

# 基于大数据分析的新能源发电预测与优化模型研究

邓泽群

吉林吉电新能源有限公司, 吉林 长春 130000

**摘要：** 全球对清洁能源需求增长, 新能源发电在能源领域越发重要, 但新能源发电的间歇性和不稳定性给电力系统的规划、运行和调度带来很大挑战。本文围绕基于大数据分析的新能源发电预测与优化模型, 阐述新能源发电类型及特性和大数据技术在能源领域的应用原理, 全面介绍大数据驱动的新能源发电预测模型以及大数据助力的新能源发电优化模型。详细分析各类模型的原理、优势与局限, 对比模型性能评估指标。目的是为新能源发电领域的研究与实践提供全面理论参考, 推动新能源发电高效又稳定的发展。

**关键词：** 大数据分析; 新能源发电; 预测模型; 优化模型; 综述

## Research on Prediction and Optimization Models for New Energy Power Generation Based on Big Data Analysis

Deng Zequn

Jilin Jidian New Energy Co., Ltd. Changchun, Jilin 130000

**Abstract:** The global demand for clean energy is growing, making new energy generation increasingly important in the energy sector. However, the intermittent and unstable nature of new energy generation poses significant challenges to the planning, operation, and scheduling of power systems. This paper focuses on a new energy generation prediction and optimization model based on big data analysis, elaborating on the types and characteristics of new energy generation and the application principles of big data technology in the energy sector. It provides a comprehensive introduction to big data-driven new energy generation prediction models and big data-assisted new energy generation optimization models. It analyzes in detail the principles, advantages, and limitations of various models, comparing performance evaluation metrics. The aim is to provide a comprehensive theoretical reference for research and practice in the field of new energy generation, promoting its efficient and stable development.

**Keywords:** big data analysis; new energy power generation; prediction model; optimization model; review

## 引言

全球能源结构加快向清洁能源转变, 新能源发电作为可持续能源发展的关键部分, 受到广泛关注。太阳能、风能、水能、生物质能等新能源清洁、可再生, 对缓解能源危机和减少环境污染意义重大。不过新能源发电受自然条件(像光照强度、风速、水文等)影响大, 有明显的间歇性和不稳定性, 这就使得新能源发电并入电网时可能对电网的稳定性、可靠性和电能质量产生不良影响。大数据分析技术快速发展, 为解决新能源发电上述问题带来新机会, 大数据有海量数据、快速流转、数据类型多样和价值密度低的特点, 并且能整合新能源发电相关的多源数据, 挖掘数据背后隐藏的规律和趋势。基于大数据分析构建的新能源发电预测模型可提前知道发电功率变化, 为电力系统的调度和运行提供准确依据; 优化模型能在考虑多种约束条件下实现新能源发电的最优配置, 提高发电效率, 降低发电成本。所以开展基于大数据分析的新能源发电预测与优化模型研究, 有重要理论意义和实际应用价值。

## 一、新能源发电与大数据技术基础

### (一) 新能源发电类型及特性

常见新能源发电类型主要有太阳能发电、风能发电、水能发电和生物质能发电等。其中太阳能发电主要通过光伏发电板把太

阳能转为电能, 它的发电特性受光照强度以及温度等因素影响明显。白天光照足时发电功率高; 夜晚或阴天时发电功率大幅下降甚至为零, 另外温度过高也会降低光伏电池的转换效率。风能发电利用风力发电机把风能转为电能, 风速大小和稳定性直接决定风能发电的功率。因风速有随机性和间歇性, 风能发电的输出功

作者简介: 邓泽群(1993.09-), 男, 汉族, 籍贯: 吉林省吉林市, 硕士, 助理工程师, 研究方向: 计算机技术在新能源电力系统中应用。

率波动大且难准确预测，同时不同地区风资源分布差异大，这对风电场选址和布局要求高。

而水能发电依靠水轮机把水流能量转为电能，其发电功率主要取决于水位落差和流量。受季节和气候影响，河流水量有明显丰枯变化，导致水能发电出力也随之波动。同时水电站建设还受地理条件和生态环境等多方面限制。生物质能发电则是利用生物质能转为电能，比如用农作物秸秆、林业废弃物等发电，生物质能发电稳定性相对较高，但原料供应受季节和地域限制且发电效率相对较低<sup>[1]</sup>。

## （二）大数据技术核心概念与能源应用原理

大数据技术核心概念围绕数据的“4V”特征，即大量（Volume）、高速（Velocity）、多样（Variety）和价值（Value），在能源领域大数据技术应用原理主要体现在以下方面：

在数据采集环节，通过各种传感器、智能电表、监测设备等去实时采集新能源发电设备运行数据、气象数据、电网负荷数据等海量信息。这些数据来源广且类型多样，包括结构化数据（如设备运行参数）、半结构化数据（如日志文件）和非结构化数据（如卫星图像、文本报告）。

数据传输方面借助高速通信网络（如5G、光纤网络）把采集到的数据快速传输到数据存储和处理中心，保证数据时效性。

数据存储采用分布式存储技术（如Hadoop分布式文件系统HDFS），能存储海量数据，并保证数据安全性和可靠性<sup>[2]</sup>。

而在数据处理与分析阶段，运用数据挖掘、机器学习、深度学习等技术，对数据进行清洗、预处理、特征提取和建模分析，挖掘数据中潜在规律和价值，进一步为新能源发电预测和优化提供数据支持。

## 二、大数据驱动的新能源发电预测模型综述

### （一）传统预测模型概述与局限

传统新能源发电预测模型主要有时间序列分析和回归分析等。其中时间序列分析基于发电功率历史数据，通过建立时间序列模型（如ARIMA模型）预测未来发电功率，该方法假设数据有平稳性和周期性，通过对历史数据统计分析预测未来趋势。但新能源发电数据受自然因素影响大并且随机性和波动性强，难满足平稳性假设，进而导致预测精度受限。而回归分析则通过建立发电功率与影响因素（如风速、光照强度等）的线性或非线性回归方程进行预测。但实际中新能源发电与影响因素关系复杂，不是简单线性关系，传统回归分析难准确描述这种复杂关系，影响预测效果。

### （二）基于机器学习的预测模型

#### 1. 神经网络模型

神经网络模型有强大非线性映射能力，能自动学习数据中复杂模式和特征，在新能源发电预测中常用神经网络模型有BP神经网络和RBF神经网络。其中BP神经网络是多层前馈神经网络，通过反向传播算法调整网络权重和阈值最小化预测值与实际值误差。它由输入层、隐藏层和输出层组成，输入层接收发电相关各

种因素数据（如气象数据、时间数据等），隐藏层对输入数据进行特征提取和非线性变换，输出层输出预测发电功率。不过BP神经网络存在训练速度慢或易陷入局部最优等问题<sup>[3]</sup>。而RBF神经网络则采用径向基函数作为激活函数，局部逼近能力强、训练速度快。它能更有效处理非线性问题，在新能源发电预测中也有较好应用效果，但RBF神经网络的径向基函数中心和宽度选择较难，需通过合适算法优化。

#### 2. 支持向量机模型

支持向量机（SVM）是基于统计学习理论的机器学习算法，在小样本、非线性问题上有独特优势，其基本思想是寻找最优分类超平面，分开不同类别数据。在新能源发电预测中SVM把发电数据和影响因素数据映射到高维空间，在高维空间寻找最优分类超平面，实现发电功率预测<sup>[4]</sup>。而SVM优点则是泛化能力强并且对小样本数据适应性好，但它对核函数选择很敏感，不同核函数导致不同预测结果且计算复杂度高，对大规模数据处理效率低。

### （三）基于深度学习的预测模型

#### 1. 递归神经网络及其变体

递归神经网络（RNN）适合处理时间序列数据，能利用历史信息预测未来值。在新能源发电预测中RNN可捕捉发电功率随时间变化趋势，但传统RNN有梯度消失和梯度爆炸问题，难处理长期依赖关系。为解决这一问题就出现长短期记忆网络（LSTM）和门控循环单元（GRU）等变体。LSTM通过引入门控机制（输入门、遗忘门和输出门），有效控制信息流动，更好处理长期依赖问题。GRU则是对LSTM的简化，把输入门和遗忘门合并为更新门，计算复杂度相对较低，在新能源发电预测中也表现出良好性能。

#### 2. 卷积神经网络及其融合模型

卷积神经网络（CNN）最初主要用于图像识别领域，通过卷积层、池化层和全连接层等结构去自动提取数据局部特征。在新能源发电预测中CNN可用于提取气象数据或发电设备图像等数据的空间特征。同时为充分利用数据时空特征，近年出现CNN与LSTM等模型的融合模型，比如把CNN提取的空间特征作为LSTM输入，让LSTM进一步处理时间序列信息，进而实现对新能源发电功率更准确预测。这种融合模型结合CNN和LSTM优势，处理复杂时空数据时表现更好<sup>[5]</sup>。

### （四）模型性能评估指标对比

在新能源发电预测模型中，常用的性能评估指标包括均方根误差（RMSE）、平均绝对误差（MAE）、平均绝对百分比误差（MAPE）等。

RMSE能反映预测值与真实值偏差的平均幅度，对较大误差更为敏感，其计算公式为：

$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$ ，其中 $y_i$ 为真实值， $\hat{y}_i$ 为预测值， $n$ 为样本数量。MAE衡量预测值与真实值误差的平均绝对值，计算简单且直观，公式为： $MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$ 。

MAPE以百分比形式展示预测误差，便于不同规模数据间比较，公式为：

$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\%$ 。

传统时间序列分析模型，如 ARIMA，由于对数据平稳性要求高，在新能源发电数据波动大的情况下，RMSE、MAE 和 MAPE 值往往较大，预测精度低<sup>[6]</sup>。

BP 神经网络模型，训练过程易陷入局部最优，导致预测结果不稳定，RMSE 和 MAE 相对较高，MAPE 也不理想。

RBF 神经网络训练速度快，但径向基函数参数选择困难，若参数不合适，RMSE、MAE 会偏大，影响预测准确性。

SVM 对小样本数据有优势，但其核函数选择敏感，核函数选择不佳时，RMSE、MAE 增大，MAPE 也会升高，且在大规模数据下计算效率低，影响其在实际中的应用。

RNN 受梯度消失和梯度爆炸问题影响，预测长期序列时误差较大，RMSE、MAE 和 MAPE 值较高。

LSTM 和 GRU 能有效处理长期依赖问题，相比 RNN，RMSE、MAE 和 MAPE 值明显降低，预测精度显著提升。

CNN 及其融合模型，在处理具有时空特征的数据时，通过提取空间和时间特征，能更准确地预测新能源发电功率，RMSE、MAE 较低，MAPE 也处于较低水平，展现出较好的性能。

### 三、大数据助力的新能源发电优化模型综述

#### (一) 优化目标与约束条件梳理

新能源发电优化目标主要有最大化发电效率、最小化发电成本、提高能源利用率等。最大化发电效率可通过合理调整发电设备运行参数（如风机叶片角度、光伏电池跟踪角度等），让发电设备在不同自然条件下保持最佳发电状态。最小化发电成本需综合考虑设备投资成本、运维成本以及能源采购成本等因素，通过优化发电计划和资源配置，降低总体发电成本。而提高能源利用率旨在减少能源浪费，进而实现能源高效转化和利用。

在优化过程中需考虑多种约束条件，发电设备运行约束包括设备功率限制、启停约束以及寿命约束等。例如风机发电功率不能超额定功率，设备频繁启停影响设备寿命，所以要合理安排设备运行时间和启停次数。另外电网接入约束主要考虑电网容量限制、电压稳定性、频率稳定性等因素，新能源发电需与电网运行要求匹配，确保电力安全稳定传输和分配。能源供需平衡约束要求新能源发电输出能满足用户用电需求，同时避免能源过度生产或短缺<sup>[7]</sup>。

#### (二) 数学规划类优化模型

##### 1. 线性规划模型

线性规划模型是经典优化模型，目标函数和约束条件都是线性函数。在新能源发电中线性规划模型可用于发电资源分配、发电计划制定等，比如在多个新能源发电站间分配发电任务时，以发电成本最小化为目标函数，以各发电站发电功率限制、能源需求等为约束条件来建立线性规划模型。求解该模型可得到最优发电分配方案，同时线性规划模型优点是求解算法成熟并且计算效率高，但只能处理线性关系，对实际中存在的非线性问题应用受限。

##### 2. 非线性规划模型

非线性规划模型适用于处理目标函数或约束条件中有非线性关系的优化问题。在新能源发电领域，很多实际问题有非线性特征，如发电设备效率曲线通常是非线性的，非线性规划模型能更准确描述这些复杂关系但求解难度大。常用求解算法有梯度下降法和拟牛顿法等，其中梯度下降法通过迭代更新变量，沿目标函数负梯度方向找最优解；拟牛顿法则通过近似海森矩阵加速收敛。不过这些算法易陷入局部最优解，对大规模或者复杂非线性规划问题来说求解效果可能不好<sup>[8]</sup>。

#### (三) 智能优化算法

##### 1. 遗传算法

遗传算法是模拟生物进化过程的随机搜索算法，通过对种群中个体（即可能的解）进行选择、交叉和变异操作，逐步找到最优解。在新能源发电优化中遗传算法可用于优化发电设备运行参数、发电调度方案等。首先对决策变量编码，转化为染色体形式；然后随机生成初始种群并计算每个个体适应度值，适应度值根据目标函数和约束条件评估；接着通过选择操作从当前种群选适应度高的个体去作为下一代种群父代；父代个体通过交叉和变异操作产生子代个体而形成新种群。不断重复上述过程，直到满足终止条件（如达到最大迭代次数或适应度值不再变化），此时得到的最优个体就是优化模型近似最优解。遗传算法优点是全局搜索能力强，对问题依赖性小但计算量大，易出现早熟收敛现象<sup>[9]</sup>。

##### 2. 粒子群优化算法

粒子群优化算法是基于群体智能的优化算法，模拟鸟群觅食行为，在新能源发电优化中把每个可能的解看作搜索空间中一个粒子，粒子在搜索空间以一定速度飞行并且通过不断调整自身位置和速度去找最优解。粒子速度和位置更新根据自身历史最优位置和群体全局最优位置进行，粒子群优化算法优点是收敛速度快且容易实现，但易陷入局部最优。为克服这一缺点就可采用多种改进策略，如引入惯性权重、自适应调整参数等。

#### (四) 模型性能评估指标对比

在新能源发电优化模型中，常用的性能评估指标有发电效率提升率、发电成本降低率、能源利用率提升率等。

发电效率提升率用于衡量优化模型对发电设备发电效率的提升程度，公式为：发电效率提升率 = 优化前发电效率优化后发电效率 - 优化前发电效率 × 100%。

发电成本降低率反映优化模型在降低发电成本方面的效果，公式为：发电成本降低率 = 优化前发电成本优化后发电成本 - 优化后发电成本 × 100%。

能源利用率提升率体现优化模型对能源利用效率的改进情况，公式为：能源利用率提升率 = 优化前能源利用率优化后能源利用率 - 优化前能源利用率 × 100%。

线性规划模型在处理线性关系的发电资源分配等问题时，能快速找到理论上的最优解，可有效提高发电效率提升率，一定程度降低发电成本，提升能源利用率。但实际中非线性问题较多，限制了其对这些指标的进一步优化。

非线性规划模型能更准确描述实际问题中的非线性关系，理

论上对发电效率提升率、发电成本降低率和能源利用率提升率的优化潜力更大。然而，由于求解难度大，易陷入局部最优，实际应用中可能无法充分发挥其优势，导致这些指标提升效果不稳定。

遗传算法全局搜索能力强，在优化发电设备运行参数和调度方案时，有可能找到较优解，提高发电效率提升率，降低发电成本，提升能源利用率。但计算量大、早熟收敛问题可能影响其最终优化效果，使得各项指标提升幅度受限。

粒子群优化算法收敛速度快，在处理一些简单的新能源发电优化问题时，能快速提高发电效率提升率、降低发电成本，提升能源利用率。但易陷入局部最优，对于复杂问题，可能导致优化后的指标提升程度有限<sup>[10]</sup>。

## 四、结论

综上所述，大数据分析技术为新能源发电领域带来了创新性的解决方案。从预测模型来看，传统的时间序列分析和回归分析模型，因难以适应新能源发电数据的随机性和波动性，在预测精度上存在明显局限，难以满足实际需求。

基于机器学习的神经网络模型，如 BP 神经网络和 RBF 神经网络，虽具备强大的非线性映射能力，但 BP 神经网络训练速度慢、易陷入局部最优，RBF 神经网络的径向基函数参数选择困

难，影响了其预测性能的稳定性。支持向量机模型在小样本、非线性问题上具有优势，然而对核函数选择敏感且计算复杂度高，限制了其在大规模数据处理中的应用。而基于深度学习的递归神经网络及其变体（LSTM 和 GRU），通过改进门控机制有效解决了长期依赖问题，显著提升了预测精度，展现出良好的发展性。卷积神经网络及其融合模型，能够充分提取数据的时空特征，在处理复杂的新能源发电数据时表现出色，同样具有较大的发展潜力。

在优化模型方面，线性规划模型虽求解算法成熟、计算效率高，但仅能处理线性关系，在面对实际中大量的非线性问题时应用受限。非线性规划模型虽能准确描述复杂的非线性关系，但求解难度大，易陷入局部最优，实际应用效果不稳定。另外遗传算法全局搜索能力强，但计算量大且易出现早熟收敛现象。粒子群优化算法收敛速度快、容易实现，却也存在易陷入局部最优的问题，这些传统的优化算法在新能源发电优化中均面临一定挑战。

未来应针对现有模型的局限性，要进一步深入研究大数据与新能源发电技术的融合。一方面要优化现有模型，如改进机器学习和深度学习模型的训练算法，提高计算效率，避免陷入局部最优；另一方面还要开发更高效、精准的预测与优化模型，结合新兴技术，如强化学习、迁移学习等，以更好地应对新能源发电的间歇性和不稳定性挑战。通过不断努力，推动新能源发电在全球能源结构中占据更重要的地位，助力能源行业的可持续发展。

## 参考文献

- [1] 王华元, 邓星野. 基于大数据分析的新能源发电数据分析与预测研究 [J]. 电气技术与经济, 2024, (07): 239-241.
- [2] 李琦. 基于大数据与智能算法的新能源项目风险分析 [J]. 集成电路应用, 2024, 41(02): 296-297.
- [3] 陈瑞. 新能源发电企业大数据中心提高数据采集质量的方法及其有效性分析 [J]. 云南电力技术, 2022, 50(06): 75-79+92.
- [4] 张扬帆, 杨伟新, 梁恺. 以数据为桥梁为新能源产业赋能增值 [J]. 华北电业, 2022, (04): 28-30.
- [5] 任东方. 多种能源发电协同发展管控模型及大数据分析研究 [D]. 华北电力大学 (北京), 2020.
- [6] 王华元, 邓星野. 基于大数据分析的新能源发电数据分析与预测研究 [J]. 电气技术与经济, 2024, (07): 239-241.
- [7] 陈志煌. 基于大数据分析的煤电与新能源互补优化调度策略研究 [J]. 前卫, 2024, (01): 0007-0009.
- [8] 李雯, 魏斌, 韩肖清, 等. 面向滚动优化调度的光伏发电功率日内超短期预测 [J]. 电力系统及其自动化学报, 2020, 32(11): 43-49. DOI: 10.19635/j.cnki.csu-epsa.000439.
- [9] 卢才云, 朱文, 张海天, 等. 新能源发电功率预测准确率计算策略优化研究与应用 [J]. 电气应用, 2024, 43(06): 54-59.
- [10] 曹婧, 王瑞. 智能电网中新能源发电的预测与调度优化研究 [J]. 张江科技评论, 2024, (07): 108-110.