

基于大数据分析的电力系统负荷预测模型研究

肖永青

广东盈浩新能源科技有限公司，广东省中山市，528400；

摘要：在电力系统中，负荷预测是保证电力系统安全稳定运行的基础，其重要性不言而喻。传统的负荷预测方法主要包括时间序列分析方法、机器学习方法（如随机森林、支持向量机、深度学习等）、数据挖掘方法（如主成分分析法、因子分析法）等，但这些方法的计算复杂度高，难以在工业级场景下实现。针对上述问题，本文提出了基于大数据分析的负荷预测模型设计方案。

关键词：大数据分析；电力系统；负荷预测模型

DOI：10.69979/3060-8767.25.09.064

引言

近年来，随着我国社会经济的高速发展，电力系统负荷预测对电力系统运行的重要性越来越突出。准确、可靠的负荷预测，能够为电力系统安全稳定运行提供重要保障，同时也有利于电力系统的经济管理和调度决策。因此，对电力系统负荷预测方法进行研究，提高其精度和效率具有重要意义。目前，国内外学者提出了多种基于大数据分析的负荷预测方法，如时间序列分析方法、机器学习方法、数据挖掘方法等。

1 电力系统负荷预测概述

在电力系统中，负荷预测是保障电力系统安全稳定运行的重要手段。在电力负荷预测中，为保证电力系统安全稳定运行，需要对负荷变化趋势进行准确预测。目前，负荷预测主要采用时间序列分析方法、机器学习方法和数据挖掘方法。其中，时间序列分析方法具有精度高、预测速度快等优点，但计算复杂度高，适用于大规模数据集；机器学习方法能够在短时间内利用大量样本数据训练出较好的模型，但受模型结构的影响较大，难以在工业级场景下实现；数据挖掘方法能够挖掘大量数据之间的相关性并将其用于负荷预测，但由于数据量过大而导致计算复杂度高，难以在工业级场景下实现^[1]。

2 传统负荷预测方法及其局限性

传统的时间序列分析方法主要有三种：单变量时间序列分析方法、多变量时间序列分析方法和混合模型。其中，单变量时间序列分析方法主要包括自回归模型（AR）、自回归移动平均（ARIMA）模型等，多变量时间序列分析方法主要包括指数平滑法、移动平均法、最小二乘估计（LS-SVM）模型等，混合模型则是将传统的时间序列分析方法和机器学习方法混合起来的一种预测模型。在电力系统中，由于负荷数据具有海量、低值和高变等特征，因此传统的负荷预测方法无法有效处理大规

模数据集，并且由于传统的负荷预测模型难以在工业级场景下实现。

3 电力系统负荷预测中的大数据分析需求

3.1 电力负荷数据特性分析

电力负荷数据是一种复杂的大规模数据，具有海量、低值和高变等特征，传统的负荷预测方法难以在工业级场景下实现。在电力负荷预测中，由于需要对海量数据进行处理，因此需要进行特征工程。特征工程是一项以特征提取为核心的数据处理方法，它将大量复杂的数据转换成一系列与原有数据具有一定相似性的特征，从而便于后续数据分析。在电力系统负荷预测中，负荷预测模型需要从海量数据中挖掘出一些与实际情况相符的特征。因此，在大数据分析中，需要对电力负荷数据进行深入分析，从中提取出所需的特征，并对这些特征进行处理以提高预测精度^[2]。

3.2 数据采集与预处理方法

在电力负荷预测中，由于数据采集和预处理不准确，导致模型误差较大。因此，在电力负荷预测中，需要对采集到的大量数据进行预处理，从而为负荷预测模型提供更为准确的数据基础。在电力系统中，数据采集通常包括物理设备监测和电子设备监测两种方式。在物理设备监测中，由于测量方法和精度等因素的影响，可能存在测量误差。而在电子设备检测中，由于缺乏统一的标准和规范，各厂商测量方法各异，甚至存在同一设备的不同厂家之间测量方法不同的情况。

3.3 多源异构数据的集成与管理

在电力系统中，存在大量的数据源，包括传感器、智能终端、手机APP等设备产生的数据。在负荷预测中，这些设备产生的数据会对模型预测结果产生重要影响。由于异构数据的多样性，传统的负荷预测方法难以对其

进行有效管理和处理。因此，需要对多源异构数据进行集成，并通过统一标准来进行数据管理和处理。在大数据分析中，需要将多源异构数据进行统一存储、管理和分析^[3]。

3.4 负荷预测模型对大数据的需求分析

随着大数据的广泛应用，大数据分析的理论也日益成熟，同时大数据技术在负荷预测领域的应用也不断深入，这些都为负荷预测模型的设计和发展提供了强有力的支撑。电力系统负荷预测模型设计中，对大数据的需求分析主要体现在以下几个方面：（1）海量数据规模性需求，主要包括数据采集量、数据量增长速度等方面。（2）大数据处理和分析技术的需求，主要包括大数据处理技术和大数据分析技术等。（3）数据处理速度要求，主要包括大数据存储速度、大数据访问速度等方面。

4 基于大数据分析的负荷预测模型设计

4.1 负荷预测模型设计原则

以用电特性为基础，通过建立负荷模型，使负荷预测更加准确，提高对负荷预测的准确性。负荷预测模型应考虑影响用电特性的因素，并通过分析其对负荷的影响程度，使预测结果更加合理。充分利用负荷数据与用电数据之间的相关性，提高预测精度。在预测模型设计过程中，应考虑各影响因素的权重、模型参数、样本容量等因素。根据实际情况选取不同的预测模型和参数进行预测分析，并将预测结果与实际数据进行比较，最终确定最优方案。为保证预测结果的准确性和有效性，在建立模型之前应对其进行多次测试和修正。

4.2 模型总体框架与流程

在负荷预测模型的设计中，为了提高预测精度，应以实际负荷数据为基础，建立负荷预测模型。模型框架的总体思路是以特征工程为核心，对负荷数据进行特征提取与优化处理，并采用不同的模型和算法进行预测，最终输出预测结果。在模型框架中，主要包括以下几个模块：（1）数据预处理模块：该模块主要包括数据采集、数据清洗、缺失值处理、特征工程等。（2）特征选择与构造模块：该模块主要包括特征提取与优化、特征选择与构造和算法选型。（3）模型训练和评估模块：该模块主要包括模型训练与评估和模型改进两个部分^[4]。

4.3 特征工程与数据处理方法

4.3.1 数据清洗与缺失值处理

在负荷预测模型的设计中，数据清洗是一项重要的工作。数据清洗的目的是消除由于设备故障、传感器故障、恶劣天气、设备故障等原因引起的数据缺失，同时将异常值剔除，提高数据质量。在数据清洗中，应采取

有效的方法来保证数据的准确性和完整性，常用的方法主要有三种：基于时间序列分析方法、基于机器学习方法和基于数据挖掘方法。其中，在基于时间序列分析方法中，可采用滑动平均法、指数平滑法和移动平均法等。在基于机器学习方法中，可采用线性回归、支持向量机和随机森林等。在基于数据挖掘方法中，可采用 KNN（K 近邻法）、Lasso 回归和随机森林等。

4.3.2 特征选择与构造

在特征工程中，通过分析负荷数据的特征，对其进行降维处理，提取出重要的特征。在负荷预测模型的设计中，需要对负荷数据进行特征选择与构造。在特征选择中，应采用一些有效的方法对负荷数据进行降维和压缩处理，提高预测结果的准确性和有效性。

4.3.3 特征提取与优化

在特征选择与构造中，应对负荷数据进行特征提取，并将其作为模型参数进行训练和评估。在特征提取中，可以采用基于网格搜索法、改进的信息增益法等方法进行特征提取。在特征优化中，应采用一些有效的方法来进行模型参数优化，如交叉验证法、基于梯度提升树算法、遗传算法等^[5]。

4.4 负荷预测算法选型与模型构建

4.4.1 时间序列分析方法

时间序列分析方法主要用于对一段时间内的负荷数据进行分析，通常是通过对该数据的统计特性进行分析，从而发现其变化规律和发展趋势，进而实现对负荷预测模型的构建。时间序列分析方法包括季节性负荷预测、日负荷预测以及周负荷预测。在进行电力系统的短期和中期预测时，时间序列分析方法可以较好地克服季节性因素的影响，对历史数据进行统计和分析，从而达到提高短期和中期预测精度的目的。但对于长期的负荷数据则需要通过模型拟合来建立长期预测模型。因此，时间序列分析方法在实际应用中通常作为短期和中期预测方法的补充。

4.4.2 机器学习方法

机器学习方法主要用于对历史负荷数据进行分析，根据历史负荷数据的特点建立相应的模型，并利用模型进行预测。其中最常见的机器学习方法有线性回归、人工神经网络和支持向量机等。其中，线性回归主要是根据历史负荷数据建立线性模型，通过拟合得到模型参数，然后利用该参数对未来一段时间的负荷进行预测；人工神经网络方法则是将大量的负荷数据输入到一个神经网络模型中，并通过神经网络模型预测未来一段时间的负荷；支持向量机方法则是利用支持向量机对历史负荷数据进行分析，通过判断每个负荷数据对支持向量机的分类情况，从而建立相应的预测模型。

4.5 模型参数优化与评估指标

在参数优化方面，首先，需要对负荷预测模型的参数进行优化。将优化后的参数与原始参数进行比较，以检验其预测能力是否优于原始参数。其次，通过对比各模型的预测结果，找出各模型之间的差异，并评估各个模型在某一时刻的预测能力。在评估指标方面，本研究采用均方根误差（RMSE）来评估预测精度。RMSE 定义为：对于一个训练集，最大误差与平均误差之比。平均误差是指在一段时间内预测值与真实值之间的相对差值，因此可以将均方根误差作为评估指标。为了便于比较，采用均方根误差平均值作为衡量预测精度的指标。

5 模型实现与实验分析

5.1 实验平台与数据集介绍

在大数据分析的背景下，利用大数据技术可以获得大量的负荷数据，如何从这些数据中提取有用信息并实现对负荷的预测，成了负荷预测领域研究的重点和热点。本文以电力系统中的负荷数据为研究对象，建立了基于大数据分析的负荷预测模型，并通过实验证明了该模型的有效性和可行性。本文所提出的基于大数据分析的电力系统负荷预测模型，主要采用了时间序列分析、机器学习和数据挖掘等方法，利用历史负荷数据对预测模型进行训练和评估。为了提高模型的预测精度，对模型进行了参数优化和评估指标改进。实验结果表明，该负荷预测模型在一定程度上提高了对未来一段时间负荷预测的精度。

随着电力系统规模不断扩大，数据规模也随之增大。如何将海量数据有效地组织起来、管理起来并充分利用这些数据已成为电力系统发展中面临的一个重要问题。因此，如何利用大数据技术对海量数据进行有效组织和管理，并充分利用这些数据来提高电力系统负荷预测精度已经成为当前研究领域的热点。本文从大数据分析的角度出发，在对负荷预测模型进行设计时，充分考虑了影响负荷的相关因素。同时，对于负荷数据中存在的一些问题进行了相应处理和改进。通过对这些问题进行分析和解决，可以有效地提高预测精度和效率。总之，基于大数据分析的电力系统负荷预测模型是目前电力系统负荷预测领域研究中一个重要发展方向。在未来，随着大数据技术在电力系统中不断深入应用，基于大数据分析的负荷预测模型将会得到更加广泛地应用。

5.2 负荷预测模型实现过程

在本实验中，数据集包含了实际电网中的一天 24 小时的电力负荷数据，每隔 8 小时获取一次数据。在本文中，用来预测未来 8 小时的数据作为训练集，用来测

试集来预测未来 1 天的数据作为测试集。在训练集中，对原始数据进行归一化处理，即去掉其中的数值型信息；对于测试集，计算预测值与实际值之间的均方误差（MSE），MSE 越小越好。由于在训练集中需要预测 1 天中每一天的数据，所以将训练集划分为训练集和测试集，通过实验对比来看预测值和实际值之间的误差是否在可接受范围内。

5.3 实验结果与性能分析

本文采用的数据集包含了大量的真实数据，采用传统方法训练的模型在实际应用中误差较大。本文在建立基于大数据分析的电力系统负荷预测模型过程中，针对传统算法中存在的缺陷，采用了改进的遗传算法来提高模型的预测精度。在实际应用中，由于遗传算法在解决复杂问题时存在一些缺陷，如全局搜索能力差，局部搜索能力强等。为了弥补这些缺陷，本文将遗传算法和 BP 神经网络结合起来使用，经过多次优化后得到了最佳组合。为了检验该模型在实际应用中的性能，本文采用了 4 种不同的天气类型作为输入数据进行仿真实验，最终得到了一个最优组合模型。

6 结语

本文首先介绍了遗传算法和 BP 神经网络理论，然后以某地区的日负荷数据为例，建立了基于遗传算法的 BP 神经网络负荷预测模型，并使用该模型对其进行仿真实验。最后根据仿真结果对模型进行了评价和分析，并与传统 BP 神经网络算法和 BP 神经网络预测模型进行对比。结果表明，本文提出的基于遗传算法的 BP 神经网络预测模型比传统算法在收敛速度、收敛精度和稳定性等方面有了较大提高。本文研究表明：遗传算法和 BP 神经网络相结合的负荷预测模型能较好地满足实际需求，能够满足电力系统负荷预测的精度要求，具有一定的实用性。

参考文献

- [1] 杨欣, 徐飞, 贺国伟, 等. 基于 BP 神经网络的能源负荷短期预测模型 [J]. 能源与环保, 2025, 47(06): 174–178.
- [2] 高嘉廉. 基于大数据分析的电力系统负荷预测与规划优化 [J]. 光源与照明, 2025, (01): 78–80.
- [3] 马秀娟. 大数据分析电力系统负荷预测与调度控制策略的研究 [J]. 家电维修, 2024, (08): 104–106.
- [4] 左继恩. 基于大数据分析的电力系统负荷预测与优化调度方法研究 [J]. 家电维修, 2024, (01): 52–54.
- [5] 王栋梁. 基于大数据分析的电力系统风险评估与预警模型 [J]. 数字技术与应用, 2025, 43(06): 116–118.