

改进启发式优化算法在数据科学预测分析中的应用

李珊珊¹ 卢杰² 贾诗音¹

1.湖北商贸学院 湖北 武汉 430000

2.桂林理工大学 广西 桂林 541000

【摘要】：数据科学预测分析已经成为复杂系统决策支持的重要技术方法，其分析精度在很大程度上取决于模型参数寻优效率和全局收敛性。针对高维、非凸性和多峰性目标函数，传统优化方法易陷入局部极值，且收敛速度受限，难以满足大规模数据分析的需要。启发式算法因其对梯度信息依赖较弱、全局寻优能力强等优点，可有效替代参数优化，但仍面临收敛精度和计算效率之间的矛盾。本研究所提出的改进启发式优化算法将为提高预测模型的泛化能力和分析可靠性提供理论依据和技术支撑。

【关键词】：改进启发式；优化算法；数据分析；科学预测

DOI:10.12417/2811-0536.26.07.057

数据科学预测分析的核心是从海量数据中抽取有效的模式，并在预测结果中建立具有前瞻性的映射关系。启发式算法模拟自然演化或群体智能行为，通过迭代搜索获得全局最优解，为复杂预测模型参数整定提供灵活的工具。现有研究在算法收敛速率、求解空间探索和发展均衡、高维自适应能力等方面不断深入，但其性能与问题特性的自适应机理尚不明确。通过动态调整搜索策略和有机嵌入先验知识，在保持算法普适性的前提下，提高对特定数据分布的适应性，促进预测分析由经验驱动向智能化驱动发展。

1 改进启发式算法在预测流程中的作用

1.1 提升数据预处理适配与稳定能力

针对多源异构、高维、含噪声和缺失数据的真实数据，本项目提出一种基于数据分布特性和业务约束的方法，自主地进行特征筛选、冗余剔除、数据规整和离群点处理，避免人为干预造成的主观偏差和效率低下等问题^[1]。在此基础上，利用自适应优化机制，快速定位出对预测结果有重要影响的关键信息，并削弱干扰，使得数据结构更加符合预测模型输入的需求。

1.2 强化参数寻优效率与精准程度

传统的人工调整或遍历优化方法容易陷入局部极值、耗时长、适应性差等缺点，而改进型启发式算法通过迭代优化和全局寻优能力，能够快速逼近复杂参数空间中的最优组合。该算法可以兼顾探索和开发能力，同时兼顾全局覆盖和局部精细化搜索，提高模型

参数和预测目标之间的匹配程度，保证模型在训练和推理过程中保持较好的状态。

1.3 优化预测模型泛化适用性能

基于约束优化、正则化指导和风险控制的改进启发式算法，可以减少过拟合和欠拟合风险。该算法可以根据数据的特点动态调整模型的拟合程度，使得学习的结果不再局限于局部特征，而是更加符合数据的内在规律。针对非平稳、非线性、时变性强的数据，算法能够增强模型对未知数据的推理能力，提高预测结果的一致性和可信性^[2]。

1.4 保障预测输出可靠与可控水平

在实际预报任务中，预报结果的稳定性直接影响着决策的判断，通过误差约束、趋势定标和结果校正等机制，有效地抑制随机干扰和异常波动对预报结果的影响，保证预报结果的连续性和稳定性。该算法可以定量评价预测结果并动态调整预测结果，辨识和控制偏差范围，提高预测结果的准确性和一致性。同时，该算法还可以根据业务需求设置约束，使得预测结果满足现实逻辑和边界约束，避免产生不合理的结果。

2 改进启发式优化算法在数据科学预测分析中的应用

2.1 优化数据预处理实践方法

针对多源多维预测数据，统一输入格式、数值单元和口径，去除重复输入的冗余数据项，利用启发式算法自带筛选规则，批量识别重复样本和重复字段，降低后续计算负担。开展缺失数据针对性处理，结合

作者简介：

第一作者：李珊珊，女（1998.04-），汉族，湖北襄阳人，硕士研究生，高级工程师，研究方向：人工智能、云计算、智能计算、优化算法。

第二作者：卢杰，男（1998.06-），汉族，江西九江人，硕士研究生，工程师，研究方向：云计算、大数据。

数据类型划分处理模式,以中位数填充连续数值型数据,以众数填充分类数据,同时,通过改进的启发式算法动态检索功能,定位离散型缺失数据分布区域之间,并根据数据相关性,实现差异性填充,避免单一填充方法带来的数据偏差。

针对异常数据,严格按照 3σ 准则和IQR规则执行筛选运算,并借助改进的启发式算法动态阈值调整功能,根据预测情景中数据波动范围,灵活调整异常识别阈值,批量剔除偏离数据分布规律的异常样本,并对关键异常数据进行暂存,防止误删有效数据^[3]。在完成数据清洗之后,进行数据规范化和归一化处理,借助算法内建的数据转换模块,对不同维度的特征数据进行统一压缩,消除数值量级差异对预测操作的影响。

同时建立预处理校验流程,利用改进的启发式算法实时校验功能对每个预处理过程进行采样,对数据的完整性和一致性进行采样,并对处理过程中产生的数据偏差进行及时修正。在大数据预测场景下,采用算法轻量级运行方式,对数据进行分批预处理,避免因批量操作造成的系统延迟,保证预处理过程的高效进行。将预处理过程和参数设定全过程记录下来,形成标准化的操作日志,保证不同预测数据集的预处理过程统一,为后续启发式算法优化应用提供一致的数据条件,从而提高预处理的准确性和稳定性。

2.2 优化特征筛选落地实践

以预测对象的需求为中心,对特征筛选的全过程进行精细化操作,简化输入维度,加强预测数据的匹配程度。通过预设特征筛选基础规则,结合预测任务类型,明晰特征筛选核心指标,以数据关联度、信息量贡献率、冗余度为核心评价准则,借助改进的启发式检索机制,实现原始海量特征的初筛,快速剔除无关联和贡献低的无效特征,压缩筛选范围^[4]。采用递进式启发式方法,将正向和反向两种操作相结合,先将有效特征逐步叠加到空白特征集上,再逐步去除冗余特征,并借助算法动态评估机制,对各特征子集的适应性进行实时判断。

在筛选过程中,对启发式算法的检索权重进行动态调整,针对不同场景(如时序、数值、分类等),分别对特征关联度评分权重进行差分调整,以保证筛选规则符合预测任务要求。通过对启发式算法迭代筛选函数进行迭代筛选,控制迭代次数和搜索范围,避免无限搜索带来的资源浪费,在每一轮筛选结束后,自动统计各特征子集的运行效率和适合度,并保留最佳筛选结果。针对高维复杂预测数据,采用算法分层

筛选模式,按特征层级划分筛选模块,逐层完成基本特征、衍生特征和关联特征筛选,层层剔除冗余特征,提高筛选精度。

开展特征筛选结果验证,采用交叉验证的方法,对筛选出的特征子集开展多轮验证,并借助改进的启发式算法错误检测功能,检测特征冗余和缺失信息,动态调整筛选规则。通过对特征个数的控制,根据预测模型的计算能力,对最终特征的个数进行合理的限制,避免因特征数量过多而导致的计算速度过慢,或者由于特征数量太少而导致信息缺失等问题。针对不同领域的预测数据,对算法参数进行微调,保证改进后的启发式算法能灵活适用于各种预测场景,不断优化其应用效率和应用效果。

2.3 优化算法参数调节实践

对启发式算法核心调整参数进行梳理,明确关键参数(如迭代次数、搜索范围、更新率、权值等)的调整区间,并结合预测数据规模确定参数上下限阈值,避免因参数超限而导致算法失效^[5]。采用分步调整的方法,优先调整基本操作参数,根据数据集的规模,调整算法的单次恢复样本量和执行速度,实现小样本下的搜索范围缩小,大数据下的预测场景扩大计算区间,以保证算法的顺利运行。

通过递进式的调试方法,对算法的全局搜索和局部搜索之间的平衡点进行逐级调整,在前期对全局搜索权重进行放大,扩大数据规则检索的范围,充分捕捉数据之间的潜在关联。后期缩小搜索范围,提高局部优化权重,聚焦核心数据规则优化计算。通过改进的启发式算法实时反馈功能,对调节后的运行状态进行实时监控,并对运转率、误差波动、工作稳定性等关键参数进行记录,并根据实际操作数据对参数进行动态修正,避免人为设定参数。借助算法完善机制,实现对波动型预报数据的参数自适应调整,实现对数据波动的实时感知,并对数据的动态变化规律进行自适应调整。

建立参数迭代优化流程,按预设调整周期,完成多轮参数调试和优化,每次调整结束后,开展短期运行试验,筛选出最佳参数组合并归档保存,得到不同预报情景下的参数配置库。对参数调整幅度进行严格控制,采用小幅度渐进调整方法,避免因参数大幅变化而引起的运算混乱,保证预测分析过程的稳定性。同时,做好参数调整的记录,对各参数的调节值、运行效果和适用场景进行详细记录,积累实践经验,为后续不同预测任务的算法参数配置提供参考,不断完

善启发式算法参数调整的标准化实践流程。

2.4 优化算法运行调度实践

为实现算法分阶段运行计划,按照预测分析流程,对数据适应、规则提取、优化操作、结果输出等运行阶段进行划分,并借助改进的启发式算法的分段调度函数,对各个阶段的运行优先级进行设置,优先完成核心计算环节,对运行资源进行合理配置。针对多任务并行预测场景,引入算法分时调度模式,对计算周期进行划分,并对多组数据进行集中计算,避免计算负荷过大造成计算延时。

对算法搜索调度策略进行优化,对传统启发式算法全局搜索模式进行改进,根据数据分布特点划分重点搜索区域和常规搜索区域,将计算重点放在核心数据区的规则挖掘上,弱化对低值数据区的计算投入,降低无效计算的开销。依托动态调度机制,对算法的运行负荷进行实时监控,当运行压力过大时,自动启动轻量级运行模式,简化非核心操作环节^[6]。在计算能力充足的情况下,启动深层检索模式,加强数据规则挖掘的深度,在运算效率和分析精度之间取得平衡。在时序类中引入时序调度机制,根据时序规则安排操作次序,保证时序数据操作的连续性和逻辑性。

完善算法的启停和调度控制操作,设置自动启停机制,在数据输入的情况下,自动启动算法的运行,在预测计算完成之后,会自动暂停,降低闲置算力的消耗。建立运行异常调度预案,当算法发生卡顿、操作中断、数据匹配异常等情况时,自动启动调度调整,重启运算模块,调整运行节奏,切换运行方式,保证预测分析的持续推进。

2.5 优化预测迭代调优实践

建立预测迭代的基本规则,确定迭代操作的启动条件、运行步骤和结束准则,以预测误差、结果稳定性和数据匹配度为核心指标,基于改进的启发式算法

实时监控功能,实时判断迭代最优启动时间,当预测偏差超过阈值时,自动启动迭代优化过程。进行分层迭代运算,第一波迭代聚焦于数据全局规则的挖掘,借助启发式算法的全局检索能力,对数据核的变化规律进行梳理,并输出初步的预测结果。

在后续迭代中,逐步缩小优化范围,聚焦于预测偏差较大的数据区间,借助改进算法的精确检索功能,细化数据波动性特征分析,有针对性地修正操作偏差,逐步减小预测误差。在迭代过程中,对启发式算法的优化策略进行动态调整,强化特征关联分析和数据规则挖掘,实现数据区间稳定,简化迭代操作步骤,提高迭代优化效率。同时,对迭代过程进行数据管控,在每次迭代结束后,自动记录操作数据、偏差值和优化方向,生成迭代日志,为后续优化调整提供数据支持,避免重复优化操作。

采用周期迭代机制,对长周期、动态变化的预测任务进行周期性迭代,并与实时更新的服务数据相结合,对操作基准进行更新,对预测模型的操作逻辑进行修正,以适应数据的动态变化。在迭代优化结束后,进行结果验证,采用多维度验证方法,对预测结果进行合理性和稳定性检验,剔除异常迭代结果,形成优化迭代算法。

3 结语

综上所述,不断探索算法改进与预测任务特征之间的内在联系,可实现自适应搜索机制与数据分布特性的动态自适应调整,发挥启发式算法在复杂预测场景下的潜力。未来,随着计算基础设施的不断完善和算法理论的逐步深化,改进后的启发式优化算法能够更好地结合深度学习等先进技术,使其能够更好地处理高维、复杂结构的预测分析任务,为数据驱动的决策提供更准确、更可靠的技术支持,促进大数据科学应用走向更高的智能化水平。

参考文献:

- [1] 张家强,芦艾,康明,余凤涓.小数据机器学习算法在材料科学领域中的研究进展[J].橡胶工业,2026,73(04):309-319.
- [2] 黄侃.人工智能驱动另类科学知识生产的现状和图景[J].自然辩证法研究,2026,42(03):89-97.
- [3] 肖濛,周骏丰,周园春.基于强化学习的科学数据特征生成算法[J].计算机研究与发展,2025,62(09):2127-2138.
- [4] 常月.数据挖掘算法在大数据网络安全防御中的应用研究[J].互联网周刊,2025,(13):18-20.
- [5] 张文魁.数字经济中的数据和算法规制[J].重庆理工大学学报(社会科学),2022,36(10):8-13.
- [6] 周佳茵,钱庆,唐明坤,吴思竹.科学数据引用识别方法研究[J].数据分析与知识发现,2023,7(06):38-49.