

创新点/亮点: 采用时间序列模型预测未来 5 年中国石油石化领域大数据算法趋势, 解决了石油石化领域大数据算法未来发展方向不明确造成的科技攻关方向无法聚力的难题, 挖掘出整个石油石化领域现有的 25 种算法中 9 种可能最适合; 勘探开发领域现有的 24 种算法中 5 种可能最适合, 油气集储领域现有的 19 种算法中 6 种可能最适合, 石油化工领域现有的 13 种算法中可能 4 种最适合。

引用格式: 李荣光, 金龙, 孙伶, 赵俊淇, 陈斯迅, 郑力会. 时间序列统计法预测中国石油石化领域大数据算法发展趋势 [J]. 石油钻采工艺, 2024, 46(5): 525-548. // LI Rongguang, JIN Long, SUN Ling, ZHAO Junqi, CHEN Sixun, ZHENG Lihui. Time series statistical method predicts the trend of big data algorithms in China's petroleum and petrochemical industry [J]. Oil Drilling & Production Technology, 2024, 46(5): 525-548.

时间序列统计法预测中国石油石化领域大数据算法发展趋势

李荣光¹, 金龙^{2,3*}, 孙伶¹, 赵俊淇², 陈斯迅¹, 郑力会^{2**}

1. 国家管网集团北方管道有限责任公司信息中心, 河北廊坊
2. 中国石油大学(北京)石油工程学院, 北京昌平
3. 中国教育科学研究院教育统计分析研究所, 北京海淀

*通信作者: 金龙, 电子邮箱: jinl@cnaes.edu.cn

**通信作者: 郑力会, 电子邮箱: zhenglihui@cup.edu.cn

基金项目: 国家科技重大专项“多气合采钻完井技术和储层保护”(编号: 2016ZX05066002)。

摘要: (目的意义) 适合中国石油石化领域的大数据算法经过多年试错探索, 依然没有找到适合石油石化领域数据特征的算法。由于反复试错不仅分散人力物力, 还影响了智能化发展进程。(方法过程) 从中文文献数据库中收集石油石化领域近 10 年来 800 多篇与算法相关文献, 涉及大数据算法相关性较强文献 87 篇, 按报道内容和关键词将文献分为勘探开发类(上游) 47 篇、油气集输类(中游) 25 篇和石油化工类(下游) 15 篇, 采用时间序列统计法拟合出的方程, 拟合优度 0.8 以上, 表明可以用来预测未来 5 年石油石化领域大数据算法发展趋势。同时, 以平均应用频率为尺度, 高于平均数的算法, 可能适用石油石化领域的的数据特征。(结果现象) 未来 5 年, 整体上石油石化领域应用算法解决具体问题时可以选择现有 25 种算法中的 9 种。上游可选择现有 24 种算法中的 5 种, 中游可选择现有 19 种算法中的 6 种, 下游可选择现有 13 种算法中的 4 种。(结论建议) 用定量的方法预测算法的未来趋势, 不仅揭示了石油石化领域大数据关键技术发展趋势, 为石油石化领域科研立项提供参考, 也为文献综述提供了一种可选择的方法。

关键词/主题词: 油气资源; 勘探开发; 油气储运; 石油炼化; 智能; 大数据; 算法; 文献综述

中图分类号: TE19; TE319; TE88 文献标识码: A DOI: [10.13639/j.odpt.202411030](https://doi.org/10.13639/j.odpt.202411030)

收稿日期: 2024-03-01; 修回日期: 2024-05-10; 录用日期: 2024-06-15; 编辑: 杨春莉

Time series statistical method predicts the trend of big data algorithms in China's petroleum and petrochemical industry

LI Rongguang¹, JIN Long^{2,3*}, SUN Ling¹, ZHAO Junqi², CHEN Sixun¹, ZHENG Lihui^{2**}

1. PipeChina North Pipeline Co., Ltd. Information Center, Langfang, Hebei, 06500, China
2. College of Petroleum Engineering, China University of Petroleum (Beijing), Changping, Beijing, 102249, China
3. China National Academy of Educational Sciences Institute of Educational Statistics Analysis, Haidian, Beijing, 100088, China

*Corresponding author. JIN Long, E-mail addresses: jinl@cnaes.edu.cn

**Corresponding author. ZHENG Lihui, E-mail addresses: zhenglihui@cup.edu.cn

Abstract: After years of trial and error in the oil and petrochemical industry, no clear consensus has been reached on which big data algorithms are most suitable for the specific characteristics of the sector's data. Repeated trial-and-error approaches have dispersed human and material resources and hindered the progress of intelligent development. Based on international and domestic literature

databases, 87 highly relevant articles were collected from over 800 publications related to big data applications in the oil and petrochemical sector over the past decade. These articles were categorized according to their content and keywords into three groups: exploration and development (47 articles), oil and gas storage and transportation (25 articles), and petrochemical processes (15 articles). Using the ARIMA model, the fitted equation, with a goodness-of-fit exceeding 0.8, indicates that it can be used to predict the development trends of big data algorithms in the oil and petrochemical field over the next five years. Additionally, the average number of algorithm application frequency was used as a benchmark; algorithms with publication counts above this average were deemed more likely to be suitable for the sector's data characteristics. Over the next five years, the number of algorithms applied to solve specific problems in the oil and petrochemical sector is projected to decrease from 25 to 9. In the exploration and development category, the number of algorithms is expected to reduce from 24 to 5. In the oil and gas storage and transportation category, the number is anticipated to decline from 19 to 6. For the petrochemical category, the number of algorithms is predicted to decrease from 13 to 4. The time series prediction of the future development trends of big data algorithms in the oil and petrochemical industry has revealed the trajectory of big data technology development and research gaps in this field. These findings provide valuable insights for choosing research project directions in the oil and petrochemical industry.

Key words: Oil and gas resources; Exploration and development; Oil & gas storage and transportation; Petroleum refining; intelligence; Big data; algorithms; literature review

<https://doi.org/10.13639/j.odpt.202411030>

Received 1 March 2024; Revised in revised form 10 May 2024; Accepted 15 June 2024

0 引言

过去十年,全球能源市场变化迅速,尤其在应对气候变化和实现碳中和目标的背景下,传统的石油天然气行业面临着前所未有的压力和机遇。此时,数字化转型为石油石化行业提供了发展新路径,且卓有成效。大数据在油气勘探、开发、生产、藏储、运输、炼化和销售等油气各个环节广泛应用,成为提升生产效率、优化资源配置、降低运营成本的关键。

中国作为世界上最大的能源消费国和石油生产国之一,在推动石油工业智能化、数字化转型进程中,面临数据、算法和算力等核心技术的诸多挑战。

为了应对挑战,相关部门出台了一系列政策支持石油石化行业大数据、人工智能和物联网等技术关键的开发应用。石油石化行业开始通过利用过去积累的数据,利用现有的算法,在现有的算力条件下,优化生产流程,提高勘探开发集输炼制的成功率,降低成本和风险。从上游的勘探开发到中游的集储输运,到下游的炼化销售,努力寻找适用石油石化数据的算法。在知网、Web of Science 等文献数据库中出现 800 多篇大数据算法的文献,其中,与石油石化相关性较强的算法研发和应用文献有 87 篇。这些文献主要聚焦石油石化领域大数据算法的具体应用,以这些文献为介质进行分析,对于了解石油石化领域大数据算法应用现状,预测行业大数据算法未来发展趋势具有指导作用。

2014 年,钱慧芳等采用 BP 神经网络(Back Propagation Neural Network)处理双远场电磁聚焦测厚仪的实验数据,实时监测石油套管破损^[1]。李明鑫等用蚁群算法(Ant Colony Algorithm, ACA)预测油气管道系统设备故障概率^[2]。金倬伊利用线性规划分析方法中的边际价值分析方法对每种原料的经济效益进行排序,从而得出吨原料的加工效益列表,实现了炼油原料结构优化^[3]。

2015 年,马林茂等将遗传算法(Genetic Algorithm, GA)与 BP 神经网络结合,优化了 BP 神经网络连接权值和阈值,预测大庆油田高含水阶段的油田产量,提升了复杂地质条件下的预测精度^[4]。骆正山等基于贝叶斯分类决策的概率人工神经网络(Artificial Neural Network, ANN),提出计算石油天然气管道可靠性的评估方法,可靠性增高^[5]。

2016 年,田亚鹏等将遗传算法(Genetic Algorithm, GA)与 BP 神经网络相结合预测页岩气产量递减,优化了 BP 算法的权重和阈值,提升了页岩气产量预测的精度^[6]。邱莹莹等基于粒子群算法(Particle Swarm Optimization, PSO)寻找多目标供应链网络方法,建立了基础油供应链生产-分销计划模型,以实现供应链总

成本和总反应时间最小^[7]。

2017年, 刘佳佳等采用遗传算法优化二氧化碳管道运输设计^[8]。LI Z L 等使用核主成分分析(Principal Components Analysis, PCA)和支持向量机(Support Vector Machine, SVM), 识别天然气管道阀门泄漏^[9]。

2018年, 陈一诺等利用随机森林算法(Random Decision Forests)分级预测管道缺陷, 识别影响管道安全的关键因素^[10]。

2019年, 刘胜娃等用人工神经网络建立了定向井机械钻速预测模型, 通过专家选定网络输入输出, 数据工程师设计网络拓扑结构和训练, 提升了定向井机械钻速的预测精度^[11]。周游等使用核主成分分析法识别了浊积岩岩性, 识别准确率90%^[12]。陈小龙等采用遗传算法和粒子群算法优化神经网络模型, 预测气体辅助重力泄油非混相开发油藏的采收率^[13]。刘彪等采用支持向量机优化钻井过程中的主控因素, 实现井漏预警和井漏风险评估, 降低钻井成本, 提高钻井效率^[14]。黄知娟等利用剥茧算法(Algorithm of Big-Data Cocooning, ABDC)无人干预地确定12项产量影响因素之间的定量关系, 确定顺北油田SHB-X井试采产液量骤降原因^[15]。艾月乔等提出了一种利用人工蜂群算法(Artificial Bee Colony, ABC)优化的支持向量机石油管道安全风险评估方法, 建立管道风险评估模型, 普适性和准确率较好^[16]。马钢等利用主成分分析法结合支持向量回归法, 建立了油气管道剩余强度预测模型, 准确性较高^[17]。曹晓红等提出混合人工神经网络建模技术用于石化装置异常工况监测预警, 实现装置异常工况早期感知^[18]。

2020年, Kim等应用对抗神经网络法建立地震波速预测模型, 断层识别准确度92%, 提升油气勘探的地震解释能力^[19]。Chen等将随机森林算法和BP神经网络结合, 建立多层记忆网络模型, 预测致密砂岩孔隙度^[20]。贾德利等使用随机森林算法, 构建定性分析的注水调整方向模型, 实现了老油田的精细注水优化, 精准预测了目标区块的产量递减率和含水上升率^[21]。王洪亮等提出综合循环神经网络和长短时记忆网络算法(Long Short-Term Memory, LSTM), 预测特高含水期产量, 预测准确度92.3%^[22]。张军华等采用支持向量机结合交叉验证方法(Cross Validation), 预测东营凹陷深部储层^[23]。徐磊等提出改进粒子群算法与BP神经网络结合的方法, 评估了三条原油管道能耗^[24]。郑度奎论述了基于神经网络、灰色预测(Grey Prediction)、支持向量机等油气管道的二氧化碳腐蚀预测方法^[25]。贺宗江等通过采用数据驱动的建模方法建立设备运行状态特征模型, 实现了基于大数据分析技术的预警预测^[26]。韩丰磊等建立了基于BP神经网络的煤制油装置中挥发性有机物泄漏的核算方法模型, 可以应用于煤制油装置中挥发性有机物泄漏的预警预测^[27]。危拓通过线性规划、分布递归等技术, 寻找炼油厂与化工厂联合运行的最优解, 提升一体化价值^[28]。

2021年, 杨午阳等基于深度神经网络模型, 识别不规则数据体环境下的断层^[29]。郑力会等采用剥茧算法筛选出影响平均漏失速率的17个主要因素, 为控制井下漏失提供数据支撑^[30]。田高鹏等利用自组织神经网络预测油气藏分布, 预测结果与实际吻合度94.1%, 提升油气藏的识别精度^[31]。谷宇峰等应用粒子群算法和极限学习机模型, 识别姬塬油田致密砂岩岩性, 识别精度90%^[32]。Pirizadeh等建立综合随机森林与决策树(Decision Tree)建立采收率预测模型, 规避大数据处理中的局部最优解问题^[33]。Cheraghi等将遗传算法、人工神经网络和支持向量机相结合, 优化提高采收率, 参数准确率92%^[34]。Wang等用卷积神经网络的图像识别模型, 识别三维剩余油宏观分布和微观富集状态^[35]。何佑伟等采用聚类分析法(Cluster Analysis)和随机森林算法预测不同页岩气井产量^[36]。李谦等使用皮尔逊相关系数法(Pearson correlation coefficient)、BP神经网络结合十折交叉验证法, 分别就低、中、高相关性参数, 预测不同的钻井速度^[37]。陶衫等利用剥茧算法, 确定顺北碳酸盐岩储层开采过程中井壁坍塌主控因素^[38]。曹继飞等利用主成分分析法降维处理声波测井数据, 建立碳酸盐岩裂缝识别模型, 提高裂缝预测准确性^[39]。马梦桐等基于核主成分分析算法、鸡群优化算法(Chicken Swarm Optimization, CSO)、相关向量机算法组合模型, 预测部分工艺管道腐蚀速率, 其预测效果优于其他预测模型^[40]。李诏阳等基于3层长短时记忆神经网络建立了润滑油加氢装置产品预测模型, 并通过随机森林分析预测数据相关性, 实现润滑油加氢装置操作在线优化^[41]。

2022 年, FANG 等引入贝叶斯分类器(Bayes classifier)和自适应双链量子遗传算法(Adaptive Double Chains Quantum Genetic Algorithm, ADCQGA), 优化 BP 神经网络岩石可钻性预测模型^[42]。Wei 等将生成对抗网络(Generative Adversarial Networks, GAN)、强化学习与决策树算法融合形成三维地震油藏预测模型, 减少反演计算时间^[43]。王泽龙等提出反向神经网络采收率预测模型, 预测蓬莱油田 600 余口注水井的采收率^[44]。Loomba 等使用遗传算法优化的聚类学习算法, 优化油田开发 workflow, 解决传统油田开发方案优化的时效性^[45]。侯亚伟等建立反向神经网络采收率预测模型, 提高了 125 组试验井的采收率, 平均采收率 56.15%^[46]。肖阳等运用皮尔逊和斯皮尔曼大数据相关性分析方法, 研究区块产量差异性^[47]。薛永超等用深度随机森林算法构建油井产量预测模型, 产量预测准确度 92.31%^[48]。王兵等采用支持度矩阵关联规则算法(Apriori), 识别钻井作业事故隐患^[49]。姜雪等结合沉积微相、粗粒相带及成岩相分布特征, 建立灰色储层构型识别模型^[50]。李超等提出粒子滤波算法(Particle Filter)预测石油石化生产涡旋压缩机热力性能, 预测精度较高^[51]。董丰莲等提出了基于变步长和复合最优下降判定的改进分布递归算法(Recursive Algorithm), 有效地求解石油产业链非线性规划问题^[52]。潘艳秋等应用 BP 神经网络算法, 预测某炼厂变换装置出口气体质量, 效果良好^[53]。祁滢等对比聚类分析结合长短期记忆网络和人工神经网络模型预测企业月能源消耗量和能源消耗总量, 发现人工神经网络模型较好^[54]。肖荣鸽等通过灰色关联分析预测管道腐蚀速率^[55]。鲁中歧等对比前馈神经网络和随机森林回归预测模型, 预测海底油气管道腐蚀泄漏, 发现神经网络预测效果更优^[56]。张新生建立因子分析和天牛须搜索算法(Beetle Antennae Search Algorithm, BAS)结合的极限学习机海洋油气管道外腐蚀速率预测模型, 预测结果更优^[57]。李明等应用广义回归神经网络对油气管道剩余强度进行预测, 精度和鲁棒性(稳定性)最好, 预测时间最短^[58]。

2023 年, 孔令维等使用 K 近邻算法(K-Nearest Neighbor, KNN)处理缺失数据, 长短时记忆网络预测油井产量^[59]。蒲堡萍等提出多层感知机神经和随机森林算法结合, 预测采收率^[60]。罗双平等利用皮尔逊相关系数定性分析法, 研究影响钻井速度的关键因素^[61]。郭秋麟等利用贝叶斯网络模型预测准噶尔盆地的油气勘探风险^[62]。邓增利等利用深度学习图像识别算法, 识别海上平台工作人员安全帽佩戴及不安全行为^[63]。盛科鸣等采用随机森林算法预测油气层敏感性损害^[64]。刘鸣慧应用深度学习模型, 实现了石化机组轴承故障诊断^[65]。冯博采用聚类算法, 预测石油管道泄漏几率^[66]。杜宇晨提出深度强化学习的腐蚀油气管道预防性维修最优化模型, 使得管道服役周期增加 10.84 年, 年平均成本为 389.63 元^[67]。闫绍栋等基于 BP 神经网络建立了油气管道振动检测模型, 报警正确率达 95%^[68]。赵年峰等通过斯皮尔曼相关系数和随机森林算法寻找腐蚀因素中相关性较高的变量, 并采用蚁群算法优化遗传算法, 预测油气管道腐蚀速率, 准确率更高^[69]。张力凡等提出迁移学习的改进型卷积神经网络与贝叶斯优化算法, 预测管道缺陷尺寸^[70]。

2024 年, 柴天祥等运用 Apriori 算法评估油气勘探事故发生阶段、事故类型和事故后果^[71]。蒲先渤等用主成分分析法预测岩石可钻性等级^[72]。杨子杰等应用多层感知器神经网络预测油气资源丰度, 结合 Boosting 集成学习算法, 确定主要地质因素^[73]。田龙等用测井大数据和无监督聚类算法评估连续地层可钻性^[74]。白生勇运用 BP 神经网络优选单井硬件设备匹配方案, 减少电动机平均装机功率, 降低抽油机井的维修维护成本^[75]。何旭晟等采用剥茧算法从 138 项工程数据中筛选影响漏失主控因素 16 项, 预测 2 井次漏点^[76]。刘明亮使用遗传算法, 计算分流后气体的流动参数^[77]。苏乾潇等使用快速傅里叶卷积(Fast Fourier Convolution, FFC), 实现西南油气田成像测井图像构建数据集并快速生成全井眼图像^[78]。张静等利用模拟退火优化-粒子群算法(Simulated Annealing-Particle Swarm Optimization, SA-PSO)求解管道泄漏位置^[79]。秦谢勋等利用改进蜂群算法和灰色模型预测油气管道剩余寿命^[80]。王东海等利用遗传算法优化避开激振源频率^[81]。罗仁泽等使用深度学习模型预测, 识别了油气管道 X 射线图像中焊缝缺陷数据集上缺陷^[82]。范蓬勃等基于核主成分分析、海鸥优化算法(Seagull Optimization Algorithm, SOA)和核极限学习机, 预测海底油气管道内腐蚀速率^[83]。黄书童等建立了混沌映射和莱维飞行改进的粒子群优化后的 BP 神经网络, 识别油气管道的缺陷类型^[84]。王栋建立灰色系统理论的两种模型评价海底管道腐蚀速

率^[85]。赵晓弘概述人工神经网络用于石油化工装置产品预测与操作优化、设备故障诊断、安全预警、节能减排等^[86]。冉安仙论述利用机器视觉算法实现人性化、智能化、数字化的石油化工场景危险行为检测和预警系统^[87]。

将 87 篇相关性较强的文献,按上游、中游和下游,即勘探开发、油气集储和石油化工销售三类,并对文章涉及算法和具体应用场景内容梳理汇总,见表 1。

表 1 2014—2024 年石油石化领域大数据算法应用文献分类

Table 1 Literature on Big Data Algorithmic Applications in Petroleum and Petrochemical Sector for Classification, 2014-2024

领域	主要应用场景	涉及主要大数据算法	涉及文献年份	文献数量
上游 (勘探开发)	油藏识别和勘探数据分析	BP神经网络、随机森林、核主成分分析法、灰色预测理论	2014, 2019, 2020	约15篇
	复杂地质条件下的产量预测	遗传算法优化的BP神经网络、长短期记忆网络、深度随机森林	2015, 2016, 2022	约12篇
	断层识别和地震数据分析	对抗神经网络、卷积神经网络、图神经网络	2020, 2021, 2023	约10篇
	钻井过程中的风险评估	支持向量机、K-means聚类、主成分分析法	2019, 2021, 2024	约7篇
	钻速预测	人工神经网络、支持向量机、决策树、深度学习模型	2019, 2021, 2024	约3篇
中游 (油气集储)	管道泄漏与腐蚀监测	灰色预测模型、贝叶斯分类器、粒子群算法	2015, 2018, 2023	约12篇
	储运安全优化	随机森林、蚁群算法、聚类分析	2017, 2021, 2024	约7篇
	储罐和设备故障预警	深度神经网络、决策树、生成对抗网络	2020, 2022, 2023	约6篇
下游 (石油化工销售)	炼油过程中的参数优化	BP神经网络、长短期记忆网络、核主成分分析	2020, 2021, 2022	约5篇
	销售预测和供应链管理	线性规划、粒子群算法、回归分析	2016, 2019, 2021	约2篇
	装置异常监测与故障诊断	混合神经网络、卷积神经网络、深度学习模型	2019, 2021, 2024	约4篇
	环保与节能减排监控	聚类分析、长短期记忆网络、回归分析	2020, 2022, 2024	约4篇

从表 1 中可以看出,大数据算法已在上游的勘探和开发环节、中游的油气储存和运输环节和下游的炼化与销售环节,均得到了大量的应用。但不同的算法的应用场景有所不同,有些算法更适合上游地震数据处理,识别复杂的地下构造辅助决策;有些更适合中游管道泄漏预测管理、能耗优化等场景。有些则更适合下游的工艺流程优化,减少能源消耗等场景。

另外,从算法应用趋势来看,尽管大数据技术在石油行业的各个环节中成果显著,但目前大数据算法的应用趋势仍不明确,主要表现在不同算法上中下游应用效果差异较大,为适应数据特征,不断开发新算法,不少算法融合,较多算法改进,也出现数学方法的引入。同时,也逐渐凸显出以下问题:

(1)不同算法在各环节的适用性分析不足。许多研究侧重于某种算法在特定领域的应用,缺乏对多种算法在不同油气环节中适用性的对比分析。

(2)未来算法发展趋势不明朗。随着机器学习、深度学习和强化学习等技术不断进步,哪些算法在未来几年内将被广泛应用,哪些将被逐渐淘汰,尚无系统性的研究。特别是目前的算法大多数源于非石油化工领域,一般适用性较差,用户对结果的满意度不高,影响大数据的发展,急需开发适合石油化工数据特征和 workflows 的算法。

(3)科研方向模糊,技术选型困难。由于缺乏对未来发展趋势的准确判断,石油企业和科研机构在技术选型和科研方向上,无法集中人力物力开发新算法,只能头疼医头,脚疼医脚,无法形成原始创造和底层基础,影响数字化转型进程和竞争力。

通过对近 10 年石油石化领域大数据算法应用文献分析发现,基于文献发表趋势分析和对文献涉及算法应用分析是解决石油石化领域大数据算法适用性和发展趋势的有效途径。以文献发表时间序列数据为抓手,采用时间序列数据处理算法分析石油石化领域大数据应用趋势,将使石油石化领域大数据算法趋势预测成为可能。所谓时间序列数据,是指按照时间顺序收集的数据,每个数据点都与特定的时间点相关联。这些数据点通常按照均匀的时间间隔(如每日、每周、每月等)测量和记录。例如,每天的股票价格、每月的销售额、每年的人口总数等,都是时间序列数据。分析时间序列数据,从中发现和揭示现象发展变化的规律,并将这些知识和信息用于预测。比如销售量是上升还是下降,销售量是否与季节有关,是否可以通过现有的数据预测未来一年的销售额等。采用时间序列分析预测时需要相应的数学模型,这种模型称为时间序列模型。

为了理清以上问题,根据梳理的近 10 年石油石化领域大数据算法应用文献发表时间序列和频率,采用时间序列预测方法,寻找时间与文章发表频率关系,并预测未来 5 年石油石化领域大数据算法发展趋势。分析不同算法在石油石化整个领域及上中下游发展趋势,为石油石化领域科研立项方向选择提供参考。

1 方法过程

1.1 室内研究

为了预测未来石油石化领域大数据算法的发展趋势。拟先将这些文章按时间顺序做趋势分析,然后再分三类做未来技术推测。

将 2014—2024 年石油石化领域大数据算法应用类文献,逐年统计每年发表数量,形成各年份发表数量,计算每年增长率,见式(1)。

$$\text{当年文献增长率} = (\text{当年文献发表数量} - \text{前一年文献发表数量}) / \text{前一年文献发表数量} \times 100\% \quad (1)$$

将每年文献发表数量、文献增长率数据绘制趋势曲线,可以得到近 10 年石油石化领域大数据算法的发

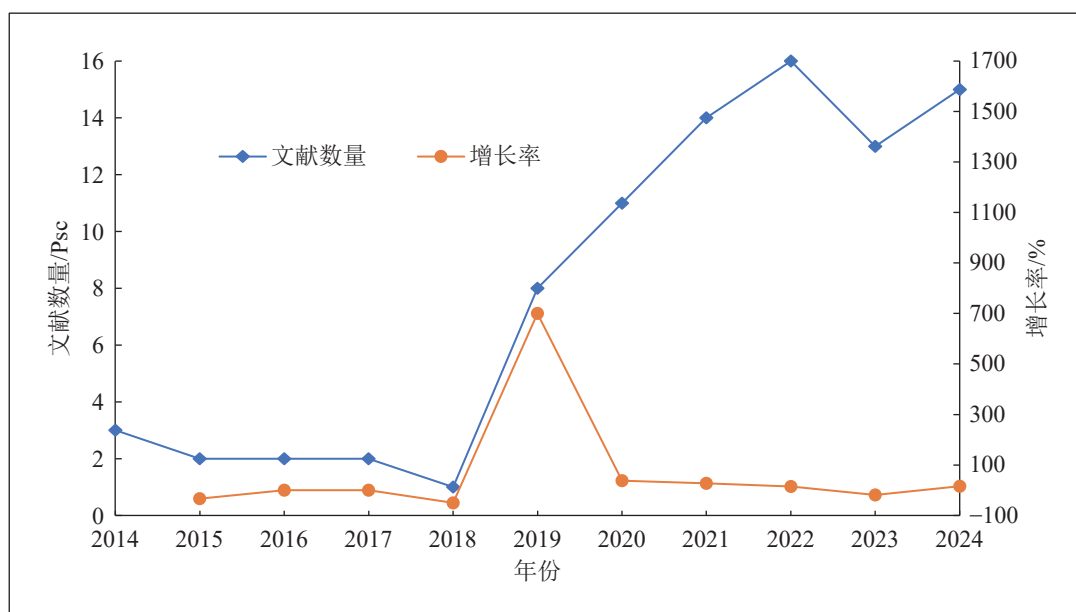


图 1 石油石化领域 2014—2024 大数据应用文献发表数量及增长率趋势图

Fig. 1 Trend of number of publications and growth rate of big data application literature in petroleum and petrochemical field, 2014-2024

展趋势, 如图 1 所示。

从图 1 中可以看出, 在 2014—2018 年, 文献数量相对较少且变化不大。2019—2022 年, 文献发表数量迅速增加。2023—2024 年, 文献数量又开始小幅下调; 从文献逐年增长率来看, 除 2019 年出现显著上升, 其他年份均处于平稳状态。

文献数量的预测通常需要更长时间数据捕捉整体趋势。增长率能更准确地反映短期内的相对变化, 对于短期预测和调整模型尤为适用。因此, 采用文献逐年增长率预测石油石化领域文献趋势。

文献增长率和年份的数据关系是典型的时间序列数据。目前, 用于时间序列预测的模型比较多, 如线性回归、移动平均、指数平滑等数学方法, 以及人工神经网络、支持向量机、长短时记忆网络、自回归差分移动平均模型(Autoregressive Integrated Moving Average Model, ARIMA)等机器学习方法。

线性回归只能捕捉线性趋势, 无法处理数据中的自相关性和复杂波动, 因此不适用于处理文献增长率这种波动性较大的时间序列数据。

虽然机器学习模型(如支持向量机、随机森林)在处理非线性关系方面有优势, 但它们对时间序列的特定模式(如自相关和趋势)捕捉较弱, 需要额外的特征工程和较大的数据集支持。尤其在当前的文献趋势分析中, 由于不涉及噪声平滑、季节性波动等问题, 无需使用移动平均、指数平滑等复杂处理方法。

相比于更复杂的模型, 如机器学习模型(如 LSTM 或神经网络), 自回归差分移动平均模型具有较低的计算复杂性和较短的训练时间。对于仅包含年度增长率的数据集, 能够快速有效地拟合数据而不需要大量计算资源。同时, 自回归差分移动平均模型的参数(如自回归和移动平均系数)有明确的统计学意义, 使得模型的可解释性更强。另外, 当前的文献增长率数据样本(2014—2024 年)较为有限, 对于时间跨度有限的数, 自回归差分移动平均模型能够通过差分和自回归捕捉到简单的趋势和周期变化, 而不需要过多参数调优。显然, 自回归差分移动平均模型更适合于文献趋势预测。

预测石油石化未来技术发展, 可以选择文献数量趋势或文章增长率趋势作为变量预测。然而, 基于时间序列的预测, 往往需要选择相对平稳的变量, 来抵抗短期异常波动, 增加预测结果的准确性。

由于文献数量有限, 采用阶数较高, 模型可能存在过拟合风险, 导致在处理长期预测时稳定性降低。且从增长率趋势图来看, 数据整体平稳, 不需要差分, 故可以采用 ARIMA(1,0,1)模型文献趋势预测。

根据 ARIMA 预测模型公式, 可得 ARIMA(1,0,1)预测拟合文献数量在不同考察周期内的变化方程, 见式(2)。

$$Y_t = \phi Y_{t-1} + \epsilon_t + \theta \epsilon_{t-1} \quad (2)$$

式中: Y_t 为时间 t 的预测值, ϕ 为自回归系数; θ 为移动平均系数; ϵ_t 为白噪声误差项。

ARIMA(1,0,1)预测模型计算和评估分四步。

步骤一, 数据准备与预处理。确保数据按时间顺序排列, 缺失值处理和基本预处理(如去掉异常值)。检查数据是否平稳。对于 ARIMA(1,0,1)模型, 数据需要相对平稳。

步骤二, 模型拟合。使用时间序列数据拟合 ARIMA(1,0,1)模型, 计算模型的系数, 如自回归系数 ϕ 和移动平均系数 θ 。

步骤三, 模型准确性和可靠性验证。通过检查残差的正态性和独立性, 可检验模型准确性和可靠性。一般采用残差标准差分析和 Ljung-Box 残差检验。1) 计算模型残差均值, 分析残差, 确保残差均值接近零且无显著自相关。如果残差显示出非白噪声特性, 模型需要重新调整, 见式(3)。

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i \quad (3)$$

式中: e_i 为第 i 个残差(实际值与预测值之差), n 为残差的数量。

2)计算残差标准差,检查残差的离散程度,如残差标准差服从正态分布,说明模型预测值可靠。残差标准差计算公式,见式(4)。

$$\sigma_e = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (e_i - \bar{e})^2}$$

(4)

式中: $\bar{e}$ 为残差的均值, e_i 为第*i*个残差。

3)在预测模型中,我们需要检查残差序列是否存在自相关性,也就是判断预测误差是否有规律。如果预测误差是随机的,说明模型预测效果较好。我们可以使用 Ljung-Box 检验来进行这个判断,计算见式(5)。

$$Q = n(n+2) \sum_{k=1}^h \frac{p_k^2}{n-k}$$

(5)

其中: n 表示样本的总数(即预测误差的数量), p_k 是第*k*阶误差的自相关系数, h 是滞后阶数(表示检查误差的滞后范围)。

计算得到的*Q*值可以结合自由度*h*通过卡方分布表查出对应的显著性水平*p*值。如果*p*值大于设定的显著性水平(例如 0.050),就可以认为残差序列没有显著的规律性,说明预测模型的效果是好的。

步骤四,根据预测模型公式,计算得到预测文献增长率和预测文献数量数据。

这样,可以直观分析 2024 年后 5 年内石油石化领域大数据算法应用文献数量的变化趋势以及未来发展方向。

1.2 现场应用

1.2.1 石油石化领域大数据算法发展趋势

为了分析未来石油石化领域大数据算法发展趋势,采用 ARIMA(1,0,1)模型预测 2025—2030 年石油石化领域大数据算法发表文献增长率,得到拟合方程,见式(6)。

$$Y_t = 0.528 \text{ } 1Y_{t-1} - 0.999 \text{ } 7_{\epsilon_{t-1}} + \epsilon_t$$

(6)

式中: Y_t 为年份*t*的文献增长率; Y_{t-1} 为年份*t*的前一年文献增长率; ϵ_t 是误差项, ϵ_{t-1} 是年份*t*的前一年文献增长率误差项。

使用计算公式计算模型残差均值及残差序列相关性进行检测模型拟合率,模型残差均值约为-0.052,接近零,符合残差为白噪声的假设。残差标准差约为 0.255,说明残差在小范围内波动。使用 Ljung-Box 检验,*p*值为约 0.859,远大于 0.050。表明残差无显著自相关性,即模型捕捉到数据的主要趋势,拟合效果较好。

根据模型拟合结果,可以得到 2025—2030 年十五年间石油石化领域大数据算法应用文献增长率及发表数量,见表 2。

表 2 2025—2030 年石油石化领域大数据算法应用文献增长率及发表数量预测结果

Table 2 Literature Growth Rate of Big Data Algorithm Application in Petroleum and Petrochemical Sector and Number of Publications Forecast Results, 2025-2030

年份	文献增长率/%	文章数量/Pcs
2025	-6.7	14.0
2026	-2.7	13.6
2027	-0.6	13.5
2028	0.6	13.6
2029	1.2	13.8
2030	1.5	14.0

将表 2 中石油石化领域 2025—2030 年大数据算法应用文献增长率及文献数量预测结果与 2014—2024 年历史数据绘图, 如图 2 所示。

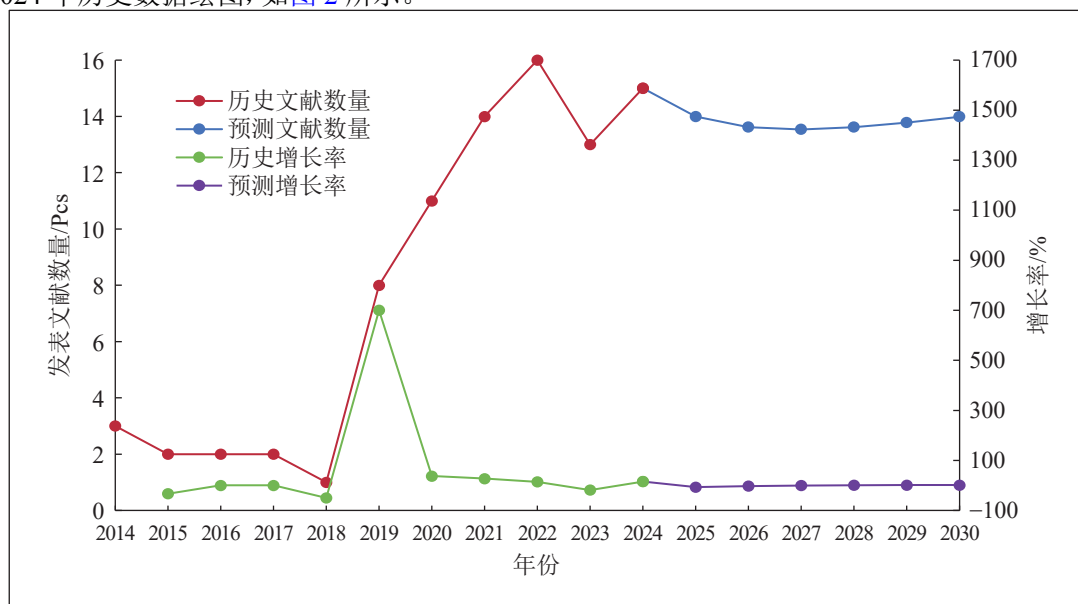


图 2 2014—2030 年石油石化领域大数据算法应用文献发表趋势

Fig. 2 Trend of published literature on big data algorithmic applications in petroleum and petrochemical fields, 2014-2030

从图 2 中可以看出, 2014—2024 年石油石化领域大数据算法应用文献呈现出先稳定、后快速增长的模式。

2014—2018 年, 文献数量相对较少且变化不大, 每年仅有 3 至 4 篇, 增长率波动较小甚至为负值。主要原因是该阶段大数据技术在石油石化领域的应用尚处于初期阶段, 行业对大数据算法还在探索, 技术应用范围较窄, 研究进展较为缓慢。大数据技术在各行业的普及度有限, 特别是复杂的油气行业对技术的适应速度较慢。因此, 石油石化行业的研究人员在尝试将大数据技术引入专业应用时, 面临数据积累不足、算法适配难度大等挑战, 影响了文献的发表数量。

2019 年起, 随着大数据技术和计算能力提升, 尤其是云计算、人工智能和物联网技术的普及, 石油石化行业逐渐发现大数据算法在生产、管理中的广泛应用价值。例如, 通过数据分析优化油田生产、降低运营成本、提高安全性等需求, 推动了算法应用的深入探索和研究。在全球能源市场波动和数字化转型的大背景下, 石油企业开始加速数字化转型, 对数据分析和智能决策的需求大幅增加, 促使行业关注大数据算法, 推动更多相关研究成果发表。文献数量迅速上升, 从 2018 年的 3 篇跃升至 2019 年的 8 篇, 实现了显著的增长。文献数量在 2020 年至 2022 年间持续增长, 到 2022 年达到了峰值, 年发表量达到 16 篇。这一增长趋势与全球数字化转型浪潮的进一步推动密切相关。尤其在 2020 年全球新冠疫情的影响下, 石油石化企业更加重视数字化手段提升企业韧性和生产效率, 大数据分析和人工智能算法被广泛应用于生产优化、风险评估、供应链管理等环节。文献数量的激增也表明, 大数据算法应用在石油石化行业的研究已逐渐进入一个成熟期, 更多的企业和研究机构投入其中, 推动了相关研究成果的快速涌现。同时, 国家出台政策支持石油石化行业的数字化转型, 鼓励智能化和自动化技术的研发, 行业内的合作项目增多, 相关研究得以快速推进并产出文献。

从预测结果来看, 2025—2030 年, 石油石化领域大数据应用文献数量基本稳定在每年 14 篇左右, 文献数量趋势相对平稳, 文献平均年增长率-1.1%。这表明在预测期内, 石油石化领域大数据算法的应用研究可能进入平稳发展的阶段, 增长不再快速, 但仍保持积极的科研输出。

随着大数据算法的应用逐渐成熟,覆盖油气勘探、生产优化、储运管理等全流程过程中,技术逐渐成熟,用户理性思考相关算法问题,增长率始终趋于平稳。石油石化领域大数据算法的应用研究主题转向深度优化与特定应用。在石油石化不同的应用场景中,更适合的算法将被选出并优化,一些不适合的算法将逐渐被淘汰。

1.2.2 上中下游业务中的大数据算法发展趋势

(1) 上游业务中的大数据算法发展趋势

为了了解石油石化上游大数据算法的应用趋势,从 87 篇文献中,将内容主要涵盖油气勘探和开发的 47 文献同样采用 ARIMA (1,0,1)模型预测 2025—2030 年文献增长率,得到拟合方程,见式(7)。

$$Y_t = 0.75Y_{t-1} - 0.65\epsilon_{t-1} + \epsilon_t \quad (7)$$

根据式(7),计算文献增长率和文献数量并绘图,如图 3 所示。

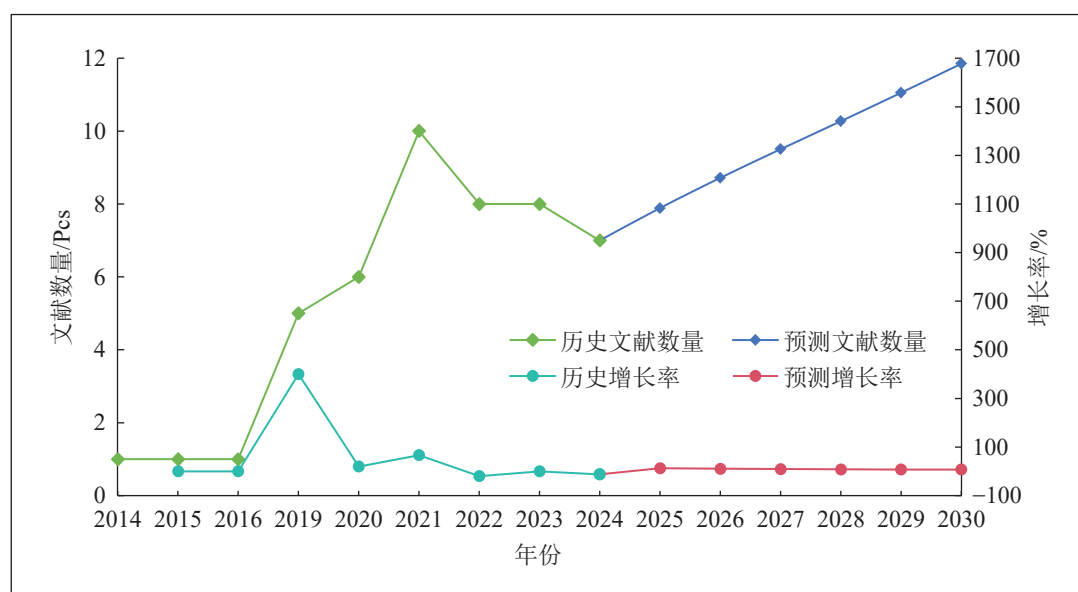


图 3 上游 2014—2030 年大数据算法应用文献发表与增长率趋势图

Fig. 3 Trend of Literature Publication and Growth Rate of Big Data Algorithm Application in Upstream Petroleum and Petrochemical Sector, 2014-2030

从图 3 中可以看出,2014 年至 2016 年,文献数量维持在较低水平,每年仅有 1 篇。此阶段大数据算法在上游业务中的应用处于起步阶段,研究热度较低。尽管大数据技术在其他行业中已经取得了初步成效,但在油气勘探开发领域,尚未规模化展开。其原因是,上游业务的数据复杂度高、数据量大且异质性强,开展过程中效果不如预期,对算法的适应性提出了更高的要求。因此,该阶段研究进展缓慢,文献数量增长乏力。

从 2018 年开始,文献数量出现较为明显的上升,尤其在 2019 年达到 5 篇,标志着大数据算法应用在上游业务中进入新的探索阶段。此时,全球范围内的数字化转型需求逐步加深,油气行业面临提升效率、降低成本的巨大压力。大数据算法被逐渐引入到油气勘探和开发的各个环节,如油藏建模、钻井优化和生产预测等,为该领域的学术研究注入新的动力。

到 2021 年,文献数量达到 10 篇的峰值。得益于大数据技术和人工智能算法的迅速成熟,如机器学习、深度学习和强化学习等技术的广泛应用,使得处理和分析复杂油气数据成为可能。上游业务需要处理大量地震数据、测井数据等复杂数据源,利用这些数据高精度预测和智能决策对油气勘探具有重要意义。因此,大数据算法的应用研究热度显著上升。

然而,2022 年起,文献数量开始呈现下降趋势,至 2024 年仅有 7 篇。表明,快速增长后,大数据算法在

上游业务中的应用研究进入了相对稳定的发展阶段。随着技术逐渐成熟, 算法应用研究从探索转向更为细致的优化研究, 关注点逐渐转向特定问题的解决效果及算法的效率提升。此外, 随着该领域的基础应用已经逐渐成熟, 新技术带来的研究增量有所减弱。

从增长率来看, 2018 年增长率出现大幅度上升, 主要与当时行业对大数据技术的需求增加密切相关。全球油气开采成本的上升及资源开发难度的增加, 促使企业和学术界对大数据算法兴趣提升。通过大数据算法, 可以实现对油藏和钻井作业的高效管理, 为油气资源的勘探和生产提供决策支持。随后 2019 年达到峰值, 大数据应用研究进入了快速增长期。

然而, 在随后的几年中, 增长率有所回落, 尤其在 2021 年后增长率出现负值, 显示出增长速度趋缓。原因是该领域的大数据算法研究逐渐进入成熟期。此时, 研究的重点已从算法的初步应用转向优化和实际问题的深度应用。随着大数据技术在上游业务的应用范围不断扩大, 简单的算法引入已经不足以吸引研究者的关注, 更多的研究聚焦于如何提高算法在实际场景中的应用效果。

从预测结果来看, 2025—2030 年, 石油石化领域上游大数据应用文献数量分别为 8 篇、9 篇、10 篇、10 篇、11 篇、12 篇, 呈现稳步增长趋势, 但增长率维持在较为平缓的水平, 年平均增长率 6.1%。符合石油石化领域上游业务中大数据算法应用的发展预期。随着大数据技术逐渐在行业中被成熟应用, 未来的研究可能集中在优化现有技术和深度应用中, 而非开创性的新技术应用。因此, 文献数量将保持增长, 但增长率将不会像初期那样显著攀升。大数据算法将更多用于提高油藏管理、钻井效率和生产优化中的智能化程度, 推动上游业务在数字化转型中的深入发展, 但大规模的研究, 几率可能不会太大。

(2) 中游业务中的大数据算法发展趋势

为了了解石油石化中游大数据算法应用趋势, 从 87 篇文献中, 将内容主要涵盖油气储存、运输和管道管理的 25 文献同样采用 ARIMA (1,0,1) 模型预测 2025—2030 年文献增长率, 得到拟合方程, 见式(8)。

$$Y_t = 0.62Y_{t-1} - 0.50\epsilon_{t-1} + \epsilon_t \quad (8)$$

根据式(8), 计算文献增长率和文献数量并绘图, 以获得 2024 年后的五年里文献增加的趋势和文献增长的增量, 以帮助研究者获得更多的算法趋势, 如图 4 所示。

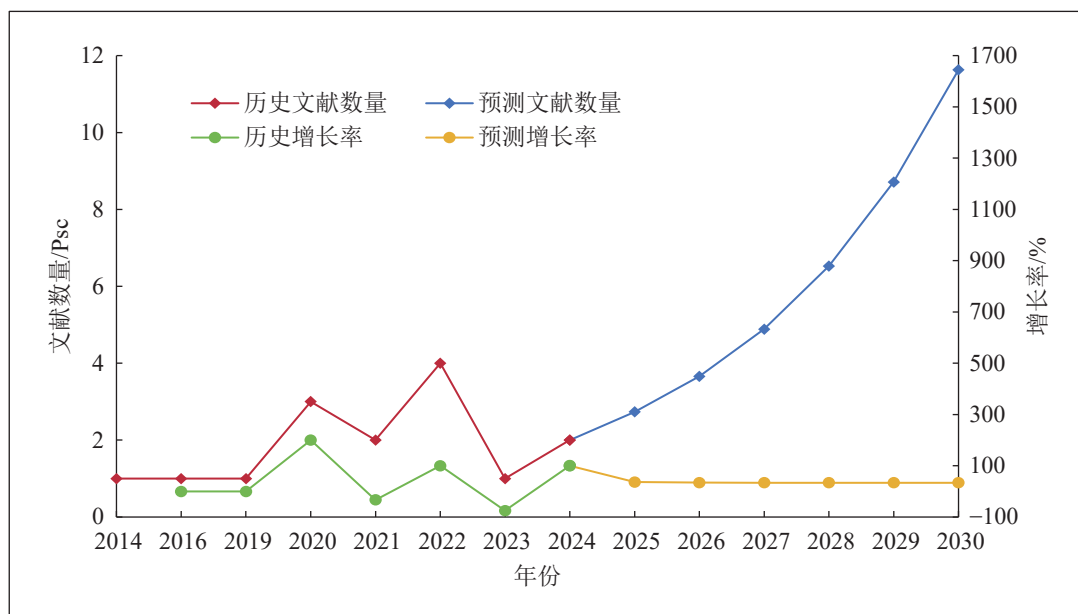


图 4 中游 2014—2030 年大数据算法应用文献发表与增长率趋势图

Fig. 4 Midstream 2014-2030 Big Data Algorithmic Applications Literature Publication and Growth Rate Trends

从图 4 中可以看出, 2014 年至 2018 年间, 中游业务中大数据算法应用类文献的发表数量处于较低水

平，每年仅有 1 到 2 篇。文献数量增长较为平缓，增长率波动不大。

2019 年开始，文献数量开始增加，特别是在 2021 年达到 4 篇，随后在 2024 年进一步上升至 6 篇。表明中游业务对大数据算法的应用需求逐渐增强。在行业数字化转型的背景下，石油石化企业开始关注如何在储运环节中更好地利用数据以提升效率、优化资源配置和保障管道运输安全。大数据算法的引入使得企业可以更有效地监控运输和存储过程中的数据，及时发现潜在的风险因素，提高了整体运营的效率和安全。这一阶段，研究人员和企业逐渐意识到大数据算法在中游业务优化中的潜力，从而推动了文献数量的增长。

从增长率趋势来看，虽然近几年上游大数据算法应用文献逐年增长，但是增长率整体平稳，主要原因在于中游业务中，早期的研究往往关注大数据算法在管道监测、储罐管理、运输调度等方面的基础应用。然而，随着研究的深入，研究方向逐渐集中于特定应用场景的优化，如异常检测、预测维护和能耗控制等。这类研究通常需要较高的技术水平和更长的开发周期，单纯依靠数量的增长难以反映研究的进展。因此，增长率整体趋于平稳，表明研究进入了深度优化阶段。

从预测结果来看，2025—2030 年，石油石化领域中游大数据应用文献数量分别为 7 篇、8 篇、10 篇、12 篇、15 篇、18 篇，年平均增长率 20.1%，呈现出更显著的逐年增长趋势。增长率在 2025 年后趋于平稳。文献数量的显著增长可能反映出行业对大数据技术应用的逐步深化，包括更复杂的算法在能源运输储集由优化、库存管理智能化向应用成果移植和精细化管理发展。

(3)下游业务中的大数据算法发展趋势

为了了解石油石化中游大数据算法的应用趋势，从 87 篇文献中，将内容主要涵盖石油化工、炼化、销售和供应链管理的 15 文献同样采用 ARIMA (1,0,1)模型预测 2025—2030 年文献增长率，得到拟合方程，见式(9)。

$$Y_t = 0.8Y_{t-1} - 0.7\epsilon_{t-1} + \epsilon_t \tag{9}$$

根据式(9)，计算文献增长率和文献数量并绘图，如图 5 所示。

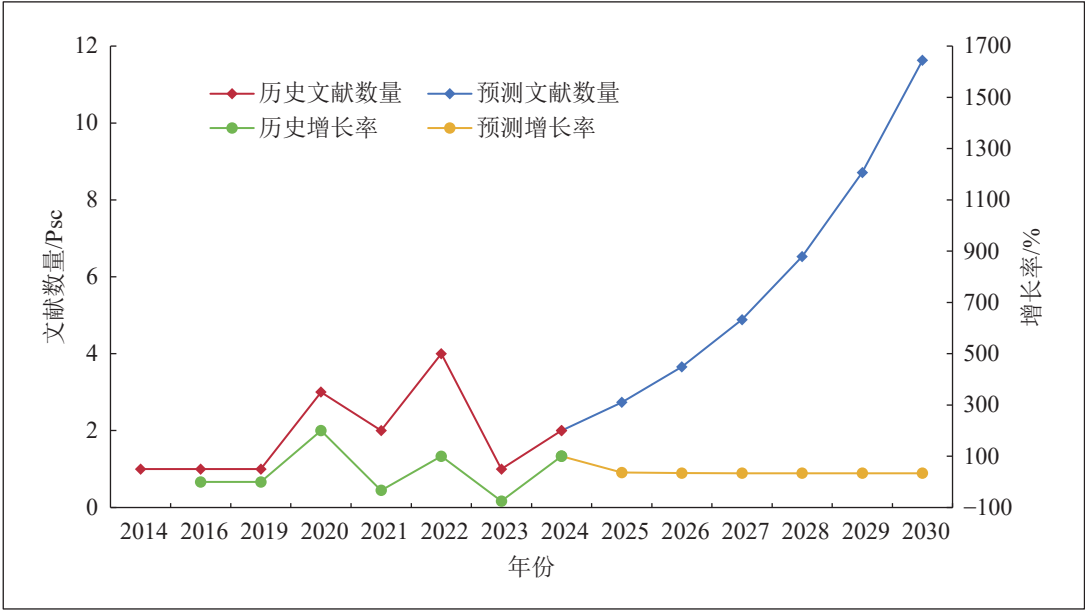


图 5 下游 2014—2024 年大数据算法应用文献发表与增长率趋势图

Fig. 5 Trend of Literature Publication and Growth Rate of Big Data Algorithm Application in Downstream Petroleum and Petrochemical Sector, 2014-2024

从图 5 中可以看出，2014 年至 2018 年，石油石化下游大数据应用相关文献趋势和上游、中游一致，每年仅有 1 至 2 篇，处于起步阶段。2020 年开始，文献数量有小幅增长，达到 3 篇，随后在 2022 年达到 4 篇的峰值。之后在 2023 年下降后，2024 年小幅回升至 2 篇。

总体来看，大数据算法应用在石油石化下游较上游和中游数量明显较少。可能是因为下游业务中的数

据需求与上游、中游不同。下游主要涉及市场预测、供应链优化等领域的传统数据分析方法,在其它行业已经相对成熟,通常不需要实时的复杂计算,技术可以通过成熟技术移植,就能够满足企业的基本需求。相比之下,上游和中游业务在勘探、储运和生产优化方面存在更多需要解决的复杂计算问题,更容易吸引学术界和工业界的关注。而在下游的化工炼化领域,数字化转型相对简单,传统的管理和预测方法依然占据主导地位,导致大数据算法在下游相关文献的研究和发表相对较少。

从预测结果来看,2025—2030 年石油石化领域下游大数据应用文献数量分别为 3 篇、4 篇、5 篇、7 篇、9 篇、12 篇,年平均增长率 34.1%,呈现出持续上升趋势,增长率趋于平稳甚至略有下降。表明大数据算法在下游业务中的应用在 2025 年后逐渐从早期的探索阶段过渡到应用的成熟期。随着技术逐渐成熟和行业接受度的提高,文献数量的增长将主要由应用的深化和优化推动,而非新的技术突破。这种趋势反映了行业对已验证技术的更广泛应用,如利用大数据进行智能供应链管理、客户需求预测和市场优化。

2 结果现象讨论

2.1 石油石化领域整体立项明确的原因

通过对石油石化整个领域及上游、中游和下游的大数据算法应用趋势分析发现,无论是整个石油石化领域,还是上中下游,对于大数据算法的应用都经历了先稳定、后快速增长的模式,然而,随着时间的推移和不同算法应用的逐渐深入,算法应用的增长率都逐渐平稳。

从具体算法来看,将 2014—2024 年石油石化领域涉及大数据算法应用的代表性文献 87 篇中,将 87 篇文献中,被应用到的大数据算法及出现的次数进行统计(含交叉应用),得出石油石化领域应用到的大数据算法 25 种,并将 25 种算法出现频率计算平均数可得算法平均应用频率,将每种算法的应用频率和所有算法平均应用频率统计绘图,如图 6 所示。

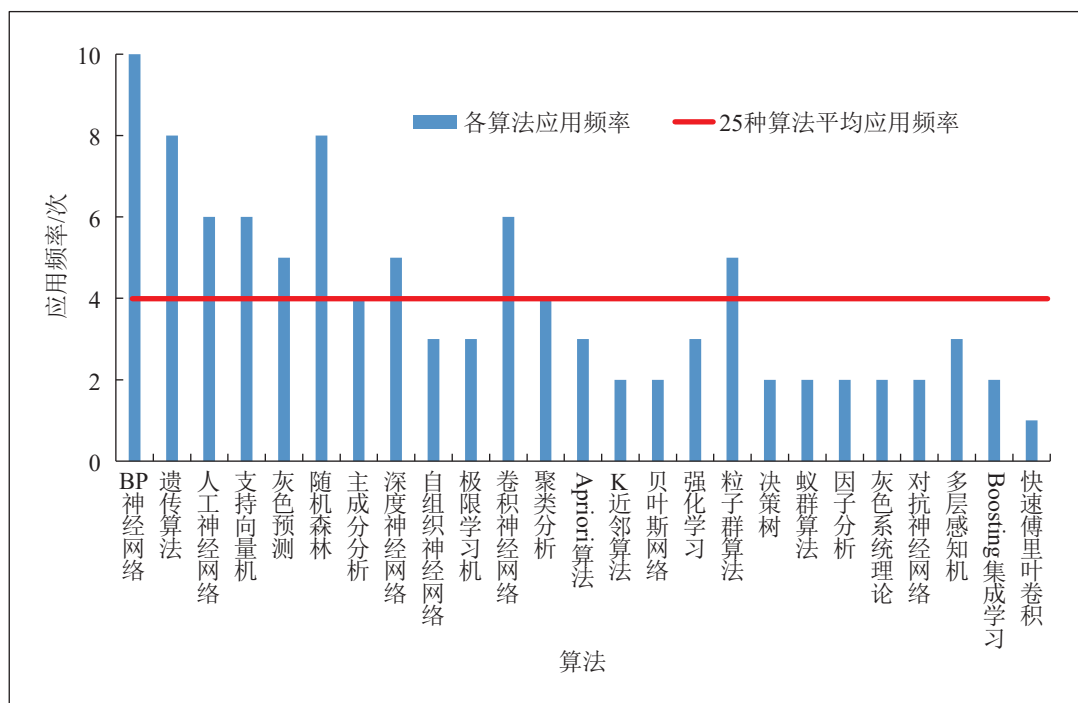


图 6 2014—2024 年石油石化领域大数据算法应用频率

Fig. 6 Frequency of application of different big data algorithms in the petroleum and petrochemical sector, 2014-2024

从图 6 中可以看出,石油石化领域应用到的大数据 25 种算法中,应用频率高于所有算法平均应用频率的算法有 9 种,分别为 BP 神经网络、遗传算法、随机森林、人工神经网络、支持向量机、卷积神经网络、灰色

预测、深度神经网络和粒子群算法,应用频率分别为 10 次、8 次、8 次、6 次、6 次、6 次、5 次和 5 次。应用频率高与所有算法平均应用频率(4 次)的算法往往具有较高的计算能力和预测精度。例如, BP 神经网络和支持向量机等算法在处理非线性问题和进行精确预测方面有良好表现。相比之下,低频算法如 Boosting、贝叶斯网络等,在石油石化行业的复杂数据场景中,难以应对高复杂度任务,尤其在动态环境中表现出局限性,可能难以提供同等水平的预测精度和可靠性。因此,随着计算资源的逐渐丰富,寻找适合的算法,企业和研究人员会更倾向于选择计算复杂度稍高但效果更佳的算法,导致低频算法逐渐被淘汰。

因此,未来在石油石化行业的复杂场景中,这些高频算法将更加广泛地应用。应用频率高的算法,如随机森林、人工神经网络、支持向量机等,具有较强的适应性和稳定性,能够适应复杂的非线性关系和多维度数据处理需求。

从数据来看,未来,石油石化领域在大数据算法的选择方面,将从原来的 25 种聚焦为 9 种,科研立项方向将更加聚焦于进一步优化和扩展 BP 神经网络、随机森林、遗传算法、支持向量机、卷积神经网络、粒子群算法的应用上,而非探索新的或不成熟的算法,从而使科研立项更加高效和针对性更强。在未来,这些算法通过改进或者应用确实不能胜任,开发新算法也有了针对问题开发的依据和应用对象。

2.2 石油石化上中下游立项更加明确的原因

在石油石化行业上游、中游和下游领域,大数据算法的应用呈现出显著的差异。不同的业务特点、数据类型以及技术需求直接影响了各类算法在这些领域的应用程度和广泛性。

2.2.1 上游立项更加明确的原因讨论

石油石化上游领域主要涉及石油和天然气的勘探、开发和生产。勘探环节包括地质勘探和地球物理数据处理,重点是发现新的油气藏;开发环节涉及油田规划与井位优化;生产环节则关注油井产量优化及现场运营管理。地质数据复杂且具有高度非线性特征,数据量庞大且时间跨度长,上游业务对能够处理高维、非结构化数据的算法需求迫切。

将 87 篇涉及石油石化大数据应用的代表性文献中,整理出涉及上游业务文献 47 篇,统计被应用到的大数据算法及出现的次数(含交叉应用),得到上游应用的算法 24 种,并将出现频率计算平均数可得算法平均应用频率,每种算法的应用频率和所有算法平均应用频率,如图 7 所示。

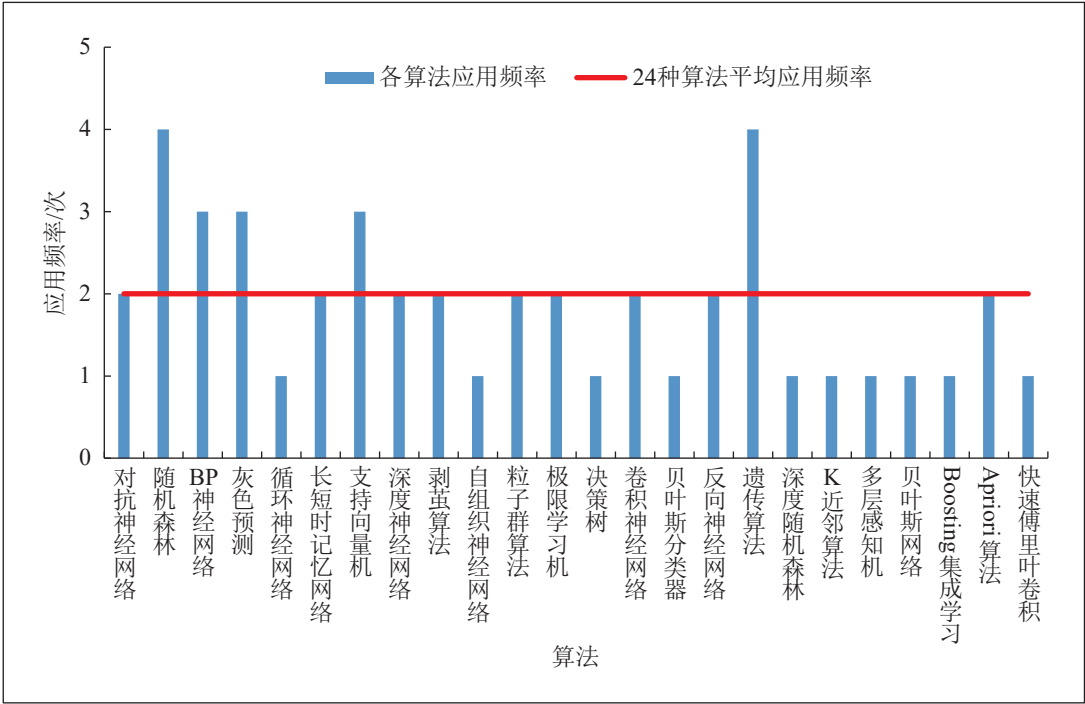


图 7 2014—2024 年上游不同大数据算法应用频率

Fig. 7 Frequency of application of different big data algorithms in the upstream petroleum and petrochemical sector, 2014-2024

从图 7 中可以看出, 上游业务中, 随机森林、BP 神经网络、灰色预测、支持向量机和遗传算法 5 种大数据算法应用频率高于所有算法(24 种)应用频率(2 次), 应用频率分别为 4 次、4 次、3 次、3 次和 3 次。这主要得益于它们在应对这些数据特征时的独特优势。

随机森林作为集成算法, 通过多棵决策树的集成, 提高了抗噪性和稳定性, 适合用于油藏分类、地质特征提取和井下异常检测等任务, 能够有效处理上游数据的高噪声和高维特性。

BP 神经网络则凭借强大的非线性建模能力, 可以捕捉复杂的地下地质条件和油藏特性之间的关系, 在储层预测、油藏模拟和生产优化方面表现优异。

灰色预测则适合处理小样本和不确定性高的数据, 常被用于短期产量预测和压力变化预测, 帮助工程师应对有限数据带来的挑战。

支持向量机因其优秀的分类和回归性能, 在油气层识别和井下数据分析中被频繁应用, 特别适合处理高维度的地质数据。

遗传算法以其强大的全局搜索和优化能力, 能够有效解决上游业务中的复杂优化问题, 如钻井路径优化和生产参数优化。

这些算法不仅提升了数据分析的精度, 还为上游业务的决策提供了重要支撑, 因此在未来有望被更加广泛地应用。相反, 像朴素贝叶斯和线性回归这样的简单算法, 对数据的非线性特征和高维度不具备足够的建模能力, 而上游数据通常具有强非线性和多维特性, 依赖这些算法容易导致误差较大。Apriori 等关联规则挖掘算法主要用于事务型数据的关联关系分析, 在石油勘探和生产过程中难以找到合适的应用场景, 缺乏对上游动态数据的适应性。

此外, 一些计算复杂度高的算法, 如 Boosting 和贝叶斯网络, 在处理大规模地质和测井数据时效率较低, 难以满足上游业务对实时性和高计算效率的要求。这些低频算法在应对上游业务的高非线性、实时性、复杂动态变化等需求上存在局限性, 因此难以在实际应用中得到广泛采纳, 逐渐被适应性更强的高频算法所取代。有的则是被复杂算法中的局部使用或者融合使用。

2.2.2 中游立项更加明确的原因讨论

石油石化中游领域涉及油气资源的运输、储存和加工, 包括管道输送、油气储存设施的管理等。中游环节注重物流管理、供应链优化、设备维护和运输风险控制。这一环节的数据相对连续且稳定, 但对时效性和安全性要求较高, 如何有效预测和控制运输过程中的风险是中游业务的核心挑战。随着传感器技术和物联网的发展, 未来中游业务中的大数据算法将朝着更加智能化、实时化和自动化的方向发展。

将 87 篇涉及石油石化大数据应用的代表性文献中, 整理出涉及中游业务文献 25 篇, 将 25 篇文献中, 被应用到的大数据算法及出现的次数进行统计(含交叉应用), 得出中游被应用到的大数据算法 19 种, 并将 19 种算法出现频率计算平均数可得算法平均应用频率, 每种算法的应用频率和所有算法的平均应用频率如图 8 所示。

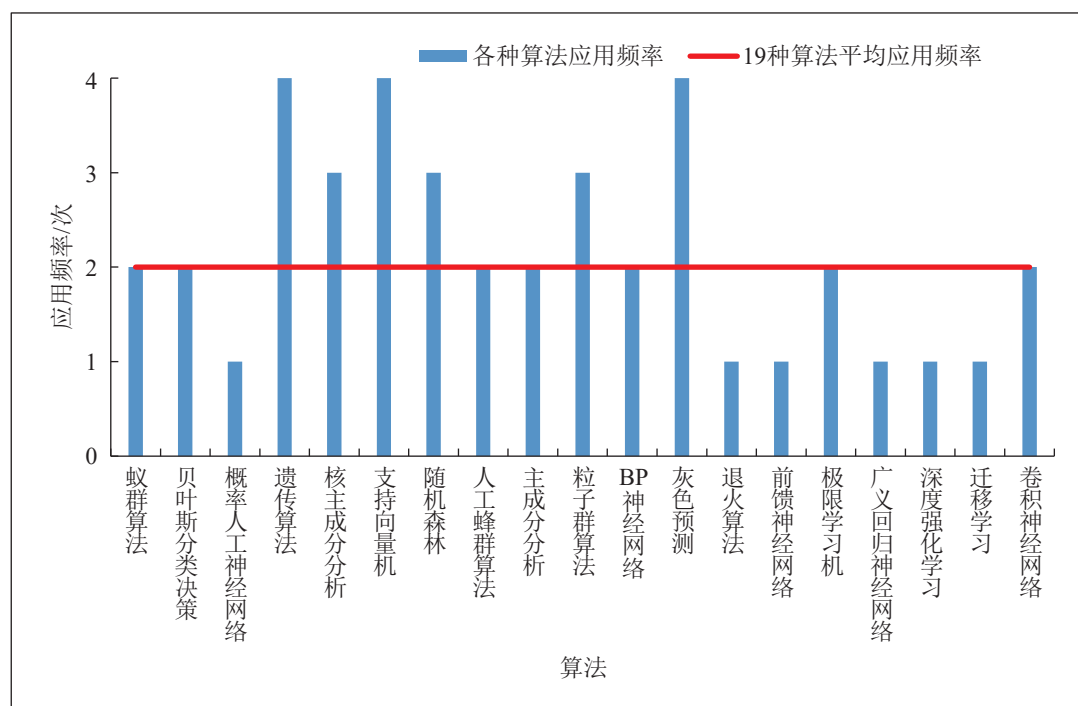


图 8 2014—2024 年中游不同大数据算法应用频率

Fig. 8 2014-2024 Frequency of application of different big data algorithms in the midstream of the petroleum and petrochemical sector

从图 8 中可以看出,在中游业务中,遗传算法、核主成分分析法、支持向量机、随机森林、粒子群算法和灰色预测算法的应用频率显著高于平均水平(2 次),应用频率分别为 4 次、4 次、4 次、3 次、3 次和 3 次。这些算法在中游业务中被广泛采用,主要是因为它们在数据处理和优化方面具备独特的优势。中游业务涉及大量管道传输、储运管理和设备监控数据,这些数据常常呈现非线性、动态变化且包含噪声。

遗传算法和粒子群算法因其强大的全局搜索和优化能力,在物流优化和管道流量预测中表现出色;核主成分分析法和支持向量机能够高效处理高维数据,适合复杂数据的降维和分类任务;随机森林则具备良好的抗噪性和解释性,能够处理多维、非线性数据,适合设备状态监控和风险预测。此外,灰色预测算法因其对不完全信息的处理能力,常用于储运环境中的短期预测任务。高频算法的选择性使用,为中游业务科研立项方向提供了明确的支撑,使研究资源更为集中,有助于提高项目的实用性和行业应用价值。

与之相比,应用低频的其他算法逐渐显现出适用性的不足,未来可能被逐步淘汰。这些低频算法在处理复杂动态数据、实时监控和优化应用时表现欠佳。例如,简单的线性回归和朴素贝叶斯模型在高维非线性数据上表现不足,缺乏对多维复杂数据的建模能力;Boosting 算法和贝叶斯网络尽管在小样本学习上有一定优势,但其计算效率低,难以满足中游业务对高效实时计算的需求。此外,一些神经网络模型(如卷积神经网络)在图像处理上表现优异,但在中游储运和管道数据的分析应用上缺乏针对性。这些算法适用场景有限,逐步被高效、鲁棒性更强的高频算法所取代,从而使科研立项方向更聚焦于真正适应行业需求的高效算法,提升研究的针对性和成果转化效率。有的则是被复杂算法中的局部使用或者融合使用。

2.2.3 下游立项更加明确的原因讨论

石油石化下游领域主要包括炼油、化工品生产和销售。下游环节的业务复杂且涉及的技术门类众多,重点在于如何通过优化炼油和化工生产工艺,实现效益最大化。此外,销售 and 市场需求预测也是下游业务的重要组成部分。与上游和中游不同,下游领域的数据更加结构化,数据量虽然庞大但更为规则。随着炼化工艺的复杂性增加以及市场环境的快速变化,算法在下游业务中的应用将更加广泛,尤其是在工艺智能化和市场

需求分析方面。

将 87 篇涉及石油石化大数据应用的代表性文献中，整理出涉及中游业务文献 15 篇，将 15 篇文献中，被应用到的大数据算法及出现的次数进行统计(含交叉应用)，得出下游应用到的大数据算法 13 种，并将 13 种算法出现频率计算平均数可得算法平均应用频率，每种算法的应用频率和所有算法平均应用频率，如图 9 所示。

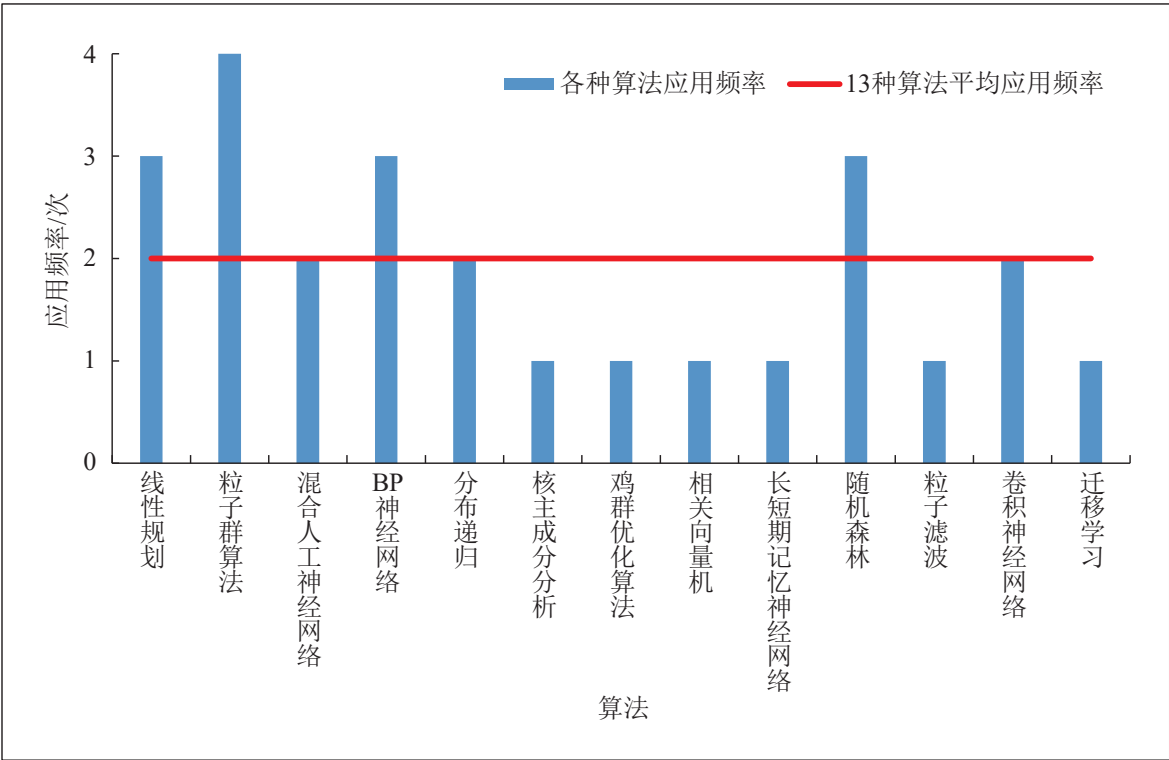


图 9 2014—2024 年下游不同大数据算法应用频率

Fig. 9 Frequency of application of different big data algorithms downstream in the petroleum and petrochemical sector, 2014-2024

从图 9 中可以看出，在下游领域，粒子群算法、线性规划、BP 神经网络和随机森林 4 种算法的应用频率较高，分别为 4 次、3 次、3 次和 3 次。表明这些算法在下游业务中得到了更广泛的应用。这主要是因为下游业务，如炼油过程优化、物流调度、市场预测和供应链管理等，涉及大量的优化和预测任务。线性规划因其在生产和物流优化中的高效性，能够解决资源配置和成本控制问题；粒子群算法在全局优化方面表现出色，适用于复杂的供应链和运输路径优化；BP 神经网络和随机森林在处理非线性数据和分类任务中具有优势，适合下游业务中的需求预测和消费者行为分析。这些高频算法符合下游业务对成本控制、市场响应和数据精细化管理的需求，使科研立项更聚焦于提升资源优化和市场分析的实际应用价值。

相反，低于平均线的其他算法在下游业务中的应用逐渐减少，主要是因为它们在实时性、计算效率和解释性方面存在不足。例如逻辑回归模型在应对复杂多变的市场数据时表现欠佳，无法满足下游业务对高精度预测的要求；卷积神经网络尽管在图像处理领域有优势，但在需求预测和物流优化等方面缺乏适应性。随着下游业务对效率和适应性的需求增长，这些低频算法的应用场景受限，将逐步被更高效、稳定性更强的算法替代。这一趋势使科研立项方向更加集中于实际业务需求和成本效益，从而推动下游领域更高效和智能化的应用创新。有的则是被复杂算法中的局部使用或者融合使用。

总体来看，无论是石油石化全领域、还是分上游、中游、下游三个领域，主要用到的大数据算法共 11 种，见表 3。

表 3 石油石化领域及上中下游适用算法对比表
Table 3 Comparison Table of Applicable Algorithms in Petroleum and Petrochemical Sector and Upstream, Midstream and Downstream

算法	全领域	上游	中游	下游
BP神经网络	★	★		★
遗传算法	★	★	★	★
随机森林	★	★	★	
人工神经网络	★			
支持向量机	★	★	★	
卷积神经网络	★			
灰色预测	★	★	★	
深度神经网络	★			
粒子群算法	★		★	★
核主成分分析			★	
线性规划				★

从表 3 可以看出，不同大数据算法在石油石化领域及子领域应用有侧重，也有交叉。BP 神经网络、遗传算法、随机森林，支持向量机、灰色预测、粒子群算法等算法在石油石化全领域及上中下游应用都普遍比较广泛，说明这些说法比较适合石油石化领域，可能成为未来主流的应用算法。而其他算法，根据其算法特性，表现出在不同子领域应用的独特性，后期在其适应的子领域可能又较好的发展和表现。未来，组合算法或将成为解决石油石化领域难题的主流趋势。

3 结论建议

(1)通过对石油石化领域大数据算法应用的深入分析和发展趋势预测，系统地梳理了上游、中游和下游各环节中大数据算法的适用性和发展潜力，为石油石化行业在不同业务阶段选择合适的算法提供了更聚焦的参考。通过明确高频算法和低频算法的优劣势，文章为上中下游各环节的大数据算法选择提供了指导，从而为石油石化领域的科研立项方向指明了更具实际价值的路径。

(2)尽管选取了近十年石油石化领域大数据应用的代表性文献进行分析，并对未来算法的发展方向进行了预测，但在文献收集过程中，可能存在遗漏现象，还有一些文献是融合了多种算法，部分预测也可能因数据局限而存在误差，分析结果应综合参考。

(3)展望未来，算法融合将成为石油石化大数据应用的主要发展趋势。无论是上游的勘探预测、中游的储运优化，还是下游的市场分析，组合多种算法以提高精度和可靠性将是大数据技术应用的关键方向，从而进一步推动石油石化行业的数字化和智能化升级。

致谢

感谢国家管网集团北方管道有限责任公司在研究工作中的大力支持，并同意将研究成果公开发表。

稿件申明

作者排名已征得全部作者的同意，学术观点方面全部作者的观点均一致，涉及的专利已经征得所有发明人的同意且已签字确认，发表的论文不涉及企业技术商业秘密，专家审稿无要求避讳问题，文章涉及的内容不存在学术不端的问题。

署名贡献声明

李荣光负责数据整理，文章策划及方向确定。
金龙负责文章方法设计和数据分析及文档整理。

孙伶负责初稿撰写, 文献核定。

赵俊淇负责方法科学性检测。

陈斯迅完成测算结果绘图及调整。

郑力会负责初稿的审阅和修改。

利益冲突说明

论文不涉及企业技术商业秘密。

专家审稿时没有需要避讳的问题。

数据可用性声明

本研究中涉及的所有原始数据、分析脚本及研究过程中产生的其他相关数据资料, 均已妥善保存并可通过合理请求获得。为确保数据的隐私与安全, 以及遵守相关伦理与法律要求, 部分敏感或个人信息已进行脱敏处理。研究者鼓励并欢迎其他学者基于本研究数据进行进一步的探索与分析。如有数据访问需求, 请通过电子邮件联系本研究的通信作者, 并在邮件中明确说明数据使用目的、研究计划以及预期的研究成果。我们将在收到请求后的合理时间内, 根据数据共享政策与程序, 对请求进行评估并作出回应。

参考文献

- [1] 钱慧芳, 罗卉, 王玉鑫. 基于 BP 网络的石油套管破损检测算法 [J]. [西安工程大学学报](#), 2014(1): 84-88. //QIAN Huifang, LUO Hui, WANG Yuxin. Study on oil casing damage detection algorithm based-on BP network [J]. [Journal of Xi'an Polytechnic University](#), 2014(1): 84-88.
- [2] 李明鑫, 王岳, 张月, 等. 基于蚁群算法的油气管道预警系统优化研究 [J]. [石油化工自动化](#), 2014, 50(1): 23-25. //LI Mingxin, WANG Yue, ZHANG Yue, et al. Optimization study of oil and gas pipeline early alarming system based on ant colony algorithm [J]. [Automation in Petro-Chemical Industry](#), 2014, 50(1): 23-25.
- [3] 金倬伊. RPMS 优化软件在炼化企业的应用研究 [J]. [广州化工](#), 2014, 42(10): 189-190,194. //JIN Zhuoyi. Research on the applications of RPMS optimization software in petrochemical enterprises [J]. [Guangzhou Chemical Industry](#), 2014, 42(10): 189-190,194.
- [4] 马林茂, 李德富, 郭海湘, 等. 基于遗传算法优化 BP 神经网络在原油产量预测中的应用: 以大庆油田 BED 试验区为例 [J]. [数学的实践与认识](#), 2015, 45(24): 117-128. //MA Linmao, LI Defu, GUO Haixiang, et al. BP neural network based on genetic algorithm applied in crude oil production forecast: Taking the BED test area of the daqing oilfield as an example [J]. [Mathematics in Practice and Theory](#), 2015, 45(24): 117-128.
- [5] 骆正山, 蒋丽云. 基于概率神经网络的油气管道的可靠性评估 [J]. [消防科学与技术](#), 2015(11): 1517-1520. //LUO Zhengshan, JIANG Liyun. Reliability assessment of oil and gas pipelines based on probabilistic neural network [J]. [Fire Science and Technology](#), 2015(11): 1517-1520.
- [6] 田亚鹏, 鞠斌山. 基于遗传算法改进 BP 神经网络的页岩气产量递减预测模型 [J]. [中国科技论文](#), 2016, 11(15): 1710-1715. //TIAN Yapeng, JU Binshan. A model for predicting shale gas production decline based on the BP neural network improved by the genetic algorithm [J]. [China Sciencepaper](#), 2016, 11(15): 1710-1715.
- [7] 邱莹莹, 叶贞成, 赵亮, 等. 基础油供应链生产-分销计划模型及其优化 [J]. [化工进展](#), 2016, 35(3): 711-716. //QIU Yingying, YE Zhencheng, ZHAO Liang, et al. Production-distribution planning model and optimization of base oil supply chain [J]. [Chemical Industry and Engineering Progress](#), 2016, 35(3): 711-716.
- [8] 刘佳佳, 范利军, 田群宏, 等. 基于遗传算法的 CO₂ 管道运输优化设计研究 [J]. [石油工程建设](#), 2017, 43(5): 14-18. //LIU Jiajia, FAN Lijun, TIAN Qunhong, et al. Optimization design of CO₂ pipeline transportation based on genetic algorithm [J]. [Petroleum Engineering Construction](#), 2017, 43(5): 14-18.
- [9] LI Z L, GHANG H F, TAN D J, et al. A novel acoustic emission detection module for leakage recognition in a gas pipeline valve [J]. [Process Safety and Environmental Protection](#), 2017, 105: 32-40.
- [10] 陈一诺, 董绍华. 基于随机森林算法的管道缺陷预测方法 [J]. [油气储运](#), 2018, 37(9): 975-979. //CHEN Yinuo, DONG Shaohua. Pipeline defect prediction method based on Random Forests algorithm [J]. [Oil & Gas Storage and Transportation](#), 2018, 37(9): 975-979.
- [11] 刘胜娃, 孙俊明, 高翔, 等. 基于人工神经网络的钻井机械钻速预测模型的分析与建立 [J]. [计算机科学](#), 2019, 46(S1):

- 605-608. //LIU Shengwa, SUN Junming, GAO Xiang, et al. Analysis and establishment of drilling speed prediction model for drilling machinery based on artificial neural networks [J]. *Computer Science*, 2019, 46(S1): 605-608.
- [12] 周游, 张广智, 高刚, 等. 核主成分分析法在测井浊积岩岩性识别中的应用 [J]. *石油地球物理勘探*, 2019, 54(3): 667-675. //ZHOU You, ZHANG Guangzhi, GAO Gang, et al. Application of kernel principal component analysis in well logging turbidite lithology identification [J]. *Oil Geophysical Prospecting*, 2019, 54(3): 667-675.
- [13] 陈小龙, 李宜强, 管错, 等. 基于量纲分析的优化神经网络模型预测 GAGD 非混相开发油藏采收率 [J]. *石油科学通报*, 2019, 4(3): 288-299. //CHEN Xiaolong, LI Yiqiang, GUAN Cuo, et al. An optimized neural network prediction model for gas assisted gravity drainage recovery based on dimensional analysis [J]. *Petroleum Science Bulletin*, 2019, 4(3): 288-299.
- [14] 刘彪, 李宸晓, 李双贵, 等. 基于支持向量回归的井漏预测 [J]. *钻采工艺*, 2019, 42(6): 17-20. //LIU Biao, LI Chengxiao, LI Shuangui, et al. Lost circulation prediction based on support vector regression [J]. *Drilling & Production Technology*, 2019, 42(6): 17-20.
- [15] 黄知娟, 潘丽娟, 路辉, 等. 大数据分析顺北油田 SHB-X 井试采产液量骤降原因 [J]. *石油钻采工艺*, 2019, 41(3): 341-347. //HUANG Zhijuan, PAN Lijuan, LU Hui, et al. The reasons for sudden production drop by big data analysis in trial production for Well SHB-X in Shunbei oilfield [J]. *Oil Drilling & Production Technology*, 2019, 41(3): 341-347.
- [16] 艾月乔, 徐亮, 陶江华, 等. 基于人工蜂群算法优化的 SVM 管道风险评估 [J]. *油气储运*, 2019, 38(5): 510-515. //AI Yueqiao, XU Liang, TAO Jianghua, et al. An optimized SVM pipeline risk assessment method based on artificial bee colony algorithm [J]. *Oil & Gas Storage and Transportation*, 2019, 38(5): 510-515.
- [17] 马钢, 李俊飞, 白瑞, 等. 基于 PCA-SVR 模型的油气管道剩余强度预测 [J]. *油气储运*, 2019, 38(10): 1119-1124. //MA Gang, LI Junfei, BAI Rui, et al. Prediction on the residual strength of oil and gas pipelines based on PCA-SVR model [J]. *Oil & Gas Storage and Transportation*, 2019, 38(10): 1119-1124.
- [18] 曹晓红, 陆鹏飞, 周业永, 等. 混合建模技术在石化装置异常工况预警中的应用 [J]. *计算机与应用化学*, 2019, 36(5): 534-538. //CAO Xiaohong, LU Pengfei, ZHOU Yeyong, et al. Application of hybrid modeling technology in early warning of abnormal working conditions in petrochemical plants [J]. *Computers and Applied Chemistry*, 2019, 36(5): 534-538.
- [19] KIM M, SHIN H. Machine learning-based prediction of the shale barrier size and spatial location using key features of SAGD production curves [J]. *Journal of Petroleum Science and Engineering*, 2020, 191: 107205.
- [20] CHEN W, YANG L Q, ZHA B, et al. Deep learning reservoir porosity prediction based on multilayer long short-term memory network [J]. *Geophysics*, 2020, 85(4): 213-225.
- [21] 贾德利, 刘合, 张吉群, 等. 大数据驱动下的老油田精细注水优化方法 [J]. *石油勘探与开发*, 2020, 47(3): 629-636. //JIA Deli, LIU He, ZHANG Jiqun, et al. Data-driven optimization for fine water injection in a mature oil field [J]. *Petroleum Exploration and Development*, 2020, 47(3): 629-636.
- [22] 王洪亮, 穆龙新, 时付更, 等. 基于循环神经网络的油田特高含水期产量预测方法 [J]. *石油勘探与开发*, 2020, 47(5): 1009-1015. //WANG Hongliang, MU Longxin, SHI Fugeng, et al. Production prediction at ultra-high water cut stage via Recurrent Neural Network [J]. *Petroleum Exploration and Development*, 2020, 47(5): 1009-1015.
- [23] 张军华, 任雄风, 赵杰, 等. 基于交叉验证支持向量机储层预测方法及应用 [J]. *科学技术与工程*, 2020, 20(13): 5052-5057. //ZHANG Junhua, REN Xiongfeng, ZHAO Jie, et al. Reservoir prediction method and its application of support vector machine based on cross validation [J]. *Science Technology and Engineering*, 2020, 20(13): 5052-5057.
- [24] 徐磊, 侯磊, 李雨, 等. 基于机器学习的原油管输能耗预测方法研究 [J]. *石油科学通报*, 2020, 5(4): 567-577. //XU Lei, HOU Lei, LI Yu, et al. Research into prediction of energy consumption of crude oil pipelines based on machine learning [J]. *Petroleum Science Bulletin*, 2020, 5(4): 567-577.
- [25] 郑度奎, 李昊然, 程远鹏, 等. 油气管道的 CO₂ 腐蚀预测模型和预测方法 [J]. *腐蚀与防护*, 2020, 41(3): 48-53, 58. //ZHENG Dukui, LI Haoran, CHENG Yuanpeng, et al. CO₂ corrosion prediction model and prediction method for oil and gas pipeline [J]. *Corrosion & Protection*, 2020, 41(3): 48-53, 58.
- [26] 贺宗江. 工业大数据技术在石化设备预警预测中的研究与实践 [J]. *当代石油石化*, 2020, 28(6): 35-40. //HE Zongjiang. Study and practice of industrial big data technology in petrochemical equipment early warning and prediction [J]. *Petroleum & Petrochemical Today*, 2020, 28(6): 35-40.
- [27] 韩丰磊, 李梦雨, 刘杰夫, 等. 基于 BP 神经网络的煤直接液化装置 VOCs 包袋法检测与预测建模 [J]. *化工进展*, 2020, 39(10): 3949-3955. //HAN Fenglei, LI Mengyu, LIU Jiefu, et al. Method of checking and accounting for the bag method of VOCs in coal direct liquefaction unit based on BP neural network [J]. *Chemical Industry and Engineering Progress*, 2020, 39(10): 3949-3955.

- [28] 危拓. 炼化一体化 MPIMS 模型的建设及应用 [J]. *石油化工技术与经济*, 2020, 36(5): 45-48. //WEI Tuo. The construction and application of MPIMS model for refining and chemical integration [J]. *Techno-Economics in Petrochemicals*, 2020, 36(5): 45-48.
- [29] 杨午阳, 杨佳润, 陈双全, 等. 基于 U-Net 深度学习网络的地震数据断层检测 [J]. *石油地球物理勘探*, 2021, 56(4): 688-697. //YANG Wuyang, YANG Jiarun, CHEN Shuangquan, et al. Seismic data fault detection based on U-Net deep learning network [J]. *Oil Geophysical Prospecting*, 2021, 56(4): 688-697.
- [30] 郑力会, 徐燕东, 邱子瑶, 等. 完整作业信息定量分析顺北油田钻完井漏失因素 [J]. *油气藏评价与开发*, 2021, 11(4): 597-604. //ZHENG Lihui, XU Yandong, QIU Ziyao, et al. Quantitative analysis on drilling and completion loss factors by all data in Shunbei Oilfield [J]. *Reservoir Evaluation and Development*, 2021, 11(4): 597-604.
- [31] 田高鹏, 林年添, 张凯, 等. 多波地震油气储层的自组织神经网络学习与预测 [J]. *科学技术与工程*, 2021, 21(19): 7931-7941. //TIAN Gaopeng, LIN Niantian, ZHANG Kai, et al. Prediction of seismic oil and gas reservoir using self-organizing neural network from multi-component seismic data [J]. *Science Technology and Engineering*, 2021, 21(19): 7931-7941.
- [32] 谷宇峰, 张道勇, 鲍志东, 等. 利用 GS-LightGBM 机器学习模型识别致密砂岩地层岩性 [J]. *地质科技通报*, 2021, 40(4): 224-234. //GU Yufeng, ZHANG Daoyong, BAO Zhidong, et al. Lithology prediction of tight sandstone formation using GS-LightGBM hybrid machine learning model [J]. *Bulletin of Geological Science and Technology*, 2021, 40(4): 224-234.
- [33] PIRIZADEH M, ALEMOHAMMAD N, MANTHOURI M, et al. A new machine learning ensemble model for class imbalance problem of screening enhanced oil recovery methods [J]. *Journal of Petroleum Science and Engineering*, 2021, 198: 108214.
- [34] CHERAGHI Y, KORD S, MASHAYEKHIZADEH V. Application of machine learning techniques for selecting the most suitable enhanced oil recovery method: challenges and opportunities [J]. *Journal of Petroleum Science and Engineering*, 2021, 205: 108761.
- [35] WANG Y W, LIU H Q, GUO M Z, et al. Image recognition model based on deep learning for remaining oil recognition from visualization experiment [J]. *Fuel*, 2021, 291: 120216.
- [36] 何佑伟, 贺质越, 汤勇, 等. 基于机器学习的页岩气井产量评价与预测 [J]. *石油钻采工艺*, 2021, 43(4): 518-524. //HE Youwei, HE Zhiyue, TANG Yong, et al. Shale gas well production evaluation and prediction based on machine learning [J]. *Oil Drilling & Production Technology*, 2021, 43(4): 518-524.
- [37] 李谦, 曹彦伟, 朱海燕. 基于人工智能的钻速预测模型数据有效性下限分析 [J]. *钻探工程*, 2021, 48(3): 21-30. //LI Qian, CAO Yanwei, ZHU Haiyan. Discussion on the lower limit of data validity for ROP prediction based on artificial intelligence [J]. *Drilling Engineering*, 2021, 48(3): 21-30.
- [38] 陶杉, 余星, 宋海, 等. 大数据方法寻找顺北碳酸盐岩储层开采过程中井壁坍塌主控因素 [J]. *石油钻采工艺*, 2020, 42(5): 627-631. //TAO Shan, YU Xing, SONG Hai, et al. Application of the big data method to search for the main factors controlling the hole collapse in the production process of Shunbei carbonate reservoir [J]. *Oil Drilling & Production Technology*, 2020, 42(5): 627-631.
- [39] 曹继飞, 邹德永, 舒腾飞, 等. 基于主成分分析的碳酸盐岩裂缝识别方法 [J]. *科学技术与工程*, 2021, 21(36): 15421-15428. //CAO Jifei, ZOU Deyong, SHU Tengfei, et al. Fracture identification in carbonate rocks based on principle component analysis [J]. *Science Technology and Engineering*, 2021, 21(36): 15421-15428.
- [40] 马梦桐, 赵琢. 基于 KPCA-CSO-RVM 模型的工艺管道腐蚀速率预测 [J]. *安全与环境工程*, 2021, 28(4): 1-7, 20. //MA Mengtong, ZHAO Zhuo. Prediction of corrosion rate of process pipelines based on KPCA-CSO-RVM model [J]. *Safety and Environmental Engineering*, 2021, 28(4): 1-7, 20.
- [41] 李绍阳, 曹萃文. 基于 LSTM 的润滑油加氢装置产品预测和基于 RF 的操作优化相关性分析 [J]. *石油学报 (石油加工)*, 2021, 37(5): 1040-1049. //LI Zhaoyang, CAO Cuiwen. Product prediction of lubricating oil hydrogenation unit based on LSTM and optimal operation correlation analysis with RF [J]. *Acta Petrolei Sinica (Petroleum Processing Section)*, 2021, 37(5): 1040-1049.
- [42] FANG X X, FENG H, WANG H. Study on intelligent prediction method of rock drillability based on Bayesian lithology classification and optimized BP neural network [J]. *Petroleum Science and Technology*, 2022, 40(17): 2141-2162.
- [43] WEI J, LIAO H L, WANG H J, et al. Experimental investigation on the dynamic tensile characteristics of conglomerate based on 3D SHPB system [J]. *Journal of Petroleum Science and Engineering*, 2022, 213: 110350.
- [44] 王泽龙, 刘先贵, 李志勇, 等. 基于贝叶斯理论的四维地震与生产动态数据的油藏历史拟合新方法 [J]. *石油科技论坛*,

- 2022, 41(2): 9-22. //WANG Zelong, LIU Xiangui, LI Zhiyong, et al. A novel bayesian-based reservoir history matching method for 4D seismic and production data [J]. *Oil Forum*, 2022, 41(2): 9-22.
- [45] LOOMBA A K, BOTECHI A V, SCHIOZER D J. Cluster-based learning and evolutionary gorithm for optimization [J]. *Geoenergy Science and Engineering*, 2023, 227: 211801.
- [46] 侯亚伟, 刘超, 徐中波, 等. 多层水驱开发油田采收率快速预测方法 [J]. *石油钻探技术*, 2022, 50(5): 82-87. //HOU Ya-wei, LIU Chao, XU Zhongbo, et al. A method for rapidly predicting recovery of multi-layer oilfields developed by Water-Flooding [J]. *Petroleum Drilling Techniques*, 2022, 50(5): 82-87.
- [47] 肖阳, 王家豪, 李志刚, 等. 基于大数据的页岩油区块产量差异分析方法研究 [J]. *钻采工艺*, 2022, 45(3): 73-78. //XIAO Yang, WANG Jiahao, LI Zhigang, et al. Study on production difference analysis method of shale oil play based on big data [J]. *Drilling & Production Technology*, 2022, 45(3): 73-78.
- [48] 薛永超, 袁志乾, 金青爽, 等. 基于深度森林算法的油井产量预测 [J]. *科学技术与工程*, 2022, 22(11): 4327-4334. //XUE Yongchao, YUAN Zhiqian, JIN Qingshuang, et al. Production prediction of oil well based on deep forest algorithm [J]. *Science Technology and Engineering*, 2022, 22(11): 4327-4334.
- [49] 王兵, 黄丹, 李文璟. 基于支持度矩阵 Apriori 算法的钻井隐患关联挖掘 [J]. *西南石油大学学报 (自然科学版)*, 2022, 44(2): 113-122. //WANG Bing, HUANG Dan, LI Wenjing. Correlation mining of hidden hazards in drilling based on support matrix apriori algorithm [J]. *Journal of Southwest Petroleum University(Science & Technology Edition)*, 2022, 44(2): 113-122.
- [50] 姜雪, 熊志武. 基于灰色理论指导的储层构型半定量表征及优质储层预测——以西湖凹陷 A 构造花港组 H3 砂层组为例 [J]. *海洋地质与第四纪地质*, 2022, 42(6): 162-172. //JIANG Xue, XIONG Zhiwu. Semi-quantitative study on reservoir configuration in grey theory—a case study of H3 sand unit of Huagang Formation in A Structure, Xihu Sag [J]. *Marine Geology & Quaternary Geology*, 2022, 42(6): 162-172.
- [51] 李超, 刘忠良, 金银霞, 等. 基于粒子滤波算法的涡旋压缩机性能预测研究 [J]. *流体机械*, 2022, 50(5): 54-61. //LI Chao, LIU Zhongliang, JIN Yinxia, et al. Performance prediction analysis of scroll compressor based on particle filter algorithm [J]. *Fluid Machinery*, 2022, 50(5): 54-61.
- [52] 董丰莲, 魏志伟, 孙鑫, 等. 石油产业链一体化优化模型系统的开发与应用 [J]. *油气与新能源*, 2022, 34(4): 87-92, 98. //DONG Fenglian, WEI Zhiwei, SUN Xin, et al. Development and application of the optimization model system regarding the integration of the petroleum industry Chain [J]. *Petroleum and New Energy*, 2022, 34(4): 87-92, 98.
- [53] 潘艳秋, 李鹏飞, 高石磊, 等. 基于 MIC 筛选规则和 BP 神经网络的变换装置建模及产品预测 [J]. *化工进展*, 2022, 41(S1): 36-43. //PAN Yanqiu, LI Pengfei, GAO Shilei, et al. Application of BP neural network based on MIC screening rules in modeling and product prediction of shift unit [J]. *Chemical Industry and Engineering Progress*, 2022, 41(S1): 36-43.
- [54] 祁滢, 李峻, 郭以东. 基于神经网络的能耗预测在油气田企业的应用 [J]. *信息系统工程*, 2022(5): 60-63. //QI Ying, LI Jun, GUO Yidong. Application of neural network-based energy consumption prediction in oil and gas field enterprises [J]. *China CIO News*, 2022(5): 60-63.
- [55] 肖荣鸽, 靳帅帅, 庄琦, 等. 基于灰色理论的油气管道腐蚀速率预测 [J]. *热加工工艺*, 2022, 51(18): 53-57. //XIAO Rongge, JIN Shuaishuai, ZHUANG Qi, et al. Prediction of corrosion rate of oil and gas pipeline based on grey theory [J]. *Hot Working Technology*, 2022, 51(18): 53-57.
- [56] 鲁中歧, 肖文生, 崔俊国, 等. 基于参数优化神经网络的海底油气管道腐蚀泄漏预测 [J]. *科学技术与工程*, 2022, 22(20): 8673-8682. //LU Zhongqi, XIAO Wensheng, CUI Junguo, et al. Corrosion leakage prediction of submarine oil and gas pipeline based on neural network for parameter optimization [J]. *Science Technology and Engineering*, 2022, 22(20): 8673-8682.
- [57] 张新生, 常潞戈. 基于 FA-BAS-ELM 的海洋油气管道外腐蚀速率预测 [J]. *中国安全科学学报*, 2022, 32(2): 99-106. //ZHANG Xinsheng, CHANG Yingge. Prediction of external corrosion rate of offshore oil and gas pipelines based on FA-BAS-ELM [J]. *China Safety Science Journal*, 2022, 32(2): 99-106.
- [58] 李明, 闻伟, 李秉军, 等. 基于 ICS-GRNN 的油气管道剩余强度预测技术 [J]. *焊管*, 2022, 45(1): 32-36. //LI Ming, WEN Wei, LI Bingjun, et al. Residual strength prediction technology of oil and gas pipeline based on ICS-GRNN [J]. *Welded Pipe and Tube*, 2022, 45(1): 32-36.
- [59] 孔令维, 许立红, 刘树辉. 基于大数据分析的机采井组最优产能模型的确定与应用 [J]. *石油石化节能*, 2023, 13(5): 49-53. //KONG Lingwei, XU Lihong, LIU Shuhui. Determination and application of optimal capacity model of machine production well group based on big data analysis [J]. *Energy Conservation in Petroleum & Petrochemical Industry*, 2023, 13(5):

- 49-53.
- [60] 蒲堡萍, 魏建光, 周晓峰, 等. 基于机器学习的低渗透砂岩聚合物驱采收率预测 [J]. *科学技术与工程*, 2023, 23(28): 12045-12056. //PU Baoping, WEI Jianguang, ZHOU Xiaofeng, et al. Predicting the recovery factor of polymer flooding in low-permeability sandstone based on machine learning [J]. *Science Technology and Engineering*, 2023, 23(28): 12045-12056.
- [61] 罗双平, 刘青, 韩巍, 等. 基于井筒工程数据的钻井提速评价分析方法——以高石梯—磨溪区块为例 [J]. *天然气技术与经济*, 2023, 17(6): 21-28. //LUO Shuangping, LIU Qing, HAN Wei, et al. Evaluation on ROP improvement based on well-bore engineering data: an example from Gaoshiti-Moxi block [J]. *Natural Gas Technology and Economy*, 2023, 17(6): 21-28.
- [62] 郭秋麟, 任洪佳, 于京都, 等. 基于贝叶斯网络的油气勘探风险预测方法——以准噶尔盆地腹部侏罗系三工河组为例 [J]. *中国石油勘探*, 2023, 28(1): 108-119. //GUO Qiulin, REN Hongjia, YU Jingdou, et al. Prediction method of petroleum exploration risks based on Bayesian network: a case study of the Jurassic Sangonghe Formation in the hinterland of Junggar Basin [J]. *China Petroleum Exploration*, 2023, 28(1): 108-119.
- [63] 邓增利, 吴南旭, 鲁小琴. 基于深度学习的安全帽识别在海上油田智能化中应用 [J]. *电工技术*, 2023(21): 45-48. //DENG Zengli, WU Nanxu, LU Xiaoqin. Application of helmet recognition based on deep learning in intelligent safety monitoring of offshore platforms [J]. *Electric Engineering*, 2023(21): 45-48.
- [64] 盛科鸣, 蒋官澄. 基于随机森林算法的油气层敏感性损害预测 [J]. *钻井液与完井液*, 2023, 40(4): 423-430. //SHENG Keming, JIANG Guancheng. Prediction of four kinds of sensibility damages to hydrocarbon reservoirs based on random forest algorithm [J]. *Drilling Fluid & Completion Fluid*, 2023, 40(4): 423-430.
- [65] 刘鸣慧, 熊建斌, 苏乃权, 等. 基于深度学习的石化机组轴承故障诊断综述 [J]. *机床与液压*, 2023, 51(6): 171-180. //LIU Minghui, XIONG Jianbin, SU Naiquan, et al. Review of bearing fault diagnosis of petrochemical units based on deep learning [J]. *Machine Tool & Hydraulics*, 2023, 51(6): 171-180.
- [66] 冯博. 基于大数据聚类算法石油管道泄漏检测技术探究 [J]. *传感器世界*, 2023, 29(7): 23-26,32. //FENG Bo. Research on oil pipeline leakage detection technology based on large cluster analysis algorithm [J]. *Sensor World*, 2023, 29(7): 23-26,32.
- [67] 杜宇晨, 张新生. 基于深度强化学习的腐蚀油气管道预防性维修策略研究 [J]. *经营与管理*, 2024(5): 73-80. //DU Yuchen, ZHANG Xinsheng. Research on preventive maintenance strategies for corroded oil and gas pipelines based on deep reinforcement learning [J]. *Management and Administration*, 2024(5): 73-80.
- [68] 闫绍栋, 刘斌, 李卫东, 等. 基于 BP 神经网络的油气管道振动检测系统 [J]. *化工设备与管道*, 2023, 60(5): 68-75. //YAN Shaodong, LIU Bin, LI Weidong, et al. Oil and gas pipeline vibration detection system based on BP neural network [J]. *Process Equipment & Piping*, 2023, 60(5): 68-75.
- [69] 赵年峰, 夏艳权, 崔凯强, 等. 基于优化 GM(1, N) 模型的油气管道腐蚀速率预测 [J]. *焊管*, 2023, 46(8): 38-44. //ZHAO Nianfeng, XIA Yanquan, CUI Kaiqiang, et al. Corrosion rate prediction of oil and gas pipeline based on optimized GM(1, N) model [J]. *Welded Pipe and Tube*, 2023, 46(8): 38-44.
- [70] 张力凡, 魏航信. 基于深度学习神经网络的油气管道漏磁检测缺陷诊断 [J]. *机电工程技术*, 2023, 52(1): 260-266. //ZHANG Lifan, WEI Hangxin. Defect diagnosis of magnetic flux leakage testing for oil and gas pipeline based on deep learning neural network [J]. *Mechanical & Electrical Engineering Technology*, 2023, 52(1): 260-266.
- [71] 柴天祥, 樊建春, 赵宏展, 等. 基于数据挖掘的油气勘探开发事故分析 [J]. *云南化工*, 2024, 51(4): 178-182. //CHAI Tianxiang, FAN Jianchun, ZHAO Hongzhan, et al. Oil and gas exploration and development accident analysis based on data mining [J]. *Yunnan Chemical Technology*, 2024, 51(4): 178-182.
- [72] 蒲先渤, 李泽群, 尹飞, 等. 基于 PCA-LM-BP 神经网络的岩石可钻性预测研究 [J]. *钻探工程*, 2023, 50(6): 63-68. //PU Xianbo, LI Zequn, YIN Fei, et al. Research on rock drill ability prediction based on PCA-LM-BP neural network [J]. *Drilling Engineering*, 2023, 50(6): 63-68.
- [73] 杨子杰, 陈冬霞, 王翹楚, 等. 基于人工神经网络方法预测油气资源丰度——以渤海湾盆地东濮凹陷文留地区古近系沙河街组三段为例 [J]. *石油实验地质*, 2024, 46(2): 428-440. //YANG Zijie, CHEN Dongxia, WANG Qiaochu, et al. Prediction of petroleum resource abundance based on artificial neural network method: a case study of third member of Paleogene Shahejie Formation in Wenliu area of Dongpu Sag, Bohai Bay Basin [J]. *Petroleum Geology & Experiment*, 2024, 46(2): 428-440.
- [74] 田龙, 朱智华, 王立伟, 等. 基于大数据和无监督聚类算法的岩石可钻性表征和预测方法 [J]. *新疆石油天然气*, 2024,

- 20(2): 29-36. //TIAN Long, ZHU Zhihua, WANG Liwei, et al. A rock drillability characterization method based on big data and unsupervised clustering algorithm [J]. *Xinjiang Oil & Gas*, 2024, 20(2): 29-36.
- [75] 白生勇. 基于大数据的抽油机井硬件匹配技术研究 [J]. *石油石化节能与计量*, 2024, 14(5): 21-24. //BAI Shengyong. Research on hardware matching technology of pumping wells based on big data [J]. *Energy Conservation and Measurement in Petroleum & Petrochemical Industry*, 2024, 14(5): 21-24.
- [76] 何旭晟, 周井红, 管桐, 等. 地质工程原生数据预测深层小尺度裂缝性地层漏失特征 [J]. *石油钻采工艺*, 2024, 46(1): 33-44. //HE Xusheng, ZHOU Jinghong, GUAN Tong, et al. Prediction of loss characteristics in deep small-scale fractured formations using geological engineering native data [J]. *Oil Drilling & Production Technology*, 2024, 46(1): 33-44.
- [77] 刘明亮. 优化遗传算法在天然气管网流量分配过程模拟计算中的应用 [J]. *计量学报*, 2024, 45(6): 848-856. //LIU Mingliang. Application of optimization genetic algorithm in the simulation calculation of flow distribution process of natural gas pipeline network [J]. *Acta Metrologica Sinica*, 2024, 45(6): 848-856.
- [78] 苏乾潇, 乔德新, 任义丽, 等. 基于傅里叶卷积的电成像测井图像修复 [J/OL]. *北京航空航天大学学报*, 2024-02-04, [2024-08-16]. //SU Qianxia, QIAO Dexin, REN Yili, et al. Inpainting of blank strips in imaging logging images based on Fourier convolution [J/OL]. *Journal of Beijing University of Aeronautics and Astronautics*, 2024-02-04, [2024-08-16].
- [79] 张静, 徐春亮, 马彦涛, 等. 基于模拟退火-粒子群算法的输油管道泄漏定位技术 [J]. *石油工程建设*, 2021, 47(4): 64-69. //ZHANG Jing, XU Chunliang, MA Yantao, et al. Oil pipeline leak location technology based on SA-PSO algorithm [J]. *Petroleum Engineering Construction*, 2021, 47(4): 64-69.
- [80] 秦谢勋, 刘文彬, 陈良超. 基于改进蜂群算法和灰色模型的管道腐蚀预测 [J]. *北京化工大学学报(自然科学版)*, 2021, 48(1): 74-80. //QIN Xiexun, LIU Wenbin, CHEN Liangchao. Pipeline corrosion prediction based on an improved artificial bee colony algorithm and a grey model [J]. *Journal of Beijing University of Chemical Technology(Natural Science Edition)*, 2021, 48(1): 74-80.
- [81] 王东海, 孟彤, 崔媛智. 基于遗传算法的不规则天然气管道支管件位置避振优化 [J]. *石化技术*, 2024, 31(7): 137-139. //WANG Donghai, MENG Tong, CUI Yuanzhi. The optimization of position vibration avoidance for irregular natural gas pipeline supports based on genetic algorithm [J]. *Petrochemical Industry Technology*, 2024, 31(7): 137-139.
- [82] 罗仁泽, 王磊. 基于深度学习模型的油气管道焊缝缺陷智能识别 [J]. *天然气工业*, 2024, 44(9): 199-208. //LUO Renze, WANG Lei. Weld defect intelligent identification for oil and gas pipelines based on the deep learning models [J]. *Natural Gas Industry*, 2024, 44(9): 199-208.
- [83] 范蓬勃, 张新生. 基于 KPCA-SOA-KELM 的海底油气管道内腐蚀速率预测 [J/OL]. *热加工工艺*, 1-6 [2024-09-11]. <https://doi.org/10.14158/j.cnki.1001-3814.20223107>. //FAN Pengbo, ZHANG Xinsheng. Prediction of internal corrosion rate of subsea oil and gas pipeline based on KPCA-SOA-KELM [J/OL]. *Hot Working Technology*, 1-6 [2024-09-11]. <https://doi.org/10.14158/j.cnki.1001-3814.20223107>.
- [84] 黄书童, 贾晓丽. 基于 PCC-LPSO-BP 的油气管道缺陷分类识别研究 [J]. *石油矿场机械*, 2024, 53(4): 1-9. //HUANG Shutong, JIA Xiaoli. Research on classification and identification of oil and gas pipeline defects based on PCC-LPSO-BP [J]. *Oil Field Equipment*, 2024, 53(4): 1-9.
- [85] 王栋, 邹亚飞, 耿慧晶. 基于灰色系统理论的油气管道腐蚀速率预测 [J]. *全面腐蚀控制*, 2024, 38(6): 162-167. //WANG Dong, ZOU Yafei, GENG Huijing. Prediction of corrosion rate of oil and gas pipelines based on grey system theory [J]. *Total Corrosion Control*, 2024, 38(6): 162-167.
- [86] 赵晓弘, 胡西娅, 纪博睿, 等. 人工神经网络在石油化工中的应用研究进展 [J]. *当代化工*, 2024, 53(5): 1228-1232. //ZHAO Xiaohong, HU Xiya, JI Borui, et al. Research progress in application of artificial neural network in petrochemical industry [J]. *Contemporary Chemical Industry*, 2024, 53(5): 1228-1232.
- [87] 冉安仙, 吴恋, 王丰茂, 等. 基于机器视觉的石油化工场景中的危险行为检测与预警系统 [J]. *物联网技术*, 2024, 14(9): 44-46. //RAN Anxian, WU Lian, WANG Fengmao, et al. Machine vision-based hazardous behavior detection and early warning system in petrochemical scenarios [J]. *Internet of Things Technologies*, 2024, 14(9): 44-46.