

摘要

传统煤炭能源所面临的资源紧缺和碳排放问题使得光伏太阳能发电量和光伏电站并网装机容量持续上升,而光伏电站出力由于受日光照射影响较大而出现的间歇性、随机性问题易引发并网点电压、频率大幅波动,在将直流逆变成交流并网的过程中使用大量电力电子逆变装置会产生谐波污染等电能质量问题,严重影响配电网的可靠安全运行。通过挖掘光伏电站电能质量的历史数据信息提前感知电能质量变化态势,及早发现存在的电能质量问题,为电能质量提前治理预留出时间并提供准确信息,成为可再生能源并网可靠运行的关键技术之一。本文基于深度神经网络混合模型挖掘光伏电站电能质量历史数据,实现对光伏电站电能质量未来变化态势的感知。具体研究工作如下:

(1) 提出光伏电站电能质量态势感知架构,完成光伏电站电能质量要素提取及态势理解。提出光伏电站电能质量态势感知由光伏电站电能质量要素提取、光伏电站电能质量态势理解和光伏电站电能质量态势预测三个阶段组成。介绍光伏电站电能质量要素提取实现过程,使用自相关系数和最大信息系数分析方法对电能质量数据进行态势理解,对电能质量数据和外部变量数据进行异常值和标准化处理,针对三种电能质量指标分别采用两种不同方法综合评估模型效果。

(2) 基于深度置信网络(Deep Belief Network, DBN)和支持向量回归(Support Vector Regression, SVR)方法构建光伏电站电能质量态势预测模型并进行算例分析。阐述 DBN 和 SVR 方法用于时序数据预测的原理,构建基于 DBN 和 SVR 方法的电能质量态势预测单一模型,介绍两种模型实现流程。以某地光伏电站电压、频率、电压总谐波畸变率三种电能质量实测数据进行算例分析,分析两种单一模型对于光伏电站电能质量态势预测的效果。

(3) 提出一种基于混合深度神经网络的光伏电站电能质量态势预测模型。针对 DBN 和 SVR 单一模型预测效果较差的情况,提出了一种卷积神经网络(Convolutional Neural Networks, CNN)和门控循环单元(Gated Recurrent Unit, GRU)混合深度神经网络态势预测模型。先通过 CNN 提取电能质量输入数据的高层特征,再利用 GRU 学习长时间电能质量序列前后关系,克服了传统单一模型预测精度有限且泛化能力较弱的缺点,解决了海量电能质量数据高层特征提取困难及模型学习长时序信息易梯度消失的问题。通过与另外几种模型进行实验对比,该混合模型在时效性较好的同时具有更高的预测准确度。

关键词: 光伏电站; 电能质量; 态势感知; 卷积神经网络; 门控循环单元

Abstract

The resource scarcity and carbon emission problems faced by traditional coal energy sources have led to a continuous increase in photovoltaic solar power generation and grid-connected installed capacity of photovoltaic power stations. However, the intermittent and random problems of the output of photovoltaic power stations due to the great influence of sunlight can easily cause large fluctuations in voltage and frequency of the grid-connection point, the use of a large number of power electronic inverters in the process of converting DC into AC for grid connection will cause power quality problems such as harmonic pollution, seriously affect the reliable and safe operation of the distribution network. By mining the historical data information of the power quality of photovoltaic power station, we can perceive the changes of power quality in advance, discover the existing power quality problems early, reserve time for the early management of power quality and provide accurate information, which has become one of the key technology for the reliable operation of renewable energy grid-connected. This paper is based on a deep neural network hybrid model to mine the historical data of the power quality of photovoltaic power station and realize the awareness of future changes in power quality of photovoltaic power station. Major studies include:

(1) **Propose power quality situation awareness framework of photovoltaic power station, and complete the extraction of power quality elements and situation understanding of photovoltaic power station.** It is proposed that photovoltaic power station power quality situation awareness consists of three stages: photovoltaic power station power quality elements extraction, photovoltaic power station situation understanding and photovoltaic power station power quality situation prediction. Introducing the process of photovoltaic power station power quality elements extraction, using autocorrelation coefficient and maximum information coefficient analysis methods to understand the situation of power quality data, conducting value standardization and outlier processing are performed on power quality data and external variable data, and two different methods are used to comprehensively assess the model effect for each of the three power quality indicators.

(2) **Construction of power quality situation prediction model based on Deep Belief Network (DBN) and Support Vector Regression (SVR) method and perform a case study.** Expounds the principle of using DBN and SVR method for time series

data prediction, constructs a single model of power quality situation prediction based on the DBN and SVR method, and introduces the implementation process of the two models. Taking the measured data of voltage, frequency and voltage total harmonic distortion rate of a photovoltaic power station in a certain place for case study, analyze the effect of two single models on power quality situation prediction of photovoltaic power station.

(3) A power quality situation prediction model of photovoltaic power station based on hybrid deep neural network is proposed. In response to the poor prediction effect of a single model of DBN and SVR, a hybrid deep neural network situation prediction model of Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) is proposed. First extract the high-level features of power quality input data through CNN, and then GRU is used to learn the long-time power quality sequence before and after relationship, which overcomes the shortcomings of the traditional single model with limited prediction accuracy and weak generalization ability, and solves the problems of difficulty in extracting high-level features of massive power quality data and the easy gradient disappearance of long-time sequence information learned by the model. Through experimental comparison with several other models, the hybrid model has higher prediction accuracy and better timeliness.

Keywords: Photovoltaic power station; Power quality; Situational awareness; Convolutional neural networks; Gated recurrent unit

缩略语对照表

缩略语	英文全称	中文名称
PMU	Phasor Measurement Unit	同步相量测量装置
LSTM	Long Short-term Memory Network	长短期记忆网络
ARIMA	Autoregressive Integrated Moving Average Model	累积式自回归移动平均模型
SVR	Support Vector Regression	支持向量回归
BP	Back Propagation	反向传播
KPCA	Kernel Principal Component Analysis	核主成分分析
GA	Genetic Algorithm	遗传算法
LSSVM	Least Square Support Vector Machine	最小二乘支持向量机
CNN	Convolutional Neural Networks	卷积神经网络
RNN	Recurrent Neural Network	循环神经网络
XGBoost	eXtreme Gradient Boosting	极端梯度提升
DTW	Dynamic Time Warping	动态时间规整算法
GRU	Gated Recurrent Unit	门控循环单元
DBN	Deep Belief Network	深度置信网络
MIC	Maximal Information Coefficient	最大信息系数
MI	Mutual Information	互信息
MAE	Mean Absolute Error	平均绝对误差
RMSE	Root Mean Square Error	均方根误差
DAAPE	Daily All Absolute Percentage Error	日总绝对百分比误差
NRMSE	Normalized Root Mean Square Error	标准均方根误差
MAPE	Mean Absolute Percentage Error	平均绝对百分比误差
GS	Grid SearchCV	网格搜索法
RBM	Restricted Boltzmann Machine	受限玻尔兹曼机
EEMD	Ensemble Empirical Mode Decomposition	集合经验模态分解

主要参数符号对照表

t	某一时刻
V_t	电压时间序列
F_t	频率时间序列
THD_t	总谐波畸变率时间序列
P_t	有功功率时间序列
R_t	太阳辐照度时间序列
r	采样频率
C	自相关系数
l	滞后时间点
p	概率
$B(n)$	网格数目上限
v_i^{act}	第 i 个时间点电压实际值
v_i^{pre}	第 i 个时间点电压预测值
f_i^{act}	第 i 个时间点频率实际值
f_i^{pre}	第 i 个时间点频率预测值
THD_i^{act}	第 i 个时间点总谐波畸变率实际值
THD_i^{pre}	第 i 个时间点总谐波畸变率预测值
r_t	重置门
z_t	更新门
x_t	输入信息
h_{t-1}	前一时刻隐含层状态
σ	sigmoid 激活函数
f	tanh 激活函数
\tilde{h}_t	候选状态
h_t	隐含层输出信息
\odot	点乘运算
l	代价函数
ξ_i	松弛变量
ξ_i^*	松弛变量
α_i	拉格朗日乘子
α_i^*	拉格朗日乘子

目录

摘要.....	I
ABSTRACT.....	II
缩略语对照表.....	IV
主要参数符号对照表.....	V
第 1 章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	2
1.3 本文主要研究内容及结构.....	4
1.3.1 主要研究内容.....	4
1.3.2 文章结构.....	5
第 2 章 光伏电站电能质量态势感知方法研究.....	7
2.1 光伏电站电能质量态势感知架构.....	7
2.2 光伏电站电能质量要素提取.....	7
2.3 光伏电站电能质量态势理解.....	8
2.3.1 电能质量自相关性理解.....	8
2.3.2 变量因素相关性理解.....	9
2.3.3 异常值处理.....	9
2.3.4 数据标准化.....	10
2.3.5 模型评估.....	10
2.4 算例分析.....	12
2.4.1 电能质量自相关性理解.....	12
2.4.2 变量因素相关性理解.....	14
2.5 本章小结.....	17
第 3 章 基于 DBN 及 SVR 模型的光伏电站电能质量态势预测.....	18
3.1 DBN 模型构建及实现.....	18
3.1.1 深度置信网络 (DBN).....	18
3.1.2 模型构建及训练.....	19
3.1.3 实现流程.....	22
3.1.4 算例分析.....	22
3.2 SVR 模型构建及实现.....	28
3.2.1 支持向量回归 (SVR).....	28

3.2.2	模型构建	29
3.2.3	实现流程	29
3.2.4	算例分析	30
3.3	本章小结	34
第 4 章	基于混合深度神经网络模型的光伏电站电能质量态势预测.....	35
4.1	深度神经网络	35
4.1.1	卷积神经网络 (CNN)	35
4.1.2	门控循环单元 (GRU)	36
4.2	CNN-GRU 混合模型构建	37
4.2.1	电压预测模型构建	37
4.2.2	频率预测模型构建	39
4.2.3	电压总谐波畸变率预测模型构建	40
4.3	CNN-GRU 混合模型训练	42
4.3.1	电压预测模型训练	43
4.3.2	频率预测模型训练	45
4.3.3	电压总谐波畸变率预测模型训练	46
4.4	实现流程	48
4.5	算例分析	49
4.5.1	电压预测	49
4.5.2	频率预测	53
4.5.3	电压总谐波畸变率预测	55
4.6	本章小结	57
第 5 章	总结与展望	58
5.1	总结	58
5.2	展望	59
	参考文献	60
	致谢	64
	个人简历、攻读硕士学位期间的学术成果及科研项目	65

第 1 章 绪论

1.1 研究背景及意义

传统火力发电所使用的煤炭能源面临着资源紧缺和环境污染两大问题，而以光伏太阳能为代表的可再生能源则有着储量巨大和清洁绿色的优点。我国在过去几十年间，由于大规模的开采煤炭资源和使用化石能源，使得一些城市出现了煤炭能源告急、拉闸限电和雾霾环境污染等一系列问题。从全球角度上看，全球气候变暖、海平面上升等正在威胁着人类赖以生存的环境，如何着手降低碳排放以共同应对地球气候变化成为世界各国正在面临和亟待解决的问题。

为了共同应对全球气候环境变化问题，确保实现全球社会可持续发展，我国在第 75 届联合国大会提出了“双碳”的目标。2021 年，中央有关部门发布的《关于完整准确全面贯彻新发展理念做好碳达峰碳中和工作的意见》明确指出，到 2030 年要实现单位 GDP 的 CO₂ 排放量相比 2005 年降低五分之三以上，非化石能源的使用占比达到四分之一左右，太阳能、风力发电总装机容量达到 12 亿千瓦以上的目标^[1]。这表明，实现能源体系的低碳转型、不断提高清洁绿色能源占比将成为未来十年间推动我国能源供给侧改革工作的重点。

能源作为一个国家赖以生存的根基，其重要性不言而喻。太阳能作为一种储量非常丰富的清洁绿色能源，是目前世界各国都在争相发展的最重要的能源之一。在欧洲一些发达国家，光伏发电量占总发电量比例达 10% 以上。我国作为世界上太阳能发电量最多的国家，近年来可再生能源发展也十分迅速，根据国家统计局公布的 2021 年前三季度数据，我国太阳能发电量增速有 10.2%，为 1362.2 亿千瓦时，太阳能发电继续保持着较快的增长势头^[2]。同时，在面临疫情巨大冲击的 2020 年，我国依旧实现了光伏并网新增装机容量 4820 万千瓦，截至 2020 年底，光伏并网的装机容量达到了 2.5 亿千瓦，占我国发电装机总容量的 11.5%，均实现了历史性突破，“十四五”期间我国光伏并网的装机容量预计将有更进一步的生长^[3]。发展以光伏发电为代表的分布式发电系统有助于改善我国的能源系统结构，增强电网系统应对极端天气灾害的可靠性。

然而，在光伏等可再生能源并网容量的持续攀升背后，其带来的负面影响也激起了世界各国人员的广泛关注。光伏电站在将直流逆变成交流并网的过程中使用大量电力电子逆变并网装置，会产生谐波污染等电能质量问题；光伏电站由于受日光照射的影响较大而产生出力的随机性和不稳定性，易引起并网点电压大幅波动、电压波形畸变率增大、频率偏差增大等电能质量问题，在并网运行时对配电网电能质量造成污染，严重影响配电网的可靠安全运行^[4]。如果不能及时发现

并解决因光伏电站运行产生的电能质量问题,则可能造成配电网出现一定的安全事故,对电网和用户侧设备造成损坏等严重影响,例如谐波会造成电网系统设备如电动机等损耗增加、过热等问题;电压正偏过高或负偏过高会导致绝缘设备损坏、用电设备效率降低甚至损坏等问题;频率偏低则会使得电动机转速降低进而影响产品质量等^[5],这些问题一旦产生会使得电网不能很好地消纳光伏发出的电能,并出现“弃光”等浪费能源、不利于可再生能源发展的现象。因此,研究通过挖掘光伏电站电能质量的历史数据信息提前感知电能质量变化态势,及早发现光伏电站存在的电能质量问题,为电能质量提前治理预留出时间并提供准确信息,极大降低光伏电站并网运行对配电网电能质量造成的污染,成为可再生能源并网可靠运行的关键技术之一^[6]。通过选用合适且性能优异的方法快速准确感知光伏电站电能质量变化态势,为光伏电站电能质量的诊断、治理和决策控制提供坚实基础和前提,是改善光伏电站电能质量的有效途径,对光伏电站并网时配电网的可靠稳定运行有着十分重要的作用。

1.2 国内外研究现状

态势感知概念最早应用在航空军事领域,指在特定的时空环境中,对动态环境中的各种因素提取、分析并预测^[7-8]。文献[9]给出了电力系统领域态势感知的关键技术和架构,为态势感知技术在智能配电网中的应用奠定了理论基础。文献[10]则给出了态势感知技术在需求侧负荷控制的涵义及技术,并对其在某些场合的应用前景进行了分析,指明了后续的推进方向。文献[11]利用历史数据对配电网将来状态进行建模感知其运行趋势,通过仿真验证了态势感知方法在三相状态估计速度及精度方面的优越性。文献[12]利用同步相量测量装置(Phasor Measurement Unit, PMU)、支持向量数据描述算法、长短时记忆网络(Long Short-term Memory Network, LSTM)算法实现了对配电网安全态势的建模、异常数据分析和预测,并在IEEE33节点系统中证实了方法的有效性。

随着可再生能源并网装机容量的提升和以微电网为代表的分布式发电的不断发展,有关电能质量方面的研究也在不断增加,文献[13]提出了基于电能质量态势感知信息的分布式发电马尔可夫决策方法,利用态势感知方法预测未来时刻的电能质量,并将感知信息作为后续运行决策控制的输入来改善电能质量。文献[14]提出一种基于数据挖掘的配电站区域电能质量态势感知方法,利用基于层间相关性的客观赋权法和基于灰色关联的优劣解距离法进行电能质量态势理解,构建基于改进集合经验模态分解(Ensemble Empirical Mode Decomposition, EEMD)和LSTM的模型进行电能质量态势预测,但该方法训练时间太长。目前关于电能质量的研究主要还集中在电能质量监测、电能质量评价、电能质量扰动及分析等

方面^[15-22]，对于电能质量态势预测方面的研究还比较少。传统预测方法主要是以累积式自回归移动平均模型（Autoregressive Integrated Moving Average Model, ARIMA）^[23-24]为代表的时间序列预测法和以 SVR^[25-26]为代表的人工智能方法等。文献[27]利用 ARIMA 模型预测有功功率，然后根据有功功率与电能质量的关系利用反向传播（Back Propagation, BP）算法对短时波动与闪变、电压偏差、总谐波畸变率等 5 种电能质量指标进行预测，结合了 ARIMA 模型适合处理时序性较强数据和 BP 算法处理非线性序列较好的特点。文献[28]提出了一种结合线性回归模型、随机时间序列模型和灰色模型的电压偏差及谐波畸变率电能质量组合预测模型，采用方差-协方差方法组合三种模型的预测结果，相比只使用其中一种模型整体预测准确度更高。文献[29]利用 BP 算法预测含可再生能源的离网系统中电压总谐波畸变率、电流总谐波畸变率、长时闪变严重程度和频率四种电能质量指标，试验结果显示误差大体在可接受程度范围内。文献[30]使用 Spearman 和 Pearson 系数计算电能质量和天气因素的相关性，将天气因素、有功负荷也作为输入变量，提出一种基于改进核主成分分析（Kernel Principal Component Analysis, KPCA）结合 GA-BP 算法的电压偏差及负序不平衡度指标预测方法，结果比传统 BP 网络和 KPCA+BP 方法表现更好，但该方法包含的算法繁杂，且 BP 神经网络预测准确度有限。文献[31]利用混沌理论对长时闪变、三相电压不平衡度等电能质量历史数据进行相空间重构，在此基础上结合最小二乘支持向量机（Least Square Support Vector Machine, LSSVM）得到优选电能质量预测模型，预测准确度高于 BP 算法。文献[32]定义了三个不同的标准来评价所有电能质量参数组合的预测模型的性能，并采用随机森林模型进行电压质量预测。但文献[28][31]未对与电能质量变化相关的变量进行分析，电能质量的变化不仅与其自身历史数据相关，还可能与外部因素变量相关。文献[27][29][32]虽考虑了电能质量与其他变量之间的关系，但变量的选择仅凭人为经验选取，未从变量间的实际时序相关性出发进行特征筛选，特征维数过高导致训练模型复杂度变高且泛化能力较差，特征维数过低则导致预测准确度下降。文献[33]采用 Relief 算法进行输入特征的相关性理解及筛选，然后利用灰色关联分析将 Elman 网络、随机森林和 RBF-SVM 三种方法的结果组合实现对电压偏差和电压谐波畸变率的预测，结果优于未进行特征相关性理解及筛选单一方法。

近年来，深度学习一直在工业界及学术界处于火热发展的状态，各种深度神经网络被用于时间序列预测，如 CNN^[34-35]、循环神经网络（Recurrent Neural Network, RNN）^[36-37]、LSTM^[38-39]等，并在负荷和功率预测领域获得了很大的成功^[40-41]。文献[42]提出一种以 XGBoost 算法和 LSTM 作为差异化基学习器的 Stacking 集成学习架构用于负荷预测，以 ENTSO 的负荷数据集进行实验的结果

表明该方法优于传统 SVM 模型，但 Stacking 的架构设计复杂使得模型消耗时间过多。文献[43]提出一种结合 CNN 和 LSTM 的短期风电功率预测方法，先利用 CNN 提取风电功率等数据的高层特征再连接至 LSTM 网络实现对风电功率序列的预测，相比未经高层特征提取的单一 LSTM 模型风电功率拟合准确度更高。对于电能质量预测，文献[44]对电压偏差数据利用 K-means 聚类算法进行基于欧式距离的分类，然后利用 LSTM 网络实现对分类后的电压偏差数据的预测，通过 IEEE-13 节点主动配电网进行算例仿真，结果显示该方法相较于 ARIMA 和 BP 模型在相对误差值上的分布更低。文献[45]利用动态时间规整算法(Dynamic Time Warping, DTW)筛选与稳态电压质量序列相关程度高的有功功率序列，将两种数据作为 LSTM 模型的输入，以福建某地变电站监测数据进行分析，DTW-LSTM 方法比单输入 LSTM 模型的电压偏差预测误差更低。但 LSTM 也存在模型参数较多、训练速度较慢的问题^[46-47]，GRU 作为 RNN 的一种变体，是 LSTM 的改进版本，相比 LSTM 的输入门、输出门、遗忘门三个门结构，GRU 只有更新门和重置门，在保证预测精度的同时模型参数更少，训练速度更快，在时效性要求较高的场合更有优势^[48]。文献[49]提出一种基于 GRU-NN 模型的短期电力负荷预测方法，结果表明，GRU 模型相比 LSTM 模型在预测准确度接近的同时训练耗时有明显的降低。

综合上述国内外研究现状可以看出，目前关于电能质量态势感知的研究文献仅有两篇，其他研究主要集中在电能质量态势预测方面，在预测方法上也以单输入电能质量时间序列和单模型方法居多，随着大数据时代数据量和数据维度的增长，传统的单输入特征和单模型方法显然已不能满足需求，在这种情况下，已有相关学者开始研究多维序列输入多模型融合方法下的负荷时序预测^[50]，因此，探索融合不同深度神经网络模型进行电能质量序列态势预测已渐渐成为一种必然趋势。

1.3 本文主要研究内容及结构

1.3.1 主要研究内容

在光伏发电并网装机容量不断提升的背景下，针对光伏电站受日光照射影响较大而产生的出力不确定性易造成并网点电能质量大幅波动的问题，本文提出了一种基于 CNN-GRU 混合模型的光伏电站电能质量态势感知方法。首先提出了光伏电站电能质量态势感知的架构分为光伏电站电能质量要素提取、光伏电站电能质量态势理解和光伏电站电能质量态势预测三个阶段。在光伏电站电能质量要素提取阶段先通过在光伏电站端安装电能质量参数采集模块和传感器单元获取原

始数据, 然后以一定间隔时间提取数据构建时间序列; 在光伏电站电能质量态势理解阶段通过计算电能质量序列自相关系数进行电能质量自相关性理解, 得出电能质量序列在时序上的关联性以指导预测模型的输入步长选择, 通过计算电能质量与外部变量间的最大信息系数进行变量因素相关性理解, 得出与电能质量序列趋势变化相关程度高的外部变量因素以指导预测模型输入维度选择; 在光伏电站电能质量态势预测阶段提出一种 CNN-GRU 混合深度神经网络模型, 将经过态势理解阶段得到的多维时间序列输入 CNN-GRU 混合模型实现对电能质量未来变化态势的预测, 结合了 CNN 的海量数据高层特征提取优势和 GRU 学习长时间序列前后信息的优势, 充分挖掘了光伏电站电能质量数据信息。以某地光伏电站电能质量实测数据进行算例分析, 对比单神经网络模型和传统 SVR 模型, 验证了本文所提 CNN-GRU 混合模型的优越性。

1.3.2 文章结构

第一章: 指出本课题的研究背景与意义, 深入分析了国内外在电能质量态势感知方面的研究现状, 阐述了本文的主要研究内容及文章架构。

第二章: 提出光伏电能质量态势感知的架构分为光伏电站电能质量要素提取、光伏电站电能质量态势理解和光伏电站电能质量态势预测三个阶段; 分析阐述光伏电站电能质量要素提取具体实现过程; 通过计算电能质量序列自相关系数和电能质量与变量因素间的最大信息系数理解电压、频率、电压总谐波畸变率三种电能质量序列的自相关性及其与变量因素间相关程度; 进行数据异常值和标准化处理, 针对三种不同电能质量数据提出分别使用不同的模型评估指标。

第三章: 利用枚举法、网格搜索法和控制变量法构建和训练基于 DBN 和 SVR 两种传统机器学习预测方法的电能质量态势预测模型, 以某地光伏电站电压、频率和电压总谐波畸变率三种电能质量实测数据在不同测试集下进行实验, 分析这两种方法用于光伏电站电能质量态势预测的性能表现。

第四章: 提出一种 CNN-GRU 混合神经网络电能质量态势预测模型。介绍了 CNN 及 GRU 深度神经网络原理; 构建和训练电压、频率、电压总谐波畸变率三种电能质量混合神经网络模型, 利用枚举法、网格搜索法和控制变量法设计混合模型结构及参数; 使用和第三章同样的光伏电站电能质量实测数据进行实验, 并与单一深度神经网络模型和 SVR 模型做对比, 验证了本文所提方法的优越性。

第五章: 总结全文所做研究及各章节研究内容, 指出本文的不足之处及后续完善方向。

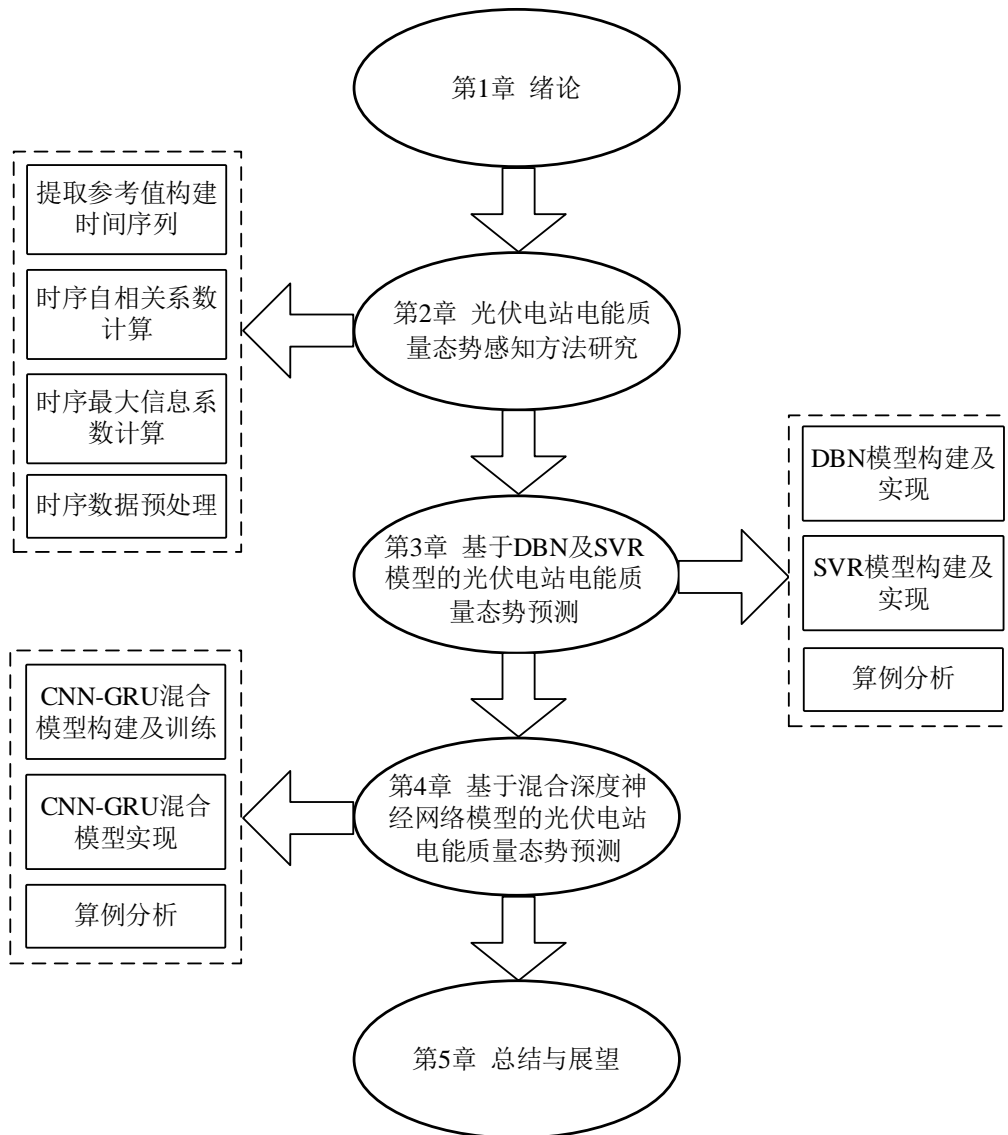


图 1-1 全文章节结构图

第 2 章 光伏电站电能质量态势感知方法研究

2.1 光伏电站电能质量态势感知架构

光伏电站电能质量态势感知分为光伏电站电能质量要素提取、光伏电站电能质量态势理解和光伏电站电能质量态势预测三个阶段，其架构如图 2-1 所示。

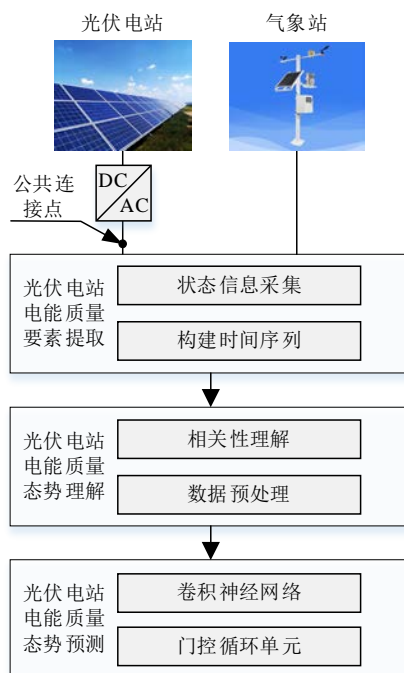


图 2-1 光伏电站电能质量态势感知架构

光伏电站电能质量要素提取是指在动态环境中，面向光伏电站端实现状态信息采集，包括实时电能质量数据、动态环境信息等，以此信息构建时间序列，是实现电能质量态势感知的数据基础。

光伏电站电能质量态势理解是指基于计算机系统管理平台，针对电能质量要素提取所提供的数据信息中隐含的相关性关系、数据特殊性进行深度挖掘，充分理解电能质量变化规律，是对提取信息的深度理解。

光伏电站电能质量态势预测是指通过智能算法学习经态势理解后的数据高层特征，挖掘长时间序列前后关系，实现电能质量发展趋势预测。

2.2 光伏电站电能质量要素提取

通过在光伏电站端安装电能参数采集模块、传感器单元，构建时刻 t 电压、频率、总谐波畸变率三种电能质量时间序列、有功功率时间序列和太阳辐照度等气象时间序列：

$$V_t = [v_1, v_2, \dots, v_j, \dots]^T, j \in \mathbf{N} \quad (2-1)$$

$$F_t = [f_1, f_2, \dots, f_j, \dots]^T, j \in \mathbf{N} \quad (2-2)$$

$$THD_t = [THD_1, THD_2, \dots, THD_j, \dots]^T, j \in \mathbf{N} \quad (2-3)$$

$$P_t = [p_1, p_2, \dots, p_j, \dots]^T, j \in \mathbf{N} \quad (2-4)$$

$$R_t = [r_1, r_2, \dots, r_j, \dots]^T, j \in \mathbf{N} \quad (2-5)$$

基于电能参数采集模块的采样频率 r ，根据态势预测时间尺度对所采集的数据以一定间隔时间 $\Delta t = 300 / r$ 提取时间序列中 Δt 时间段内的平均值作为参考值，以 $t_i \sim t_{i+1} (i \geq 0)$ 时间段电压和有功功率时间序列为例，有：

$$V_{t_i \sim t_{i+1}} = [V_1, V_2, \dots, V_m]^T, m \leq \frac{t_{i+1} - t_i}{\Delta t} \quad (2-6)$$

$$P_{t_i \sim t_{i+1}} = [P_1, P_2, \dots, P_m]^T, m \leq \frac{t_{i+1} - t_i}{\Delta t} \quad (2-7)$$

式 (2-6)、式 (2-7) 中， V_1 和 P_1 分别对应时间段 $t_i + \Delta t$ 内的电压平均值和有功功率平均值， V_2 和 P_2 分别对应时间段 $t_i + \Delta t \sim t_i + 2\Delta t$ 内的电压平均值和有功功率平均值，以此类推得出模型的电压输入时间序列和有功功率输入时间序列。

2.3 光伏电站电能质量态势理解

虽然光伏电站由于受日光照射影响而出力不确定使得其电能质量大幅随机波动，但电能质量数据作为一种时序数据仍可能存在一些态势变化的规律，同时光伏电站电能质量的变化也与外部变量因素如气象因素等关系密切，因此在对电能质量这种时序数据进行态势理解时既要挖掘电能质量态势变化的内在规律，又要考虑其他外部变量因素对其态势变化的影响^[51]。

2.3.1 电能质量自相关性理解

电能质量数据作为一种时序数据，首先需要分析其在时序上的相关性，以提供态势预测模型的输入时间步长获得最佳预测效果，即用距当前时刻多少分钟内的历史数据作为输入。自相关系数 $C(l)$ 可以度量在两个不同时间点的电能质量时序数据相关程度。 $C(l)$ 的取值范围为 $[-1, 1]$ ，若 $C(l) > 0$ ，则两电能质量时序数据正相关，反之则负相关； $|C(l)|$ 越大，相关程度越高。公式如式 (2-8) 所示。

$$C(l) = \frac{\sum_{t=1}^{m-l} (q_t - \bar{q})(q_{t+l} - \bar{q})}{\sum_{t=1}^m (q_t - \bar{q})^2} \quad (2-8)$$

式 (2-8) 中, $C(l)$ 为自相关系数, l 是滞后时间点, q_t 表示时刻 t 的电能质量值, q_{t+p} 表示时刻 $t+p$ 的电能质量值, \bar{q} 是整个电能质量序列均值, m 为总时刻。

2.3.2 变量因素相关性理解

光伏电站的电能质量变化可能与多种影响因素有关, 为充分考虑其他变量因素的影响, 指导态势预测模型输入维度选取, 提升预测精度, 需进行变量因素相关性理解。目前多数文献中都是采用 Pearson、Spearman 等线性相关性理解方法, 未考虑变量因素间的非线性关系, 可能导致模型输入维度冗余或缺失^[52]。最大信息系数 (Maximal Information Coefficient, MIC) 是一种优秀的变量关联性计算方式, 用于衡量两个变量之间的线性及非线性关联强度。MIC 不需要对数据分布进行任何假设来评估变量之间的函数和统计关系, 具有计算复杂度低、鲁棒性强的特点。MIC 的基本思想是, 如果变量之间存在相关性, 可以在两个变量的散点图上画一个网格来划分和封装它们之间的相关性, MIC 通过互信息 (Mutual Information, MI) 和网格划分方法计算^[53]。给定两个具有离散值的随机变量, $x = \{x_i, i = 1, 2, \dots, n\}$ 和 $y = \{y_i, i = 1, 2, \dots, n\}$, 其中 n 是这些变量中元素的数量, 互信息 $I(x, y)$ 被定义为:

$$I(x, y) = \int p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} dx dy \quad (2-9)$$

对互信息进行标准化, 求出最大信息系数如下:

$$MIC(x, y) = \max_{a*b < B(n)} \frac{I(x, y)}{\log_2 \min(a, b)} \quad (2-10)$$

式 (2-10) 中, $p(x, y)$ 为变量 x 和 y 之间的联合概率, a, b 是在 x, y 轴方向上划分格子的数目, $B(n)$ 为划分网格数目上限值, 一般为数据量 n 的 0.6 次方左右。MIC 取值区间为 $[0, 1]$, 其值越大, 电能质量与外部变量因素间的相关程度越高。

2.3.3 异常值处理

态势预测模型输入序列数据的正确和完整性重要影响着模型的预测准确度, 但由于历史电能质量数据和外部因素数据通过硬件设备采集而得, 受硬件设备本身质量稳定性、硬件设备所处环境对数据传输的影响, 所采集的电能质量数据或外部因素数据可能存在明显偏离或缺失等情况, 称之为异常数据。如果电能质量异常数据偏离过大或缺失值较多, 会明显改变电能质量时间序列的平均值和标准差, 使得某些异常值在模型训练过程中贡献度发生变化进而影响模型的训练效果, 造成模型预测误差增大, 达不到相应的准确度要求。因此有必要将这些异常数据

在模型训练前即进行处理，消除错误数据对模型造成的影响。

首先对于异常值的检出常用无监督学习方法，含有数据密度和聚类分析，即根据时序数据本身的特性以电能质量序列数据密度和电能质量序列数据间的距离对正常值和异常值做出辨别。其次对于异常采集值的处理，常用的处理方法包括去除异常采集值、邻近值插补异常采集值、局部加权回归插补异常采集值、中位数插补异常采集值、众数插补异常采集值、均值插补异常采集值等，为保证电能质量数据和外部因素数据在时序上的连续性，局部加权回归插补法对缺失值进行填补是更为有效的方法^[54]，对于异常偏离值则是先去除后局部加权回归插补。

2.3.4 数据标准化

在通过相关性理解得出数据输入维度之后，如果直接将原始数据输入模型进行计算会使得梯度下降求解速度变慢，模型收敛时间变长^[55]，同时，由于不同维度数据之间的量纲和数量级不一样，直接将其输入模型会导致某些维度数据在模型训练中贡献度过大，且在训练迭代的过程中可能导致计算溢出^[56]，因此为了提高模型预测准确度和收敛速度，有必要对电能质量数据和外部因素数据进行值标准化处理，计算如下所示：

$$q_t' = \frac{q_t - q_{\min}}{q_{\max} - q_{\min}} \quad (2-11)$$

式(2-11)中， q_t' 表示时刻 t 的值标准化数据， q_t 表示时刻 t 的原始数据， q_{\max} 和 q_{\min} 分别表示各维度时序数据中的最大值和最小值。

2.3.5 模型评估

评估模型的好坏是电能质量预测值相对电能质量真实值的偏离大小和模型稳定性两个方面的综合评价，具体来说即误差用来衡量模型预测值偏离大小，方差用来衡量模型预测稳定性，这样能更好体现模型的优劣。本文对模型预测电压、频率、总谐波畸变率三种电能质量序列分别选择合适的误差评估指标和方差评估指标以衡量模型在预测准确度和鲁棒性两方面的表现。

(1) 模型预测电压质量评估

针对电压数据的数量级和变化特点采用平均绝对误差(Mean Absolute Error, MAE)和均方根误差(Root Mean Square Error, RMSE)来评估模型的表现，能更好地体现模型的优劣。计算公式如式(2-12)、式(2-13)所示。

$$V_{MAE} = \frac{1}{n} \sum_{i=1}^n |v_i^{act} - v_i^{pre}| \quad (2-12)$$

$$V_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (v_i^{act} - v_i^{pre})^2} \quad (2-13)$$

式 (2-12)、式 (2-13) 中, n 是预测日电压样本总数, v_i^{act} 为预测日第 i 个时间点的电压实际值, v_i^{pre} 为预测日第 i 个时间点的电压预测值。

(2) 模型预测频率质量评估

根据频率数据的数量级和波动频次、幅度大小, 如果使用 MAE 和 RMSE 评估模型优劣的话, MAE 和 RMSE 在数值上会变得非常小, 无法直观对比出各频率预测模型的效果。为此, 本文提出一种日总绝对百分比误差 (Daily All Absolute Percentage Error, DAAPE) 指标来评估频率预测模型的表现, 同时为全面评估频率预测模型的表现再使用频率标准均方根误差 (NRMSE) 指标。计算公式如 (2-14)、式 (2-15) 所示。

$$F_{DAAPE} = \sum_{i=1}^n \frac{|f_i^{act} - f_i^{pre}|}{f_i^{act}} \times 100\% \quad (2-14)$$

$$F_{NRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (f_i^{act} - f_i^{pre})^2}}{f^{\max} - f^{\min}} \quad (2-15)$$

式 (2-14)、式 (2-15) 中, n 是预测日频率样本总数, f_i^{act} 为预测日第 i 个时间点的频率实际值, f_i^{pre} 为预测日第 i 个时间点的频率预测值, f^{\max} 和 f^{\min} 分别为预测日频率实际值的最大和最小值。

(3) 模型预测总谐波畸变率评估

根据总谐波畸变率的数量级和波动幅值的特点, 采用平均绝对百分比误差 (MAPE) 和标准均方根误差 (NRMSE) 作为模型的评估指标能更好对比各模型预测总谐波畸变率的情况。计算公式如 (2-16)、式 (2-17) 所示。

$$T_{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|THD_i^{act} - THD_i^{pre}|}{THD_i^{act}} \quad (2-16)$$

$$T_{NRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (THD_i^{act} - THD_i^{pre})^2}}{THD^{\max} - THD^{\min}} \quad (2-17)$$

式 (2-16)、式 (2-17) 中, n 是预测日总谐波畸变率样本总数, THD_i^{act} 为预测日第 i 个时间点的总谐波畸变率实际值, THD_i^{pre} 为预测日第 i 个时间点的总谐波畸变率预测值, THD^{\max} 和 THD^{\min} 分别为预测日总谐波畸变率实际值的最大和最小值。

2.4 算例分析

在本文中，基于澳大利亚 DKA 光伏电站 2020 年 1 至 5 月的电能质量及外部变量因素数据^[57]进行电能质量自相关性理解及变量因素相关性理解。包括并网点电压、频率、电压总谐波畸变率三种电能质量数据以及太阳辐照度、环境温度、相对湿度、有功功率、风速、降雨量等外部变量因素。同时因为缺少该光伏电站电流总谐波畸变率数据，本文仅对电压总谐波畸变率进行了实验。

2.4.1 电能质量自相关性理解

(1) 电压自相关性理解

通过电压自相关性理解能够指导电压态势预测模型的最佳输入步长选择，电压自相关系数如图 2-2 所示。

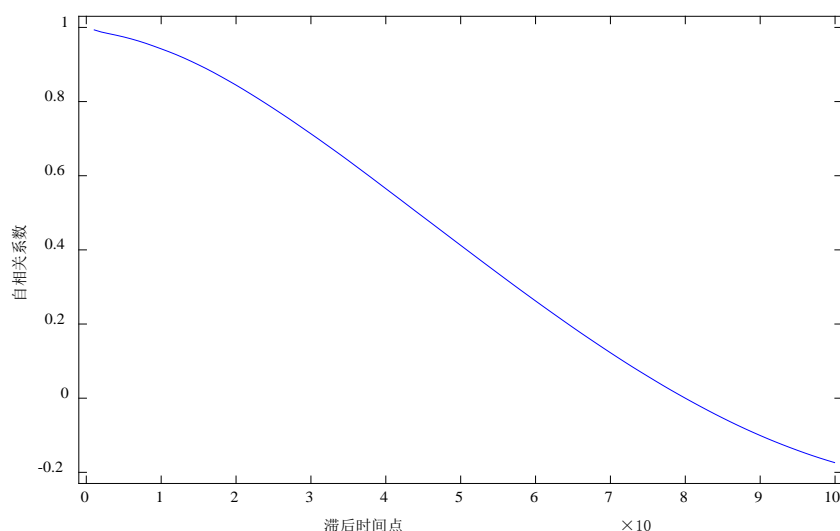


图 2-2 电压自相关系数曲线

从图 2-2 可以看出，电压自相关系数曲线整体呈现头部斜率平缓而后斜率增加的趋势，自相关程度高的电压时间序列主要集中在滞后时间较短的区间，且越靠近当前时刻自相关程度越高，因此考虑将当前时刻前几十分钟内的真实电压数据作为输入，通过枚举 1 至 10 的值确定模型输入时间步长为 6，即只用距当前时刻 6 个时间步的电压数据预测未来时刻的电压数据。

(2) 频率自相关性理解

频率自相关系数如图 2-3 所示。

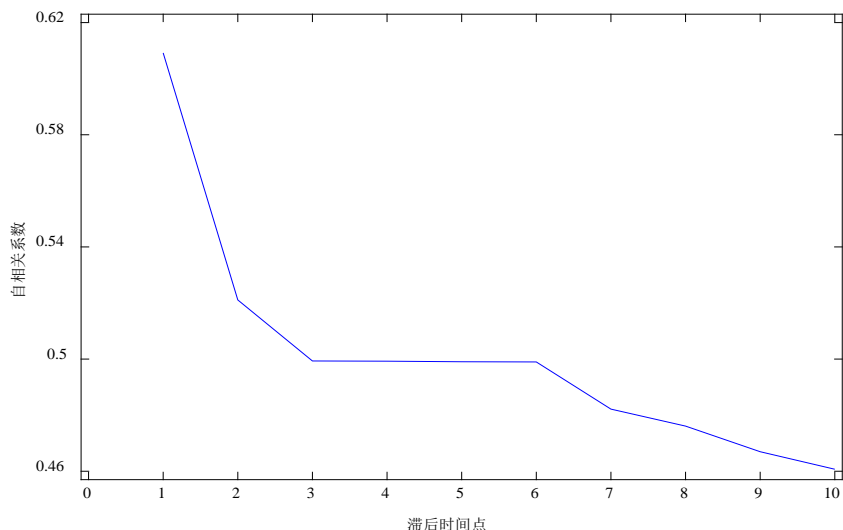


图 2-3 频率自相关系数曲线

从图 2-3 可以看出，频率自相关程度在一阶滞后的时候为最大值，在二阶滞后时自相关系数下降比较明显，当滞后三阶时频率自相关系数 $C(3) \approx 0.499$ ，相关程度一般，因此确定频率态势预测模型输入时间步长为 2。

(3) 电压总谐波畸变率(THDu)自相关性理解

电压总谐波畸变率自相关系数如图 2-4 所示。

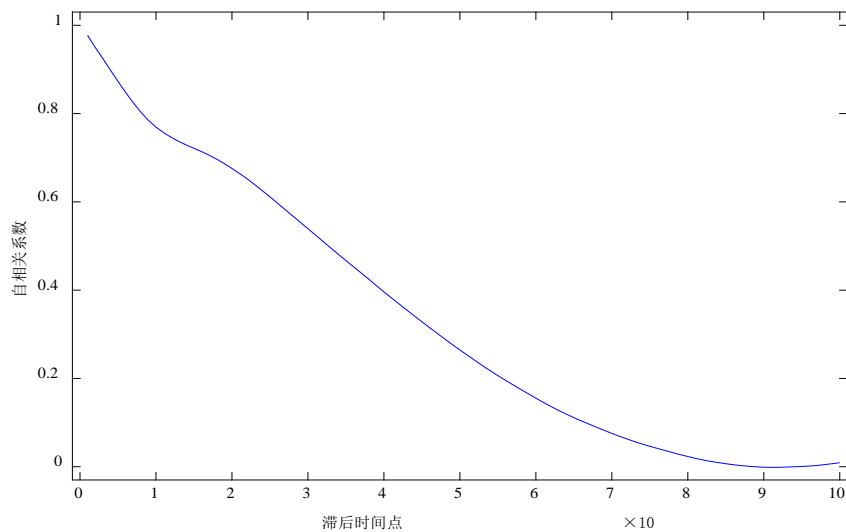


图 2-4 THDu 自相关系数曲线

从图 2-4 可以看出，电压总谐波畸变率自相关程度总体呈现出开始下降斜率较大而后有一个短暂的相对平缓下降再下降斜率增大的过程，相关性程度最强的电压总谐波畸变率时间序列同样集中在滞后 10 个时间点之内，通过枚举 1 至 10 的值确定电压总谐波畸变率态势预测模型输入时间步长为 4。

2.4.2 变量因素相关性理解

(1) 电压与变量因素相关性理解

为分析不同变量因素对电压态势预测的影响，本文分别计算太阳辐照度、环境温度、相对湿度、有功功率、风速和降雨量等 6 种变量因素与电压之间的 MIC，结果如热力图 2-5 所示。

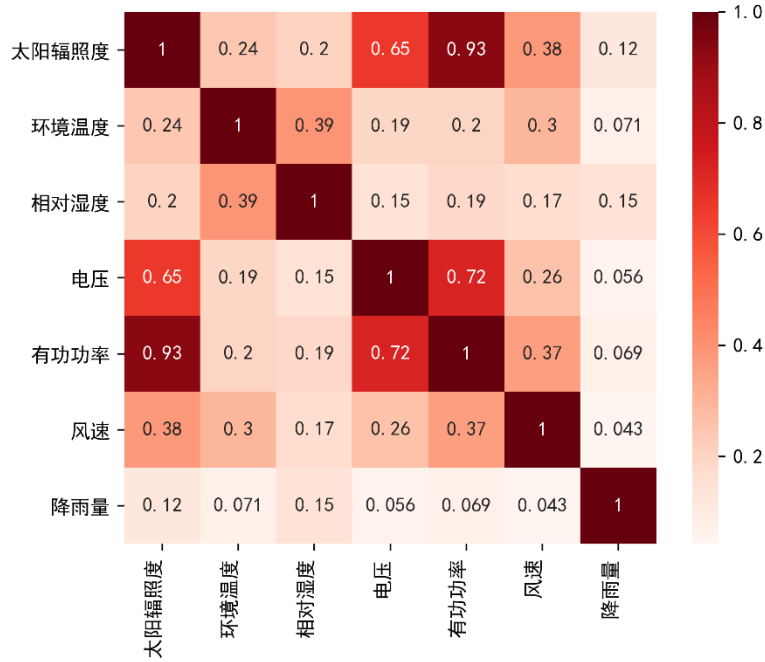
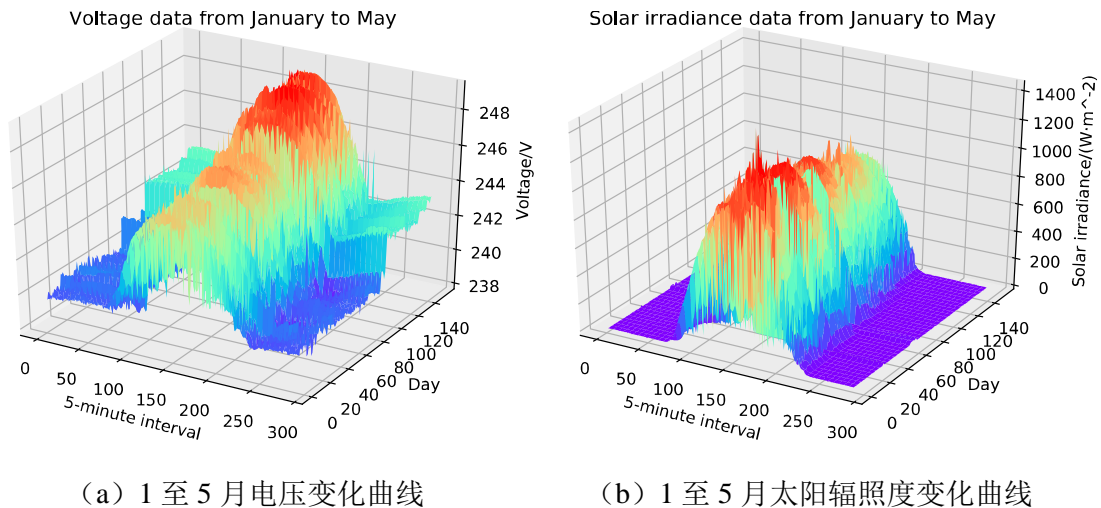
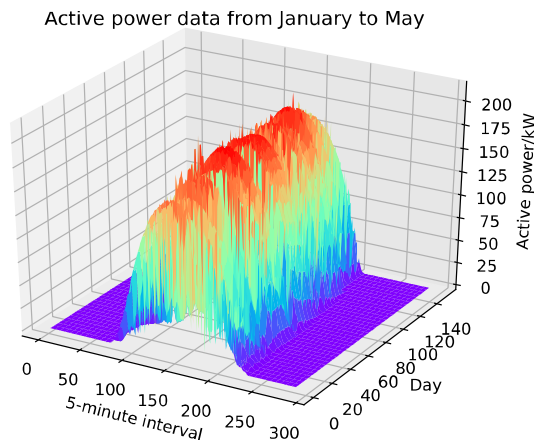


图 2-5 电压与其他变量因素间的 MIC

图 2-5 中，有功功率、太阳辐照度和电压之间的相关程度最高，MIC 分别为 0.72 和 0.65。因此，本文为了降低建模复杂度，提高模型泛化能力，只考虑将有功功率、太阳辐照度对电压态势预测的影响，将其作为电压态势预测模型的输入变量。光伏电站 2020 年 1 至 5 月电压、太阳辐照度、有功功率变化曲线如图 2-6 所示。





(c) 1至5月有功功率变化曲线

图 2-6 1至5月模型输入变量变化曲线

(2) 频率与变量因素相关性理解

频率与外部变量因素间的 MIC 如热力图 2-7 所示。

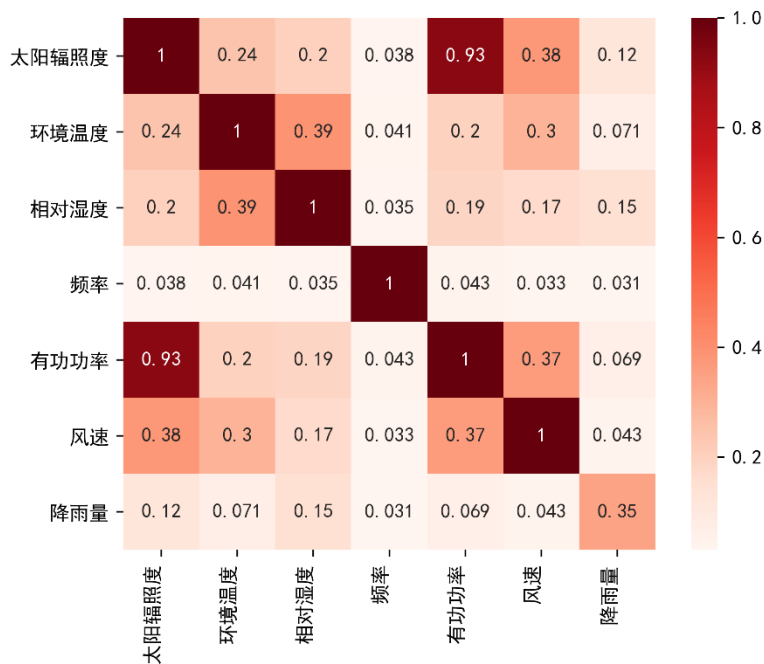


图 2-7 频率与其他变量因素间的 MIC

从图 2-7 可知，频率与太阳辐照度等 6 种外部变量因素间的 MIC 均在 0.1 以下，相关程度弱，因此频率态势预测模型的输入变量仅考虑频率数据本身。光伏电站 2020 年 1 至 5 月频率变化曲线如图 2-8 所示。

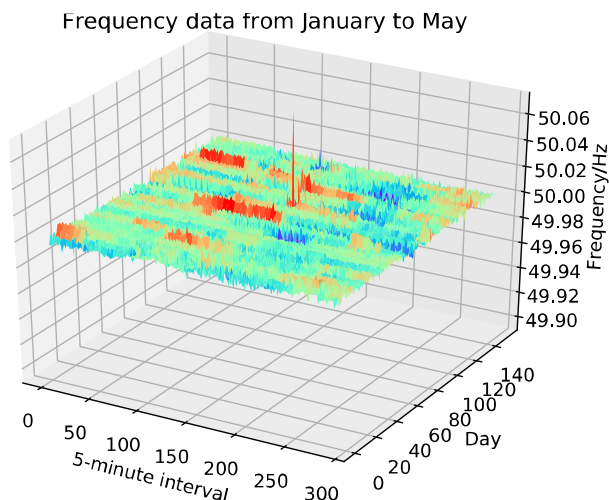


图 2-8 1 至 5 月频率变化曲线

(3) 电压总谐波畸变率与变量因素相关性理解

电压总谐波畸变率 (THDu) 与外部变量因素间的 MIC 如热力图 2-9 所示。

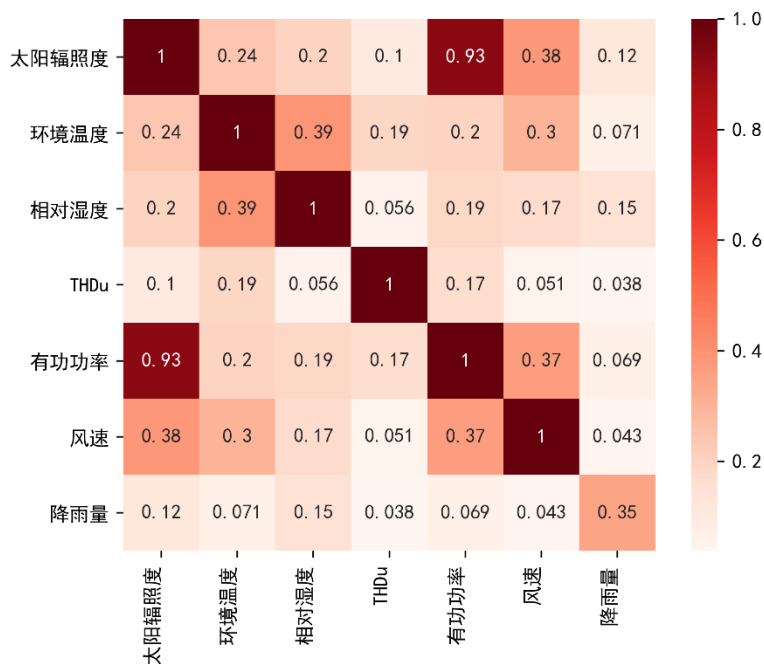


图 2-9 THDu 与其他变量因素间的 MIC

图 2-9 中，电压总谐波畸变率与环境温度间的最大 MIC 值仅为 0.19，外部变量因素与电压总谐波畸变率的关联程度较弱，因此电压总谐波畸变率态势预测模型仅考虑将电压总谐波畸变率历史数据作为输入。光伏电站 2020 年 1 至 5 月电压总谐波畸变率变化曲线如图 2-10 所示。

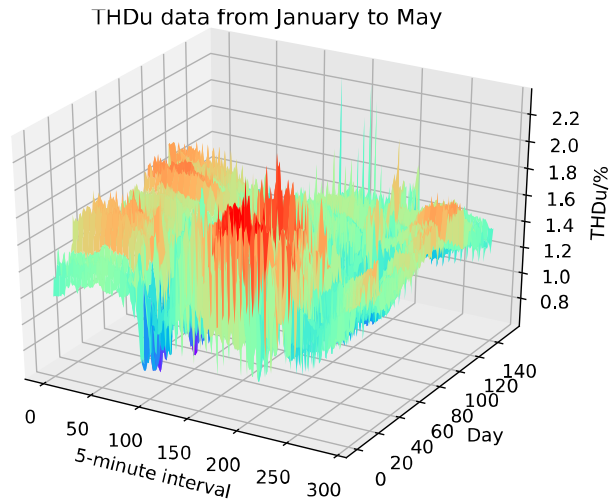


图 2-10 1 至 5 月电压总谐波畸变率变化曲线

2.5 本章小结

本章首先提出光伏电站电能质量态势感知架构,指出光伏电站电能质量态势感知由光伏电站电能质量要素提取、光伏电站电能质量态势理解和光伏电站电能质量态势预测三个阶段组成,然后阐述光伏电站电能质量要素提取具体实现过程,使用自相关系数和 MIC 分析方法对某地光伏电站电压、频率、电压总谐波畸变率三种电能质量实测数据进行态势理解,最后对光伏电站电能质量数据和外部因素数据进行异常值和标准化处理,并针对三种不同电能质量序列分别采用两种不同的模型评估指标全面评估模型性能,为后续章节构建有效的电能质量态势预测模型进行了充分的研究分析。

第 3 章 基于 DBN 及 SVR 模型的光伏电站电能质量态势预测

近年来,随着深度学习的快速发展,许多时序预测场景已使用深度学习模型取得了很好的效果,例如负荷预测等。而电能质量的态势预测作为时间序列预测任务的一种,自提出以来已有许多基于机器学习算法的预测模型被用来解决这个问题,其中 DBN 可以通过设置多个隐含层来从输入数据中提取隐藏特征,SVR 可以解决时序预测中的非线性特征拟合问题,因此本章首先考虑采用传统 DBN 和 SVR 单一模型对电能质量时序数据进行预测。

3.1 DBN 模型构建及实现

3.1.1 深度置信网络 (DBN)

DBN 由多层受限玻尔兹曼机 (Restricted Boltzmann Machine, RBM) 的堆栈和顶部的回归层组成,通过正向学习结合梯度下降的反向微调机制,可以实现更低的模型训练误差。RBM 是一种可以用随机神经网络来解释的概率突变模型,在经典的 RBM 结构中,处在同一层的隐含单元彼此没有相关关系,该结构是在玻尔兹曼机 (BM) 的基础上发展而来的,解决了传统 BM 训练速度慢的缺点,提高了网络的训练速度。RBM 由两层隐含单元组成,不同层隐含单元之间存在无向全连接,同一层隐含单元之间不存在连接,数据由可视层输入,经隐含单元和权值矩阵训练后由隐含层输出,如图 3-1 所示。

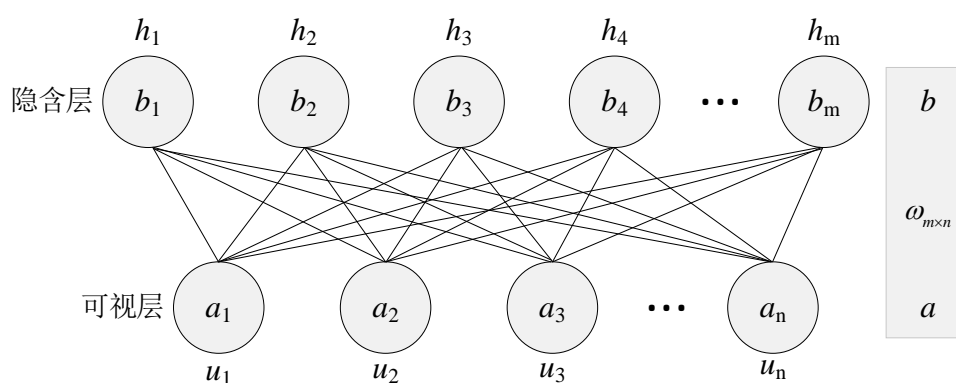


图 3-1 RBM 结构

图 3-1 中 $h_1 \sim h_m$ 为隐含层单元节点, $u_1 \sim u_n$ 为可视层单元节点, $a_1 \sim a_n$ 为可视层节点偏移, $b_1 \sim b_m$ 为隐含层节点偏移, $\omega_{m \times n}$ 为隐含层和可视层之间的权值矩阵。

由多个堆叠的 RBM 构成一个典型的 DBN 网络模型,如图 3-2 所示。与浅层神经网络相比,这种堆叠的 DBN 结构具有更深的网络层次和更好的模型泛化

能力。传统的神经网络依赖于数据特征的选择，而 DBN 可以通过设置多个隐含层来从输入数据中提取隐藏特征。DBN 由级联 RBM 和顶层采用的反向传播算法组成，算法训练过程分为预训练和参数微调两部分。预训练是指输入数据在底层 RBM 无监督的情况下逐层训练，将前一层的输出作为上层 RBM 的输入数据，这种结构可以有效地筛选出特征信息。参数微调过程包括整体调优和监督训练，通过输出数据和真实数据之间的误差向下逐层传播，以微调整个网络的参数。

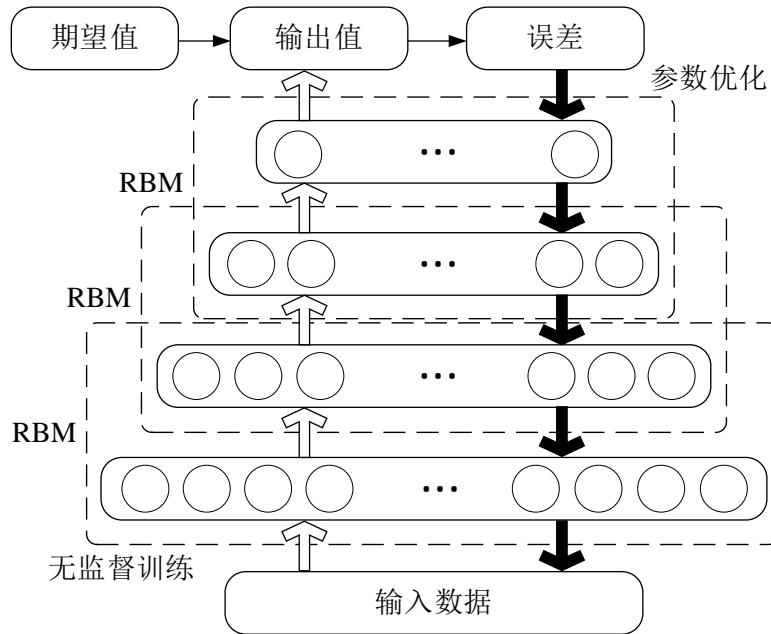


图 3-2 DBN 结构

3.1.2 模型构建及训练

DBN 模型构建过程中包括 RBM 层数及每层所含隐含单元数的确定。以 5 月 11 日至 6 月 10 日光伏电站电压数据集为训练目标寻找最优的隐含层数及隐含单元数，逐层枚举隐含单元数确定最优模型结构。实验中，首先在第一层分析不同隐含单元数实验得到的 MAE 和 RMSE 并确定最优数目；然后固定第一层最优数目再添加层数，寻找第二层最优数目；以此类推，直至模型的 MAE 和 RMSE 不再下降为止。实验中预设模型迭代训练 40 轮，每轮批大小为 50。不同 RBM 层数及隐含单元数对模型预测性能的影响如表 3-1 所示。

由表 3-1 中 MAE 和 RMSE 结果可知最优的 DBN 模型结构参数为：三层隐含 RBM，单元数分别为 32，128，128。

接着在 DBN 模型训练过程中找到迭代次数 (epochs) 以及批大小 (batch_size) 的最优组合，迭代次数即将模型在训练集上完整地遍历多少轮，设置合适的迭代次数能够保证拥有较高预测准确度的同时拥有较快的收敛速度。批大小即模型每轮迭代训练处理的数据量，批大小决定着在每轮迭代中训练集将被分成多少批进

行处理，合适的批大小在提升模型效率方面有着重要作用。对于 DBN 模型训练迭代次数和批大小的最优选取采用网格搜索法（Grid SearchCV, GS）^[58]进行寻优，寻优区间如表 3-2 所示。迭代次数以 20 为步长从 20 轮搜索至 120 轮，批大小以 50 为步长从 50 搜索至 300，参数类型均为整型。

表 3-1 DBN 模型结构不同时对预测性能的影响

隐含层数	隐含单元数	MAE/V	RMSE/V
1	16	0.458	0.706
	32	0.388	0.594
	64	0.445	0.570
	128	0.438	0.637
2	16	0.394	0.603
	32	0.358	0.560
	64	0.401	0.481
	128	0.349	0.443
3	16	0.525	0.849
	32	0.322	0.535
	64	0.299	0.505
	128	0.262	0.417
4	16	0.404	0.676
	32	0.319	0.551
	64	0.338	0.466
	128	0.371	0.467

表 3-2 DBN 寻优区间

超参数	寻优区间	类型
迭代次数	[20: 20: 120]	整型
批大小	[50: 50: 300]	整型

GS 创建一个超参数网格并遍历网格区间中 DBN 模型的所有迭代次数和批大小的组合得到最优迭代次数和批大小分别为 60 和 100，然后使用控制变量并画箱线图的方法来验证该参数组合在 DBN 预测模型训练中的优越性。

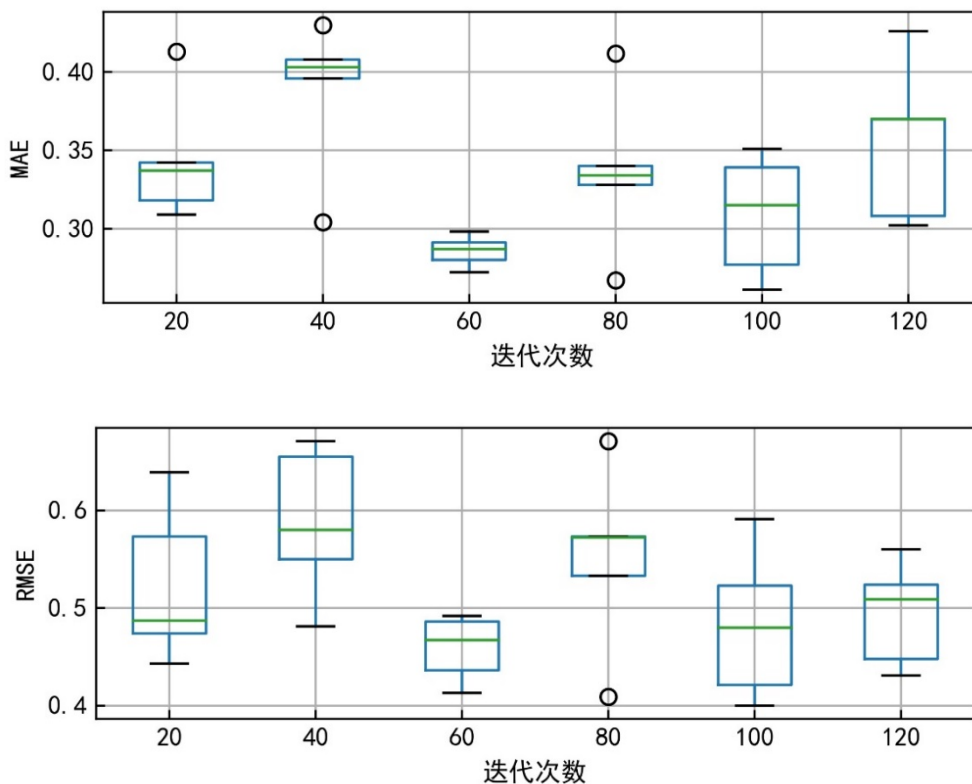
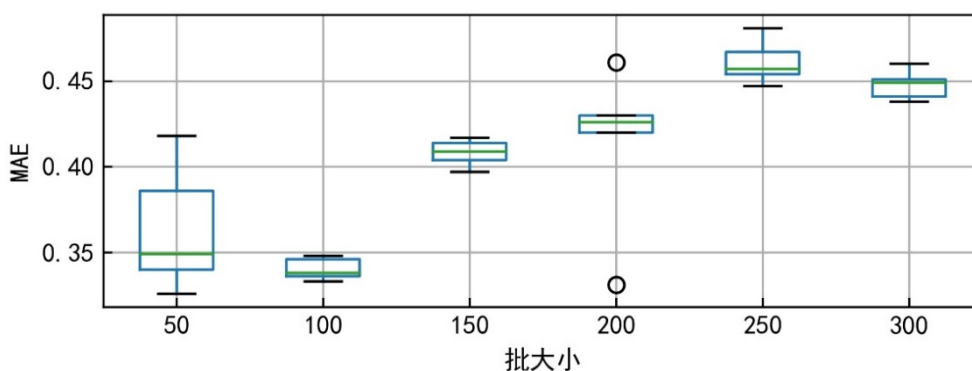


图 3-3 不同迭代次数下 DBN 模型训练情况

图 3-3 为 DBN 模型在不同迭代次数下的 MAE 和 RMSE 情况对比，任一箱体表示模型在 5 月 11 日至 6 月 10 日电压数据集上五次预测的 MAE 和 RMSE，横坐标为迭代区间中的不同取值。从图中可以看出在 60 轮迭代时，DBN 模型有较稳定的 MAE 和 RMSE。



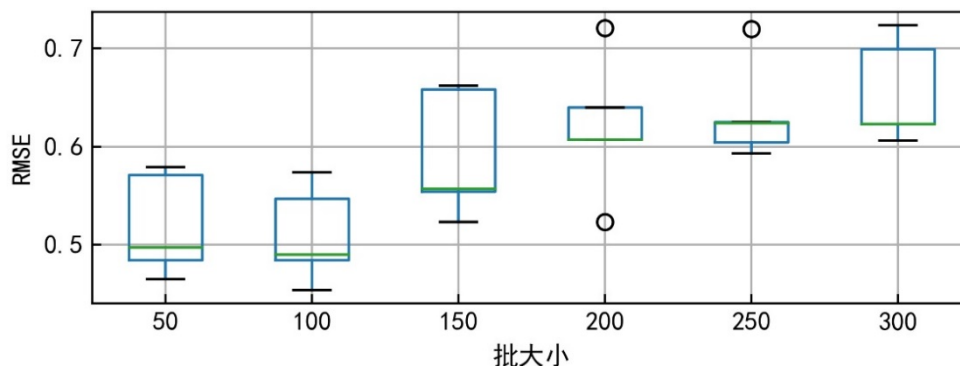


图 3-4 不同批大小下 DBN 模型训练情况

图 3-4 为 DBN 模型在不同批大小的 MAE 和 RMSE 情况对比，任一箱体表示模型在 5 月 11 日至 6 月 10 日电压数据集上五次预测的 MAE 和 RMSE，横坐标为批大小区间中的不同取值。由图可知在批大小等于 100 时，DBN 模型能取得最低的 MAE 和 RMSE。

3.1.3 实现流程

在介绍了 DBN 模型原理和构建、训练过程后，是实现 DBN 模型完成光伏电站电能质量态势预测。DBN 模型光伏电站电能质量态势预测实现步骤为：

步骤一，根据态势预测时间尺度以时间间隔 Δt 提取光伏电站采集数据中 Δt 时间段内的平均值作为参考值构建时间序列；

步骤二，对光伏电站电能质量时间序列进行自相关性理解得出模型输入时间步长，对光伏电站电能质量时间序列进行外部变量因素相关性理解得出输入数据特征维度；

步骤三，利用数据密度和聚类分析方法检出时序数据中的异常值并采用局部加权回归插补法对异常值进行处理；

步骤四，对处理完异常值的时序数据进行 min-max 标准化；

步骤五，构建含三层 RBM，隐含单元数分别为 (32, 128, 128) 的 DBN 光伏电站电能质量态势预测模型；

步骤六，将标准化后的时序数据结合输入时间步长和输入数据维度得到输入训练集和测试集，将训练集输入所构建 DBN 模型迭代训练 60 轮并在测试集上进行电能质量态势预测，综合评估模型性能。

3.1.4 算例分析

为评估模型的性能，基于第二章电能质量态势理解后的数据进行算例分析，包括并网点电压、频率、电压总谐波畸变率三种电能质量时序数据，时序间隔为

5 分钟。选取 2020 年 1 月 1 日至 2020 年 5 月 10 日的共 37728 条数据用于模型训练,并选取两日具有不同变化情况的电能质量时序数据作为测试集,记为 Day1 和 Day2。Day1 和 Day2 的电压、频率、电压总谐波畸变率三种电能质量时序数据变化情况如图 3-5 至图 3-7 所示,其中 Day1 的值选取于 5 月 23 日,Day2 的值选取于 5 月 30 日。对于电压,5 月 23 日的波动较为平缓,而 5 月 30 日的波动幅度更大,波动次数更为频繁;对于频率,5 月 23 日和 5 月 30 波动都非常频繁,但变化趋势有所不同;对于电压总谐波畸变率,5 月 23 日波动幅度更大,5 月 30 日波动次数更频繁。

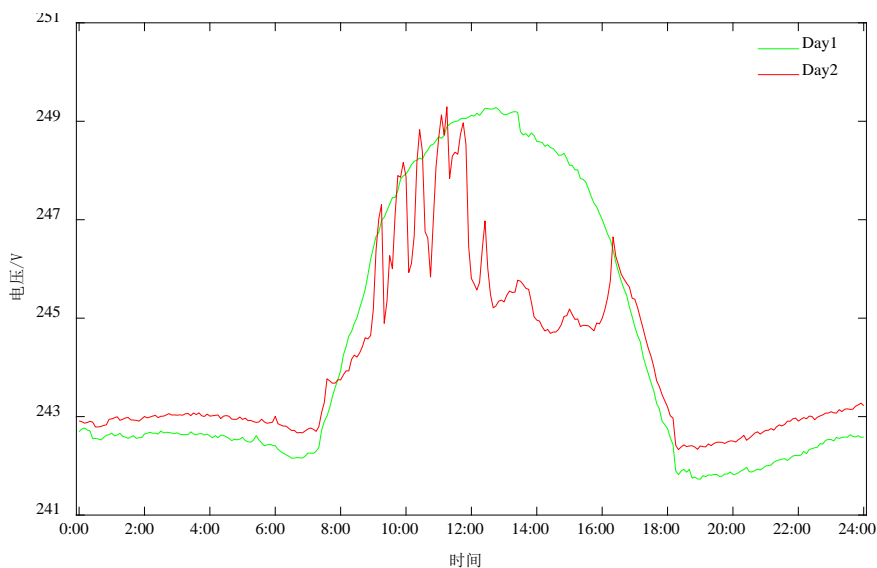


图 3-5 电压测试集

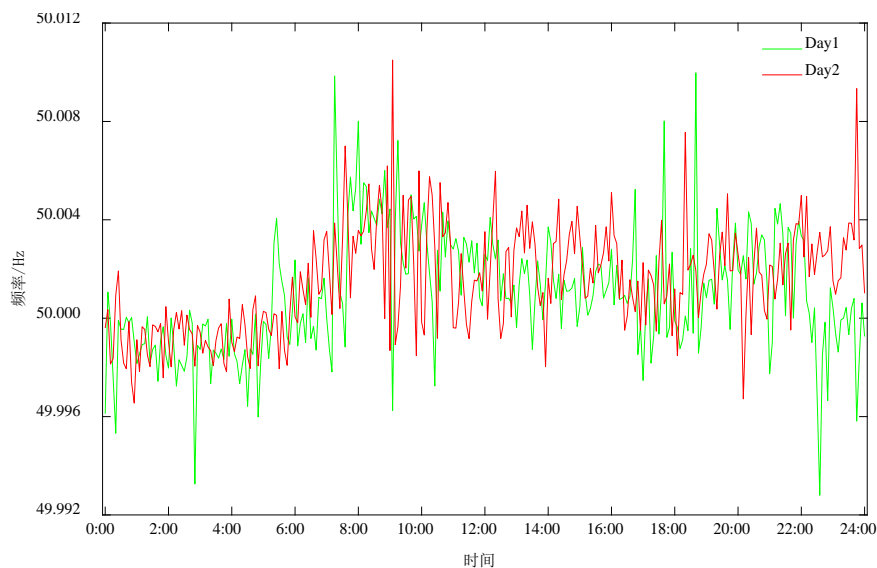


图 3-6 频率测试集

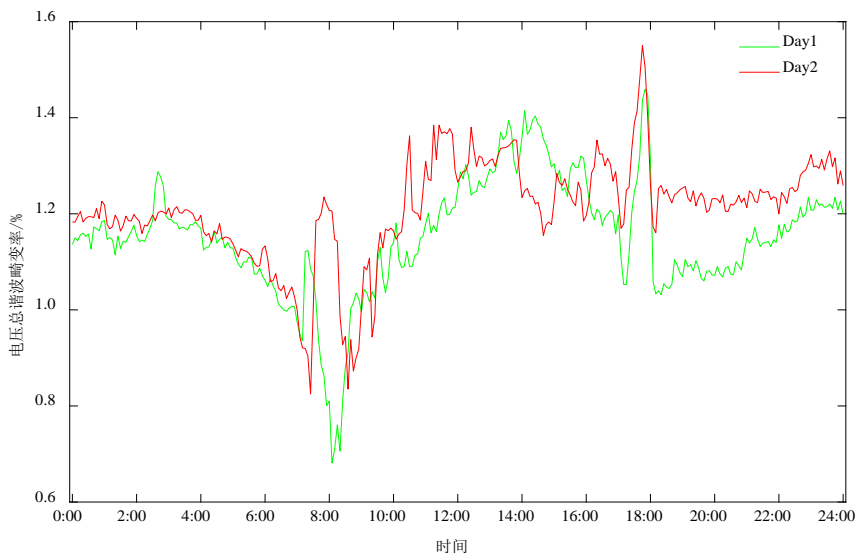


图 3-7 电压总谐波畸变率测试集

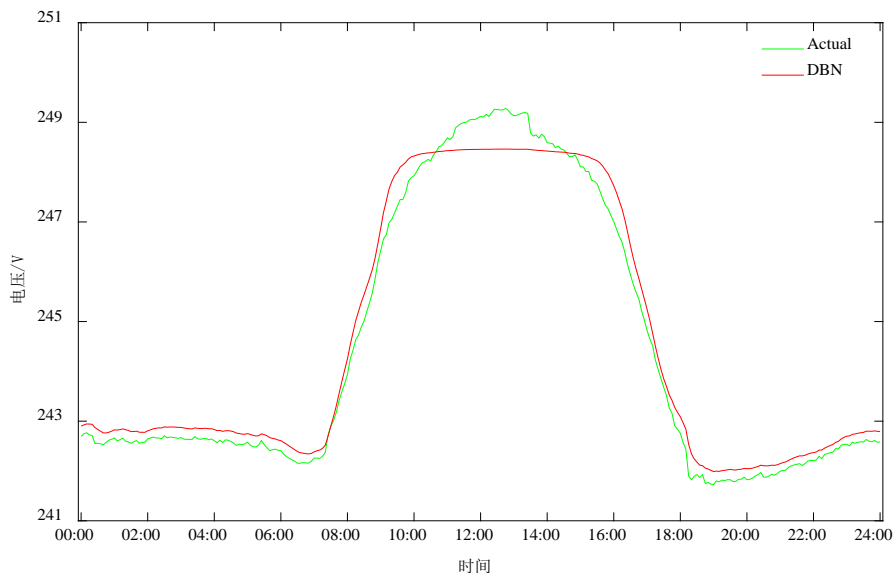
(1) 电压预测

通过选用两日不同变化模式的电压时序数据作为测试集，能够评估模型在预测电压波动剧烈、频繁和电压波动平缓情况下的表现。DBN 模型在 Day1、Day2 电压测试集上三个时间步预测的 MAE 和 RMSE 如表 3-3 所示。

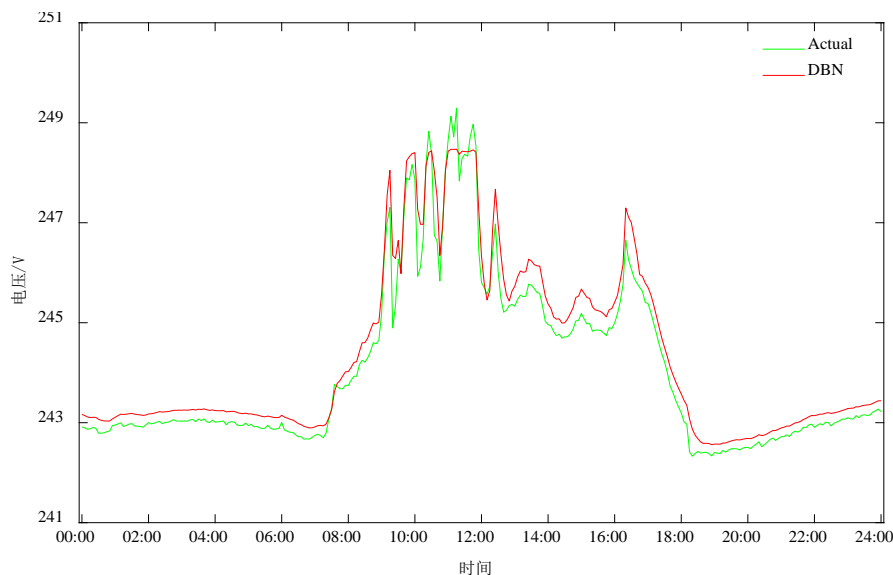
表 3-3 电压预测结果

DBN	MAE-1/V	MAE-2/V	MAE-3/V	RMSE-1/V	RMSE-2/V	RMSE-3/V
Day1	0.305	0.337	0.364	0.374	0.422	0.463
Day2	0.364	0.419	0.483	0.515	0.629	0.724

图 3-8 为 DBN 模型在 5 月 23 日和 5 月 30 日提前 5 分钟预测电压拟合情况。



(a) 5 月 23 日 DBN 电压预测曲线情况



(b) 5月30日 DBN 电压预测曲线情况

图 3-8 Day1、Day2DBN 电压预测曲线情况

由表 3-3 和图 3-8 可知，DBN 模型在 Day2 电压测试集的 MAE 和 RMSE 均比 Day1 电压测试集要高，说明模型对于幅度更大、次数更为频繁的电压波动预测效果变差，且不论是在电压波动较为平缓的 Day1 测试集还是波动更为频繁、幅度更大的 Day2 测试集，DBN 模型在提前 5 分钟预测两日测试集中 09:00-17:00 电压峰值时段时，电压预测值偏离程度都明显增大，模型稳定性较差。同时 DBN 模型在提前 15 分钟预测 Day1 和 Day2 电压相比提前 5 分钟预测时 MAE 和 RMSE 增长都较大，说明该模型在电压频繁、剧烈波动的多时间步预测输出时性能较差，在 Day1 和 Day2 电压峰值时段内的的预测值参考价值不大。

(2) 频率预测

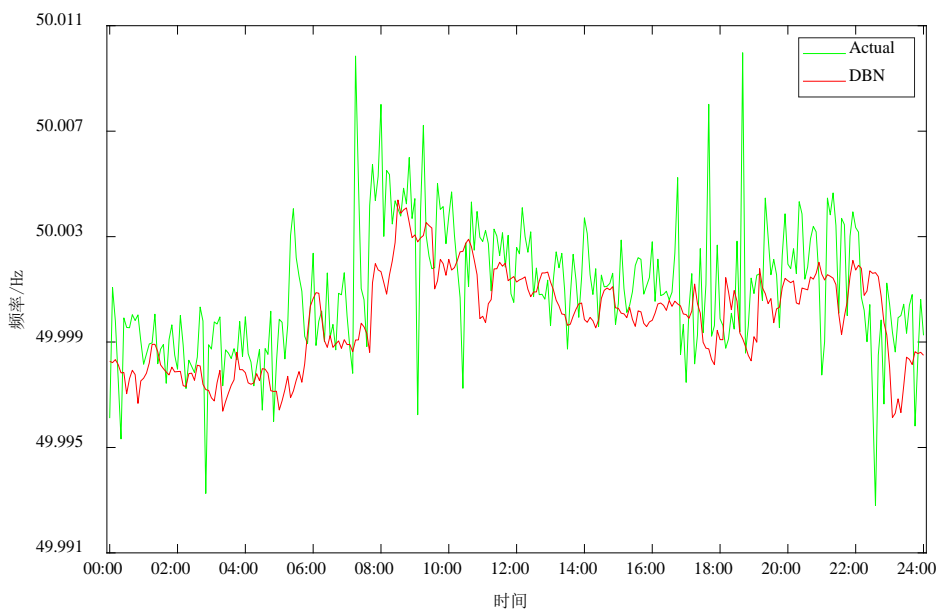
对于两日测试集 Day1 和 Day2 的频率，波动都非常频繁，DBN 模型在 Day1、Day2 频率测试集上三个时间步预测的 DAAPE 和 NRMSE 如表 3-4 所示。

表 3-4 频率预测结果

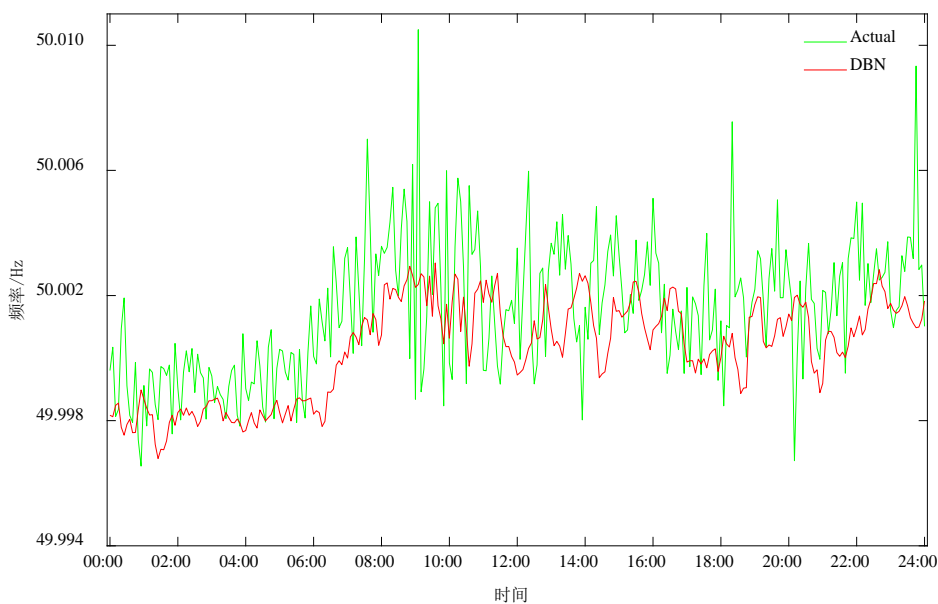
DBN	DAAPE-1/%	DAAPE-2/%	DAAPE-3/%	NRMSE-1/Hz	NRMSE-2/Hz	NRMSE-3/Hz
Day1	1.066	1.095	1.102	0.178	0.182	0.183
Day2	1.057	1.120	1.133	0.210	0.220	0.222

图 3-9 为 DBN 模型在 5 月 23 日和 5 月 30 日提前 5 分钟预测的频率拟合情况。

由表 3-4 可知，DBN 模型在 Day2 频率测试集上的 DAAPE 和 NRMSE 略大于 Day1 频率测试集，因为两日频率变化走向虽有所不同，但都非常剧烈。从图 3-9 的频率预测曲线拟合情况可以看出，不论是在 Day1 测试集还是在 Day2 测试



(a) 5月23日 DBN 频率预测曲线情况



(b) 5月30日 DBN 频率预测曲线情况

图 3-9 Day1、Day2DBN 频率预测曲线情况

集上，频率预测曲线与实际频率曲线偏离较大，模型在 Day1 测试集上几乎没有学习到频率数据的变化趋势，在 Day2 测试集上拟合频率数据变化趋势较差。同样，表 3-4 中提前 15 分钟预测 Day1 和 Day2 测试集的频率时，相比提前 5 分钟预测 Day1 和 Day2 测试集频率的 DAAPE 和 NRMSE 都有所上升，图 3-9 中提前 5 分钟频率拟合效果已经比较差，可以知道在输出时间步更长时已然不能较好地拟合实际频率曲线，DBN 模型不适用于该频率频繁波动场景下的频率预测。

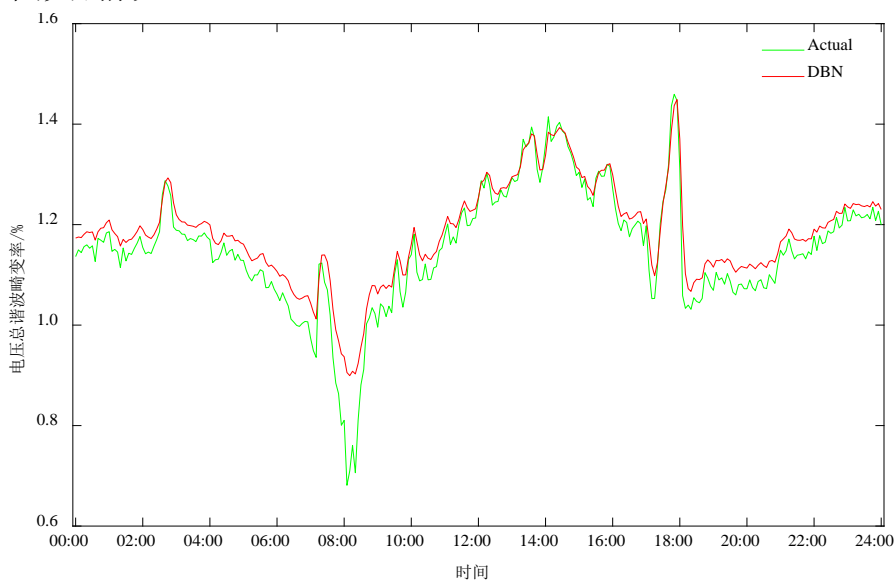
(3) 并网点电压总谐波畸变率预测

DBN 模型在 Day1、Day2 电压总谐波畸变率测试集上三个时间步预测的 MAPE 和 NRMSE 如表 3-5 所示。

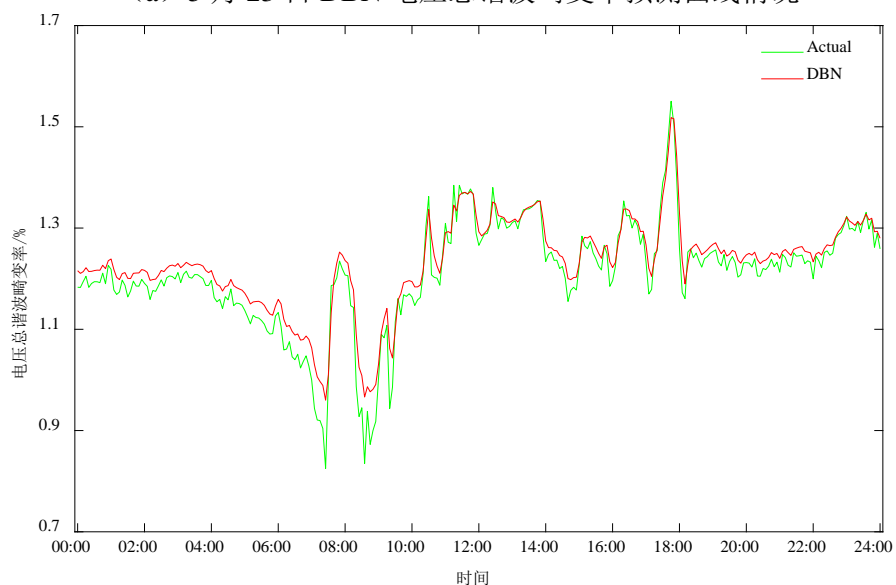
表 3-5 电压总谐波畸变率预测结果

DBN	MAPE-1/%	MAPE-2/%	MAPE-3/%	NRMSE-1/%	NRMSE-2/%	NRMSE-3/%
Day1	3.787	4.499	5.762	0.071	0.088	0.109
Day2	3.111	3.796	4.711	0.069	0.087	0.105

图 3-10 为 DBN 模型在 5 月 23 日和 5 月 30 日提前 5 分钟预测的电压总谐波畸变率拟合情况。



(a) 5 月 23 日 DBN 电压总谐波畸变率预测曲线情况



(b) 5 月 30 日 DBN 电压总谐波畸变率预测曲线情况

图 3-10 Day1、Day2DBN 电压总谐波畸变率预测曲线情况

由表 3-5 可知, DBN 模型在 Day1 电压总谐波畸变率测试集上的 MAPE 大于 Day2 电压总谐波畸变率测试集, 但 NRMSE 在两个测试集上差别不大, 说明模型在 Day1 测试集上的预测准确度比 Day2 测试集低, 但模型在两个测试集上的稳定性表现差别很小。从图 3-10 可以看出, 模型在实际电压总谐波畸变率曲线快速下降时并不能很好地拟合其变化趋势, 在电压总谐波畸变率实际曲线波谷处表现不佳, 不能为谐波治理提供精确的信息。同样, 表 3-5 中提前 15 分钟预测电压总谐波畸变率时的 MAPE 和 NRMSE 比提前 5 分钟预测时要增加不少, 可以知道在提前 15 分钟预测电压总谐波畸变率时的情况较差, 综合来看, DBN 模型在电压总谐波畸变率测试集上的表现不佳。

3.2 SVR 模型构建及实现

3.2.1 支持向量回归 (SVR)

给定光伏电站电能质量训练集 $\{x_i, y_i\}, i=1, 2, \dots, n$, $x_i \in R^n$ 为输入的电能质量及相关量, $y_i \in R$ 为相应的输出电能质量。SVR 的基本思想是找到一个函数将输入量从低维空间映射到高维特征空间中, 此函数可定义为

$$f(x) = (\omega \cdot \varphi(x)) + b \quad (3-1)$$

式 (3-1) 中 x 为输入量, $\varphi(\cdot)$ 表示 $R^n \rightarrow R^m (m > n)$ 的非线性变换, ω 为权值矩阵, b 为偏置项。通过最小化回归风险 $f(x)$ 可描述为求解优化问题:

$$\min \frac{1}{2} \|\omega\|^2 + A \sum_i^n \ell(f(x_i) - y_i) \quad (3-2)$$

式 (3-2) 中 ℓ 为代价函数, A 为常数, 是训练误差最小化和模型复杂度最低化之间的权衡因子。通过引入松弛变量 ξ_i , ξ_i^* 以适应输入训练集中不规则的因子可将优化问题更新为

$$\min \frac{1}{2} \|\omega\|^2 + A \sum_i^n (\xi_i + \xi_i^*) \quad (3-3)$$

$$\text{s.t.} \begin{cases} y_i - \omega \varphi(x) - b \leq \varepsilon + \xi_i \\ \omega \varphi(x) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (3-4)$$

式 (3-4) 中, ε 为允许误差。求解带有不等式约束的二次优化问题后, 得到式 (3-1) 中的权值矩阵 ω 如式 (3-5) 所示。

$$\omega = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \varphi(x_i) \quad (3-5)$$

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/577126062052006031>