªÖ²âÖμ
X = full(compet(X))                   % ¾°ÖùÊä³õ

%-----

% ½â¹ûİ¾¼Æ

Result = ~sum(abs(X-dn_test))          % ÖýÈ··ÖÀàİÖÊ¾¼¼1
Percent = sum(Result)/length(Result);   % ÖýÈ··ÖÀàÂÊ

% testData1 = load('testdata2.txt');    % ËÄÇ;Êý¾Ý
% num_input=length(testData1)/nnn;
%
% testData1=testData1';
% testData2=testData1(1,:)+testData1(2,:)+testData1(3,:)+testData1(4,:);
% testData1=testData2'/4;
%
% for i=1:num_input

```

```

%      testData(i,:)=testData1(1+(i-1)*nnn:i*nnn);

% end

% testData=testData';

%

% % testData=testData'

% X11 = sim(net,testData);

% Ô²âÆäËüËËÊÂ

testData2 = load('2021auopen_8.txt');    %ËÄÇ;ËÝ³⁄₄Ý

%%% ÓÃÆ½³⁄₄üÖµ

testData2=testData2';

% testData3=testData2(1,:)+testData2(2,:)+testData2(3,:)+testData2(4,:);

rawData_temp=zeros(1,16*nnn);

for i=1:num_round

    rawData_temp=rawData_temp+testData2(i,:);

end

% for i=1:num_round

%      testData3=testData3+testData2(i,:);

% end

testData2=rawData_temp/num_round;

testData2=testData2(1:nnn*num_sample);

%%% ÓÃ×î¹½üÖ»ÂÖ

```

```

% testData2=testData2';
% testData3=testData2(4,:);
% testData2=testData3';
%%
for i=1:num_sample
    predict_Data(i,:)=testData2(1+(i-1)*nnn:i*nnn);
end
predict_Data=predict_Data';

% testData=testData'
predict_result = sim(net,predict_Data);
predict_result = full(compet(predict_result))           % ¼°ÛÊä³ö

# 4. 5-Fold cross-validation
kf = KFold(n_splits=5)
for train,test in kf.split(feature):
    x_train = feature[train]
    y_train = rank[train]
    x_test = feature[test]
    y_test = rank[test]
    mlp.fit(x_train,y_train)
    pred = mlp.predict(x_test)
    print('each folds MSE=',np.mean((y_test-pred)**2))

```

参考文献

- [1]姜一鹏.第28届奥运会中国奖牌预测[J].天津体育学院学报,2004(02):86-87.
- [2]江立平.用统计回归方法预测我国代表团在2008年奥运会上的奖牌数[J].南京体育学院学报(自然科学版),2007(02):53-55.
- [3]王小春.网球运动员时空预测特征及ERP研究[D].上海体育学院,2012.
- [4]蒋彬,王亚兵.第12届全运会网球项目奖牌分布情况预测与分析[J].辽宁体育科技,2013,35(03):80-82+86.DOI:10.13940/j.cnki.lntykj.2013.03.031.
- [5]杨志敏.男子网球单打比赛成绩预测方程建立[J].北京体育大学学报,2010,33(04):143-145.DOI:10.19582/j.cnki.11-3785/g8.2010.04.039.
- [6]任波.李娜与世界优秀女子网球运动员的技术统计分析——以2013年四大满贯、中网、WTA年终总决赛为例[J].体育科技文献通报,2014,03:66-68.
- [7]邢晓圆,郭开强,陈万军,冯石.对2006年澳大利亚网球公开赛女单选手技战术指标的统计分析[J].辽宁体育科技,2006,06:59-60.
- [8]邢晓圆,郭开强,陈万军,焦雪华.对中国女网单打选手参加06、07年“大满贯赛事”的实力分析[A].中国体育科学学会.第八届全国体育科学大会论文摘要汇编(一)[C].中国体育科学学会:,2007:1.
- [9]罗溢杰.世界优秀男子网球运动员硬地赛事“关键分”技战术分析[D].湖南师范大学,2012.
- [10]李国立.2013年中国网球公开赛女子单打比赛制胜指标的模型建立及评价[J].河南师范大学学报(自然科学版),2014,01:180-184.
- [11]张潮.影响中国女子网球单打技战术因素的因子分析[J].体育科技文献通报,2014,11:51-52.
- [12]郭立亚,袁毅,关晓燕,陈马强.世界顶级网球男子单打比赛制胜技术因素分析[J].北京体育大学学报,2010,02:122-124.
- [13]杨海兵.世界优秀男子单打网球选手三项指标的研究[D].北京体育大学,2009.
- [14]付饶,孙卫星.硬地女子网球双打比赛三项指标实力评估法的建立与应用[J].中国体育科技,2008,04:79-83.

- [15]谢相和. 草地女子单打网球运动员制胜指标分析及评价[J]. 成都体育学院学报, 2012, 02:74—78.
- [16]高徐, 田广. 世界优秀红土场女子单打网球运动员制胜指标及实力评价[J]. 西南师范大学学报(自然科学版), 2014, 04:198—204.
- [17]陈国豪. WTA 选手比赛技战术应用数据统计研究[D]. 江西师范大学, 2016.
- [18]Claudino JG;Capanema DO;de Souza TV;Serrão JC;Machado Pereira AC;Nassis GP. Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: a Systematic Review. [J]. Sports Medicine – Open, 2019, Vol.5(1):28.
- [19]Adam Maszczyk;Artur Golas;Przemyslaw Pietraszewski;Robert Roczniok;Adam Zajac;Arkadiusz Stanula. Application of Neural and Regression Models in Sports Results Prediction[J]. Procedia – Social and Behavioral Sciences, 2014, Vol.117(0):482–487.
- [20]Rocio de la Iglesia, Isabel Espinosa-Salinas , F Javier Lopez-Silva rrey, J Jose Ramos-Alvarez , J Carlos 23.Segovia , Gonzalo Colmenarejo , Elena Borregon-Rivilla , Helena Marcos-Pasero , Elena Aguilar-Aguilar, Viviana 24,Loria-Kohen , Guillermo Reglero , Ana Ramirez-de Molina . A Potential Endurance Algorithm Prediction in the Field of Sports Performance[J]. Frontiers in genetics, 2020, Vol.11(0):711.
- [21] Loeffelholz, Bernard, Bednar, Earl, Bauer, Kenneth W. Predicting NBA Games Using Neural Networks[J], Journal of Quantitative Analysis in Sports, 2009, 5(1): 7.
- [22]John McCullagh. Data Mining in Sport: A Neural Network Approach[J]. International Journal of Sports Science and Engineering, 2010, 4(3): 131–138.
- [23]蒋启飞, 郑贺. 男子网球单打技术与综合实力回归预测模型构建分析[J]. 吉林体育学院学报, 2015, 31(02):39–43. DOI:10.13720/j.cnki.22-1286.2015.02.009.
- [24]蒋勇, 戴煜, 郁红高等. 运动成绩预测的灰色模型方法[J]. 体育科学, 1988, (4):41–43, 95.

- [25]杨宗友,夏思永,曾凡波.第30届奥运会男子400米自由泳冠军成绩预测[J].体育文化导刊,2010(07):34-36.
- [26]陈亮.田径世锦赛男子成绩发展态势及灰色预测研究[J].体育科技,2006,第27卷(1):47-50,66.
- [27]张银满.世界优秀男子网球单打选手硬地赛制胜因素[J].北京体育大学学报,2009,32(10):135-137.DOI:10.19582/j.cnki.11-3785/g8.2009.10.040.
- [28]周政.BP神经网络的发展现状综述[J].山西电子技术,2008(02):90-92.
- [29]韩力群.人工神经网络理论、设计及应用[M].北京:化学工业出版社,2004.
- [30]王杰.基于资源库的运动技战术训学研平台构建[J].上海体育学院学报,2018,第42卷(3):74-78.
- [31]龚剑.基于人工神经网络2008奥运会中国男篮成绩预测实验研究[D].武汉体育学院,2007.
- [32]宋爱玲,刘静,李征,陈锴.基于BP神经网络的高水平女子100m栏专项成绩预测模型[J].首都体育学院学报,2012,24(03):265-267.
- [33]张宇,袁晓曦,弓小倩.基于BP神经网络算法的体育成绩预测研究[J].科技通报,2013,29(06):149-151.
- [34]夏祥伟,毛丽娟,黄金玲,杨剑,刘单,贾璐.中国高校研究生体育锻炼与全面健康的相关性——基于BP神经网络的实证研究[J].全球教育展望,2018,47(04):111-128.
- [35]徐子哲.局部阴影条件下光伏阵列MPPT控制系统设计[D].东北农业大学,2017.
- [36]杜娜娜,杜少武,林慧.我国城市运动会田径速度类项目成绩发展趋势及灰色预测研究[J].中国体育科技,2012,48(06):10-13.DOI:10.16470/j.csst.2012.06.004.
- [37]曲淑华,张晓东,尹贻杰.世界田径锦标赛标枪成绩发展态势分析及灰色预测[J].北京体育大学学报,2017,40(11):93-97.DOI:10.19582/j.cnki.11-3785/g8.2017.11.016.
- [38]王萍.基于人工神经网络的旅游需求预测理论与实证研究[D].西北师范大学,2004.
- [39]吴丹.《管理决策方法——理论、模型与应用》:河海大学出版社,2014年

- [40]张小龙. 时间序列自回归(AR)模型在体育预测中的应用[J]. 北京体育大学学报, 2010, 33(02):86-88. DOI:10.19582/j.cnki.11-3785/g8.2010.02.023.
- [41]张锐, 张弛, 吴飞. 1985—2014年7次中国青少年体质健康监测的速度素质研究及长期发展预测[J]. 北京体育大学学报, 2019, 42(08):16-26. DOI:10.19582/j.cnki.11-3785/g8.2019.08.003
- [42]魏春玲, 孙晋海. 奥运会田径项目成绩发展趋势及灰色预测研究[J]. 中国体育科技, 2005(02):20-22+29. DOI:10.16470/j.csst.2005.02.005.
- [43]王国凡, 赵武, 刘徐军, 丰淑慧, 薛二剑, 陈林, 王波. 基于GA和回归分析的奥运会成绩预测研究[J]. 中国体育科技, 2011, 47(01):4-8+16. DOI:10.16470/j.csst.2011.01.002.
- [44]周政. BP神经网络的发展现状综述[J]. 山西电子技术, 2008(02):90-92.
- [45]蒋婷, 李庆. 职业网球赛事男子单打比赛胜负影响因素研究——基于2014—2018年四大满贯赛事[J]. 中国体育科技, 2021, 57(07):62-68.
- [46]郭立亚, 袁毅, 关晓燕, 陈马强. 世界顶级网球男子单打比赛制胜技术因素分析[J]. 北京体育大学学报, 2010, 33(02):122-124. DOI:10.19582/j.cnki.11-3785/g8.2010.02.033.
- [47]葛哲学, 孙志强. 神经网络理论与Matlab R2007实现[M]. 北京: 国防工业出版社, 2007: 36.
- [48]陶志翔, 祁兵, 胡亚斌, 裘荣, 董保健. 对网球发球体系的探究[J]. 北京体育大学学报, 2004(12):1695-1697. DOI:10.19582/j.cnki.11-3785/g8.2004.12.042.
- [49]曾繁繁, 李庆有. 优秀职业网球运动员比赛得分分析[J]. 湖北体育科技, 2006(01):67-69.
- [50]李苏, 徐薇薇. 我国女子网球运动发展研究[J]. 体育文化导刊, 2011(01):64-66.