

基于大数据分析的新能源电力项目储备潜力评估与预测模型构建

马斌¹, 王玲²

1. 国网乌鲁木齐供电公司, 新疆, 乌鲁木齐 830001

2. 国网昌吉供电公司, 新疆, 昌吉 831100

DOI:10.61369/EPTSM.2025010022

摘要 : 近年来, 新能源电力产业发展迅猛。本文深入探讨了大数据分析在新能源电力领域的应用。一方面, 详细阐述新能源电力项目储备潜力评估模型的构建, 涵盖评估指标选取的科学性、系统性、可操作性和动态性原则, 以及模型选择与构建要点和易被忽视的问题; 另一方面, 对预测模型进行研究, 包括模型选择、构建训练、评估优化等环节。旨在借助大数据分析技术, 为新能源电力项目的资源配置优化、投资风险降低提供有力支撑, 推动能源产业可持续发展。

关键词 : 新能源电力项目; 储备潜力评估; 预测模型; 大数据分析; 评估指标

Construction of a Reserve Potential Evaluation and Prediction Model for New Energy Power Projects Based on Big Data Analysis

Ma Bin¹, Wang Ling²

1. State Grid Urumqi Power Supply Company, Urumqi, Xinjiang 830001

2. State Grid Changji Power Supply Company, Changji, Xinjiang 831100

Abstract : In recent years, the new energy power industry has developed rapidly. This paper delves into the application of big data analysis in the field of new energy power. On one hand, it elaborates on the construction of a reserve potential evaluation model for new energy power projects, covering the scientific, systematic, operable, and dynamic principles of evaluation index selection, as well as key points and overlooked issues in model selection and construction. On the other hand, it explores the prediction model, including model selection, construction training, evaluation optimization, and other aspects. The aim is to provide strong support for optimizing resource allocation and reducing investment risks in new energy power projects through big data analysis technology, thereby promoting the sustainable development of the energy industry.

Keywords : new energy power projects; reserve potential evaluation; prediction model; big data analysis; evaluation indicators

引言

在全球能源结构加速向清洁化、低碳化转型的大背景下, 新能源电力产业迎来前所未有的发展机遇。风力发电、太阳能光伏、生物质能等新能源技术不断突破, 装机容量持续攀升, 成为推动能源革命的核心力量。然而, 新能源电力项目具有投资大、周期长、不确定性因素多等特点, 科学评估项目储备潜力并进行精准预测, 对于合理配置资源、降低投资风险、保障能源可持续供应至关重要。随着信息技术的飞速发展, 大数据分析技术以其强大的数据处理和挖掘能力, 为新能源电力项目储备潜力评估与预测提供了新的思路和方法。

一、大数据分析在新能源电力领域的应用现状

据国家能源局统计, 2024年上半年, 全国全社会用电量达到46575亿千瓦时, 同比增长8.1%。与此同时, 以风电、光伏为代表的新能源产业迅速崛起。2024年上半年, 全国风电、光伏发电量合计达到1.56万亿千瓦时, 同比增长22%。新能源的蓬勃发展

为电力行业注入了新活力^[1]。

目前, 大数据分析深度融入新能源电力领域, 已取得了诸多成果。在新能源发电预测方面, 通过收集历史气象数据、机组运行数据等, 运用机器学习算法建立预测模型, 能够较为准确地预测风力发电、光伏发电的功率输出, 帮助电网企业合理安排调度计划, 提高电力系统的稳定性和可靠性。例如, 一些地区利用卫

星云图数据、气象雷达数据和历史光照强度数据，对光伏发电功率进行短期预测，预测精度显著提升。

在设备状态监测与故障诊断领域，大数据分析发挥着重要作用。通过对新能源发电设备的实时运行数据、振动数据、温度数据等进行采集和分析，能够及时发现设备潜在故障，实现预防性维护，降低设备故障率和维修成本。另外，在电力市场交易中，大数据分析可以对市场供需数据、电价数据等进行分析预测，帮助发电企业制定合理的竞价策略，提高经济效益。

尽管大数据分析在新能源电力领域已取得一定进展，但在项目储备潜力评估方面的应用仍相对较少。现有评估方法大多依赖传统的经验判断和简单的指标分析，难以充分考虑众多复杂的影响因素，导致评估结果的准确性和可靠性不足^[2]。二、新能源电力项目储备潜力评估模型构建

(一) 评估指标选取原则

1. 科学性原则

评估指标的选取必须基于新能源电力项目的客观发展规律与内在逻辑，确保指标能够真实、准确地反映项目储备潜力的本质特征。所采用的指标概念应清晰明确，计算方法需科学严谨，数据来源要真实可靠。

2. 系统性原则

新能源电力项目储备潜力受到多种因素的综合影响，涵盖资源条件、技术水平、经济成本、政策环境等多个维度。因此，指标体系应具有系统性，全面涵盖项目潜力评估的各个方面，避免出现重要因素的遗漏。各指标之间应相互关联、相互补充，形成一个有机的整体。

3. 可操作性原则

选取的评估指标应便于数据的收集、整理与分析。一方面，指标的数据来源要稳定、可获取，能够通过现有的统计渠道、监测设备或调研方式获得；另一方面，指标的计算方法不宜过于复杂，要在保证评估准确性的前提下，尽量简化计算过程，提高评估工作的效率^[3]。

4. 动态性原则

新能源电力行业处于快速发展之中，技术不断创新，市场环境、政策要求也在持续变化。因此，评估指标应具有动态性，能够及时反映行业发展的新趋势、新变化。

(二) 评估模型选择与构建

在新能源电力项目储备潜力评估体系里，模型的挑选与搭建极为关键，必须紧密贴合大数据特性以及新能源电力项目的复杂本质^[4]。以往的评估模型多依赖有限样本，基于经验假设构建，在过去数据量少、类型单一的情况下尚可勉强应用。但步入大数据时代，新能源电力领域数据呈现海量、多源异构特征。据统计，仅一个中型风电场，每天产生的设备运行数据就可达数万条，还涵盖气象、地理、市场交易、政策法规等多元数据，不同来源数据在格式、结构和质量上差异巨大^[5]。传统模型难以高效处理和深度挖掘这些数据，严重影响评估结果的准确性与可靠性，因此，机器学习与深度学习模型凭借独特优势，在该领域备受关注。

新能源电力项目中各类数据存在复杂非线性关系，传统线性模型难以捕捉，而神经网络模型模拟人类大脑神经元结构和工作原理，由大量神经元和复杂连接权重组。通过对海量数据的学习训练，它能挖掘出数据间隐藏的复杂潜在关系^[6]。比如预测分布式光伏发电项目发展趋势时，神经网络模型可同时分析光照强度动态变化、光伏发电设备技术更新、当地电力市场供需波动以及政策导向调整等多种因素，精准预测项目未来走向，为项目规划和决策提供可靠依据。

构建评估模型时，数据质量极为关键。新能源电力项目涉及多源数据，必须对其清洗、预处理，剔除噪声与异常值，统一格式和标准，保障数据准确一致^[7]。同时，要合理确定评估指标体系，除常见的技术可行性、经济收益等指标外，还应纳入环境影响、社会效益等因素。比如，项目对生态环境的潜在破坏程度，以及投产后对当地就业和经济发展的带动作用等，都应作为重要评估依据。

在评估模型选择与构建过程中，存在一些易被忽视的问题。其一，模型可解释性问题，复杂的机器学习和深度学习模型如同“黑箱”，难以直观阐释输出结果的内在逻辑，而在新能源电力项目评估中，决策者需清晰了解评估依据和过程，否则难以信任评估结果^[8]。其二，数据隐私与安全问题，大数据环境下，数据涉及多方利益与敏感信息，在采集、传输、存储和使用过程中，必须采取严格安全措施，防止数据泄露和非法利用。其三，模型适应性和动态更新问题，新能源电力行业发展迅速，政策、技术、市场环境不断变化，评估模型不能一成不变，需依据实际情况及时调整优化，确保评估结果的时效性与准确性。

二、新能源电力项目储备潜力预测模型构建

(一) 预测模型选择

传统预测模型中，回归分析模型基于历史数据建立变量间的线性关系，常用于趋势相对稳定、影响因素较为明确的场景，但其对非线性关系的刻画能力较弱；时间序列模型，如 ARIMA，依赖数据的平稳性假设，在处理新能源电力数据的波动性和季节性时存在局限性^[9]。相比之下，机器学习模型展现出更强的适应性，随机森林算法通过构建多个决策树进行集成学习，能够有效处理高维数据和非线性关系，降低过拟合风险；支持向量机在小样本数据预测中表现优异，可通过核函数将数据映射到高维空间，实现复杂边界的划分。深度学习模型中的 LSTM，因其独特的门控机制，能够有效捕捉时间序列数据中的长期依赖关系，在新能源发电功率预测等领域应用广泛。但模型选择时，常因对数据特性把握不足而陷入误区^[10]。另外，过度追求模型复杂度，选择不适合数据规模和特征的深度学习模型，不仅会增加计算成本，还可能导致模型难以训练和解释。

(二) 模型构建与训练

数据是模型构建的根基，在实际操作中，需要从多个维度采集和新能源电力项目储备潜力相关的数据，像新能源资源数据（例如太阳能辐射量、风速、水能流量等）、地理环境数据（包

括地形地貌、土地利用类型、生态保护红线等）、社会经济数据（涵盖区域人口数量、GDP、电力需求增长率等）、政策法规数据（包含新能源补贴政策、碳排放政策、电力市场规则等），还有技术发展数据（比如发电设备效率、储能技术成本等）。由于这些数据来源广泛，常常存在格式不统一、数据有缺失或者包含异常值等情况，所以必须进行预处理。具体来说，数据清洗环节会筛除重复数据、修补缺失数据、处理异常数据；通过数据集成技术，把不同渠道的数据整合成一个统一的数据集；再利用数据转换方法，比如归一化、标准化等，让原本格式不一的数据具有可比性。

在新能源电力项目储备潜力预测中，气象数据作为核心输入变量，其准确性直接关系到模型预测精度，然而实际应用中，风速、光照强度等关键气象参数的测量误差往往被低估^[10]。以风速数据为例，在风速测量中，仪器的校准偏差和传感器老化是导致数据失准的主要原因。假设真实风速为 v_t ，由于校准偏差和传感器老化等因素，实际测量风速 v_m 可表示为 $v_m = v_t + \delta_v + \varepsilon_v$ ，其中 δ_v 为校准偏差， ε_v 为服从正态分布 $N(0, \sigma_v^2)$ 的随机噪声。这种误差会直接影响风力发电功率模型的准确性，风力发电功率通常表示为 $P_o = 0.5\rho A v^3 C_p$ （其中 ρ 为空气密度， A 为扫风面积， C_p 为功率系数）。当风速存在测量误差时，发电功率的预测值与真实值之间会产生偏差，偏差的方差为 $V\text{ar}(\Delta P_o) = (1.5\rho A v_t^2)^2 \sigma_v^2$ ，这使得模型难以准确学习气象条件与新能源发电的关联，进而影响对发电功率的预测精度。

模型训练是实现预测模型从理论构建到实际应用的关键步骤，其核心在于通过对历史数据的学习，使模型能够准确捕捉新能源电力项目储备潜力的影响因素及内在规律。模型训练过程中，采用机器学习算法如支持向量机（SVM）、随机森林（RF）等。以支持向量机为例，其核心是寻找一个超平面来最大化样本间的间隔。对于线性可分数据，设训练数据集为 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ，其中 $x_i \in R^d$ ， $y_i \in \{-1, 1\}$ ，超平面方程为 $w^T x + b = 0$ ，通过求解优化问题 $\min_{w,b} \frac{1}{2} \|w\|^2$ ，约束条件为 $y_i(w^T x_i + b) \geq 1, i=1,2,\dots,n$ ，得到最优超平面参数 w 和 b 。对于非线性数据，则引入核函数 $K(x_i, x_j)$ ，将数据映射到高维空间进行处理。在训练过程中，利用交叉验证法对模型参数进行调优，以提高模

型的泛化能力，确保预测结果的准确性和可靠性。同时，训练数据的划分也需谨慎，若训练集与测试集存在数据泄露，即测试集数据在训练过程中被模型提前“学习”，会导致模型评估结果虚高，无法真实反映模型的泛化能力。

（三）模型评估与优化

模型评估指标的选择需综合考虑预测任务的特点和需求。对于新能源电力项目储备潜力预测，除了常用的 MSE、MAE、R² 等指标，还应关注预测误差的分布情况，如均方根误差（RMSE）能反映预测值与真实值的平均偏离程度，而平均绝对百分比误差（MAPE）则更直观地展示预测误差的相对大小，适用于不同规模数据的比较。在评估过程中，模型的泛化能力是核心关注点，过拟合模型在训练集上表现良好，但在实际应用中预测精度大幅下降。为解决过拟合问题，可采用正则化技术，如 L1 和 L2 正则化通过对模型参数添加惩罚项，限制模型复杂度；Dropout 方法在训练过程中随机丢弃部分神经元，增强模型的鲁棒性。此外，模型优化还可从数据增强、模型融合等方向着手。数据增强通过对原始数据进行变换，如旋转、缩放、平移等，扩充训练数据量；模型融合将多个不同模型的预测结果进行组合，充分发挥各模型的优势，提高预测准确性。但模型优化过程中，容易忽视优化方法对模型可解释性的影响，例如复杂的深度学习模型经过多重优化后，其决策过程更加难以理解，不利于向决策者解释预测结果。

三、结束语

大数据分析为新能源电力项目储备潜力评估与预测带来了新机遇与方法。通过构建科学合理的评估与预测模型，能够更精准地把握项目潜力，为新能源电力产业的资源配置、投资决策提供有力支持。然而，在模型构建与应用过程中，仍面临数据隐私安全、模型可解释性、动态更新等诸多挑战。未来，需要持续优化模型，加强数据管理与安全保障，深入研究如何提高模型的实用性和可靠性，以更好地适应新能源电力行业快速发展的需求，推动全球能源结构向清洁化、低碳化转型。

参考文献

- [1] 刘洋. 新能源发展背景下的电力营销市场开拓策略 [J]. 销售与市场, 2024, (18): 40–42.
- [2] 伍璇. 大数据技术在新能源电站智能化运营监管中的应用 [J]. 时代汽车, 2024, (12): 31–33.
- [3] 王民, 秦天勇. 基于改进因子分析法的新能源电力技术环境影响综合评价方法研究 [J]. 环境科学与管理, 2024, 49(07): 184–188.
- [4] 陈鸿鑫, 惠恒宇, 孙铭磊, 等. 考虑动态分区可靠性的新能源电力系统备用优化模型 [J]. 电力系统自动化, 2024, 48(23): 65–75.
- [5] 王策. 基于分布式模型预测控制的新能源电力系统调频策略研究 [D]. 华北电力大学(北京), 2023. DOI: 10.27140/d.cnki.ghbbu.2023.000160.
- [6] 罗政杰, 任惠, 辛国雨, 等. 基于模型预测控制的高比例可再生能源电力系统多时间尺度动态可靠优化调度 [J]. 太阳能学报, 2024, 45(06): 150–160. DOI: 10.19912/j.0254-0096.tynxb.2023-0177.
- [7] 陈道平, 廖海凤, 谭洪. 考虑可再生能源配额的风火水多能源电力系统年度调度模型 [J]. 运筹与管理, 2024, 33(02): 71–77.
- [8] 戴璐平, 沈嘉怡, 张飞飞. 基于时间序列算法的能源电力需求自动预测模型 [J]. 自动化技术与应用, 2024, 43(01): 49–51+65. DOI: 10.20033/j.1003-7241.(2024)01-0049-04.
- [9] 占萌, 张亚耀, 马锐. 以新能源为主体的新型电力系统模型体系和同步稳定研究综述 [J]. 北京师范大学学报(自然科学版), 2023, 59(06): 878–886. DOI: 10.12202/j.0476-0301.2023146.
- [10] 孙杨. 面向新能源电力领域创新技术知识模型构建及系统开发 [D]. 北京邮电大学, 2024. DOI: 10.26969/d.cnki.gbydu.2024.002760.