

分类号: TM615

学校代码: 10363

密 级: 公开

学 号: 2210310102



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题目 基于灰色关联分析与 FCM
的光伏阵列健康状态评估

论 文 作 者 王传志

指 导 教 师 汪石农

学 科 (专 业) 电气工程

研 究 方 向 电力系统及其自动化

论文提交日期: 2024 年 5 月 31 日

分类号: TM615

密 级: 公开

单位代码: 10363

学 号: 2210310102

题 目 基于灰色关联分析与 FCM 的光伏阵列健康状态评估

英文并列

题 目 **HEALTH STATUS ASSESSMENT OF PV ARRAY BASED
ON GREY RELATIONAL ANALYSIS AND FCM**

学生姓名: 王传志

指导教师: 汪石农

专 业: 电气工程

研究方向: 电力系统及其自动化

论文答辩日期: 2024 年 5 月 24 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律结果由本人承担。

学位论文作者签名：王传志

日期：2024 年 5 月 24 日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐，在 ____ 年解密后适用本版权书。

本学位论文属于

不保密 ☒。

学位论文作者签名：王传志

指导教师签名：王传志

日期：2024 年 5 月 24 日

日期：2024 年 5 月 24 日

基于灰色关联分析与 FCM 的光伏阵列健康状态评估

摘 要

随着“双碳”政策的提出以及新型电力系统的推进，新能源发电技术的应用越来越广泛。中国的太阳能资源丰富，因此在众多新能源发电技术中光伏发电占比越来越大；而且光伏发电具有对环境污染小、建设周期短、边际发电成本低等优点。但是，光伏发电系统一般建设于户外，因此，光伏阵列的性能受外界环境条件的影响很大，如日照强度和组件工作温度；此外，光伏阵列性能还受系统部件异常状况等因素的影响，如 EVA 老化、钢化玻璃和电池之间的粘合剂失效等。因此，光伏阵列健康状态的评估对光伏发电系统的安全运行和效率提升具有重要的意义。

首先，分析光伏发电系统的工作原理，以及光伏发电系统的各种类型，再利用 MATLAB/SIMULINK 建立 3×3 的光伏阵列仿真模型。通过对正常和故障状态下的光伏阵列进行仿真模拟，得到光伏阵列在不同状态下的仿真 I-V 特性曲线和 P-V 特性曲线以及最大功率点电压和功率的数据。再对仿真得到的不同状态下的 I-V 特性曲线和正常运行状态下的 I-V 特性曲线进行灰色关联分析(Grey relational analysis, GRA)，得到两者之间的灰色关联度(Grey relation degree, GRD)；利用三角隶属度函数将得到的灰色关联度与健康等级进行匹配，并由模糊隶属度计算出光伏阵列的健康指数(Health index, HI)，从而划分出光伏阵列的健康、亚健康、异常和故障四种健康等级。

其次，针对亚健康或异常等级下的光伏阵列，采用模糊 C 均值聚类算法(Fuzzy C-mean clustering, FCM)对相应的光伏阵列特征参数进行聚类分析。为了消除不同状态光伏阵列的参数变化对输出特征参数的影响，使模糊 C 均值聚类算法能正确诊断出亚健康或异常等级下光伏阵列的故障类型，对特征参数进行归一化处理，进而对处理后的数据进行聚类得到聚类中心；再利用聚类中心与测试数据代入高斯隶属度函数，诊断出光伏阵列的具体工作状态类型。

然后，利用概率神经网络(Probabilistic neural network, PNN)对不同工作状态下的光伏阵列最大功率点特征数据进行训练，建立基于概率神经网络的光伏阵列故障诊断模型，并与灰色关联分析与 FCM 结合所得的光伏阵列诊断结果进行对比验证。

最后，将灰色关联分析与 FCM 结合所得出的结果作为光伏阵列健康状态评估的依据。灰色关联分析仅需要对光伏阵列的 I-V 特性曲线进行分析，算法的计算量小。而 FCM 算法虽然相对复杂，但仅针对亚健康或异常等级下的光伏阵列进行故障诊断，大大减少了聚类所需的数据量，同时能精确地诊断出光伏阵列的故障类型。因此，灰色关联分析与 FCM 算法融合后具有较高的适用性，在减少了光伏阵列健康状态评估的工作量的同时，实现了光伏阵列健康状态的准确评估。

关键词：光伏阵列，I-V 特性曲线，灰色关联分析，模糊 C 均值聚类，概率神经网络，健康状态评估

HEALTH STATUS ASSESSMENT OF PV ARRAY BASED ON GREY RELATIONAL ANALYSIS AND FCM

ABSTRACT

With the introduction of the "dual-carbon" policy and the promotion of new power systems, the application of new energy generation technologies is becoming more and more widespread. China is rich in solar energy resources, so photovoltaic (PV) power generation accounts for an increasing proportion of the many new energy generation technologies; moreover, PV power generation has the advantages of low environmental pollution, short construction period, and low marginal power generation cost. However, PV power generation systems are generally constructed outdoors, so the performance of PV arrays is greatly affected by external environmental conditions, such as the solar irradiance and the operating temperature of the modules; in addition, the performance of PV arrays is also affected by factors such as the abnormal condition of system components, such as the aging of EVA, and the failure of the adhesive between the toughened glass and the batteries. Therefore, the health status assessment of PV array is of great significance for the safe operation and efficiency improvement of PV power generation systems.

Firstly, the working principle of PV power generation system and various types of PV power generation system are analysed, and then a 3×3 PV array simulation model is established using MATLAB/SIMULINK. By simulating the PV array under normal and faulty states, the simulated I-V characteristic curve and P-V characteristic curve of the PV array under different states as well as the data of voltage and power at the maximum power point are obtained. Then the grey relational analysis is carried out on the simulated I-V characteristic curves under different states and the I-V characteristic curves under normal operation state to obtain the grey relation degree between the two; the grey relation degree obtained is matched with the health level using the triangular affiliation function, and the health index of the PV array is calculated from the fuzzy affiliation degree, which can be divided into four kinds of PV arrays, namely, healthy, sub-healthy, abnormal, and faulty.

Secondly, for the PV arrays under the sub-healthy or abnormal class, the fuzzy C-mean clustering algorithm is used to cluster and analyse the corresponding characteristic parameters of the PV arrays. In order to eliminate the influence of the parameter changes of PV arrays in different states on the output characteristic parameters, so that the fuzzy C-mean clustering algorithm can correctly diagnose the type of faults of PV arrays under the sub-healthy or abnormal class, the characteristic parameters are normalised, and then the processed data are clustered to obtain the

clustering center; and then the clustering center is used to diagnose the specific PV arrays using the test data substituting into the Gaussian subordinate degree function.

Then, the probabilistic neural network is used to train the maximum power point feature data of PV arrays in different operating states, to establish a PV array fault diagnosis model, and to compare and verify the PV array diagnosis results obtained from the combination of grey relational analysis and FCM.

Finally, the results from the combination of grey relational analysis and FCM are used as the basis for PV array health assessment. The grey relational analysis only needs to analyse the I-V characteristic curve of the PV array, and the calculation volume of the algorithm is small. Although the FCM algorithm is relatively complex, it only diagnose the faults of PV arrays with sub-healthy or abnormal levels, which greatly reduces the amount of data required for clustering, and at the same time, it can accurately diagnose the fault types of PV arrays. Therefore, the fusion of grey relational analysis and FCM algorithm has high applicability and achieves accurate assessment of PV array health status while reducing the workload of PV array health status assessment.

Wang Chuan Zhi(Electrical Engineering)

Supervised by Wang Shi Nong

KEY WORDS: photovoltaic array, I-V characteristic curve, grey relational analysis, fuzzy C-mean clustering, probabilistic neural network, health state assessment

目录

第 1 章 绪论	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	3
1.3 论文框架.....	7
1.4 本文结构组织.....	9
第 2 章 光伏发电系统分析与光伏阵列建模	10
2.1 光伏发电理论.....	10
2.2 光伏发电系统类型.....	11
2.2.1 离网光伏发电系统.....	11
2.2.2 并网光伏发电系统.....	12
2.2.3 光伏发电系统中的设备.....	12
2.3 光伏阵列建模.....	13
2.3.1 光伏电池的数学模型.....	13
2.3.2 基于单二极管的光伏阵列数学模型.....	16
2.3.3 光伏阵列建模与输出特性分析.....	16
2.4 光伏阵列故障输出特性分析.....	20
2.5 本章小结.....	28
第 3 章 基于灰色关联分析的光伏阵列健康等级划分	29
3.1 灰色关联分析原理.....	29
3.2 光伏阵列健康状态分析.....	30
3.3 不同状态下光伏阵列健康等级划分.....	32
3.3.1 光伏阵列的局部阴影状态.....	32
3.3.2 光伏阵列的老化状态.....	34
3.3.3 光伏阵列的短路与开路状态.....	35
3.4 本章小结.....	37
第 4 章 基于 FCM 的光伏阵列故障诊断	38
4.1 模糊 C 均值聚类算法原理.....	38
4.2 选取不同样本特征量的对比与分析.....	41
4.3 高斯隶属度原理与故障诊断流程.....	44
4.4 FCM 聚类结果分析	47
4.4.1 FCM 聚类仿真分析	47
4.4.2 高斯隶属度诊断.....	48
4.5 本章小结.....	50
第 5 章 基于概率神经网络的光伏阵列故障诊断	51
5.1 概率神经网络原理.....	51
5.1.1 贝叶斯决策理论和概率密度函数.....	51
5.1.2 概率神经网络模型结构.....	52
5.2 基于概率神经网络的光伏阵列故障诊断模型.....	54

5.2.1 基于概率神经网络的光伏阵列故障诊断流程.....	54
5.2.2 基于概率神经网络的光伏阵列故障诊断.....	55
5.3 本章小结.....	57
第 6 章 总结与展望	59
6.1 研究内容总结.....	59
6.2 研究工作展望.....	60
参考文献	61
攻读学位期间发表的学术论文目录	66
致谢	67

第 1 章 绪论

1.1 研究背景与意义

随着“双碳”政策的提出，国家高度重视新能源技术发展，新型清洁能源也正在逐渐替代传统化石能源^[1]。传统化石能源虽然获取方式较为容易，利用技术也非常成熟，产出的能量稳定持久；但化石能源储备会随着时间的推移而耗尽，煤、石油和天然气等化石能源的开采和使用会导致空气污染和环境破坏^[2]。因此，世界各国均在发展新型清洁能源。

太阳能资源是众多新型清洁能源之一，它取之不尽用之不竭，属于可再生的清洁能源。光伏发电可根据需要在各种地方安装，不仅可用于住宅建筑物的屋顶，也可以布设在工业设备上。光伏发电在运行过程中无需消耗燃料，不会排放温室气体及其他废气，同时不会造成噪音污染。光伏发电系统无需机械传动部件，稳定性和可靠性较高。只要设备拥有太阳能电池组件，光伏发电就能稳定发电，且广泛应用自动控制技术，保证了发电系统的稳定性。因此，光伏发电技术被广泛应用和推广。

近几年中国光伏产业快速增长，为光伏电站的建设打下坚实基础。根据国家能源局数据，2013 至 2023 年间我国累计光伏装机容量与新增光伏装机容量变化趋势如图 1-1 所示；2013 至 2023 年间我国光伏发电量与光伏发电量所占总发电量的比重如图 1-2 所示。可以发现，随着对可再生能源的开发，2013 到 2023 这十年间我国的光伏装机容量与光伏发电量整体上都在稳定的增加。2013 年我国光伏发电装机总量只有 0.2 亿千瓦，到 2018 年，我国的光伏发电装机总容量增加至 1.74 亿千瓦，这时中国的太阳能发电装机容量已经在全球排名第一^[3]。在 2020 年，虽然受疫情以及经济的影响，暂缓了光伏电站的部署，但我国光伏新增装机总量仍有约为 0.48 亿千瓦，光伏发电累计装机容量达到 2.53 亿千瓦，其所发电量占总发电量的 2.7%^[4]。截至 2023 年，我国的光伏发电装机总容量增加至 5.6 亿千瓦，是 2013 年光伏发电装机总量的 28 倍；2023 年光伏发电量占总发电量的 5.4%，而 2013 年光伏发电量仅占总发电量的 0.23%，由此可知光伏发电的总容量与光伏发电量在总发电量上的占比都在大大增加。同时根据数据统计，在

2010 年到 2020 年的十年时间，大型光伏电站的发电成本下降了 85%，这表明光伏发电具有巨大的发展前景^[5-6]。

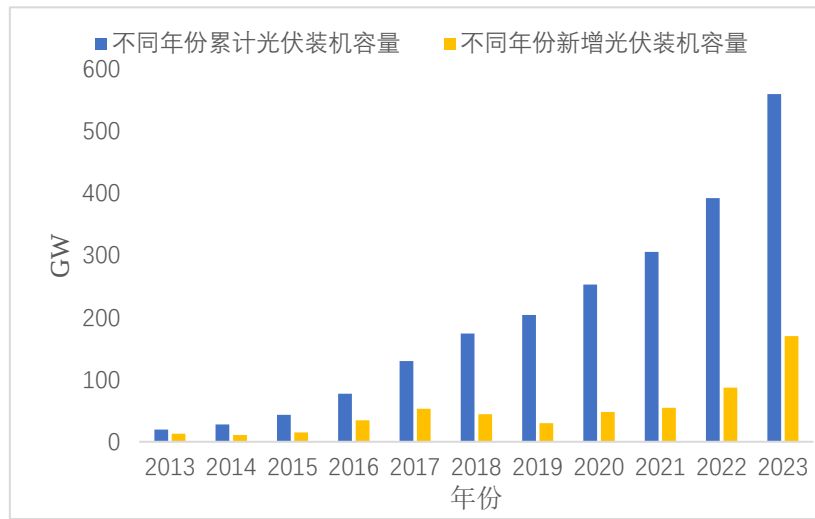


图 1-1 2013-2023 累计光伏装机容量与新增光伏装机容量

Fig1-1 2013-2023 cumulative PV installed capacity and new PV installed capacity

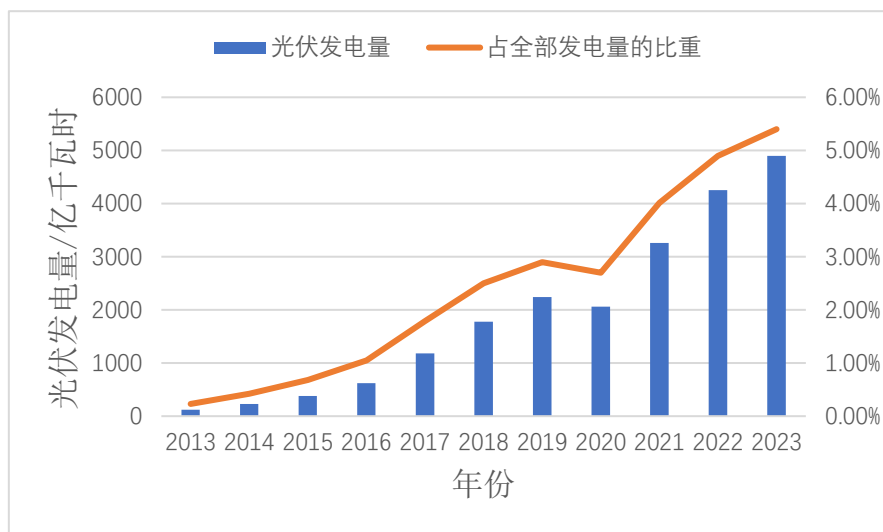


图 1-2 2013-2023 光伏发电量与占全部发电量的比重

Fig1-2 2013-2023 PV power generation and its proportion in total power generation

随着光伏发电量的增加与光伏发电在整个发电领域比重的增加，光伏发电已成为电力生产领域中至关重要的组成部分，同时也是促进“双碳”目标实现的有利因素^[7-8]。在正常情况下，光伏电站的寿命一般在 25 到 30 年之间，其年平均性能退化速率为 0.8%左右^[9]。为了使光伏阵列受到更多的光照，光伏电站大多建设于阳光充足的户外。由于光伏电站建设于户外，环境对光伏阵列运行的影响很大，如使用期间的风吹日晒以及风沙、雨雪等极端天气带来的影响，将导致光伏阵列

加快老化与受到损坏,严重影响设备的寿命;此外,光伏阵列使用期间还会遇到灰尘、异物等长时间遮挡部分组件的情况,这将影响光伏阵列的发电效率^[10]。因此,在众多因素的共同影响下,光伏阵列出现故障的频率可能会逐渐升高,对光伏阵列进行维护的费用也会增加,进而对光伏发电的经济收益造成负面影响。

为了光伏电站能够稳定持续的电力输出,就需要对光伏电站进行健康诊断和故障处理。但是,早年间大多数光伏电站的运维采用人工巡检的方式,人工巡检的局限性很大,而且耗费人力财力,增加发电成本^[11]。目前,光伏电站的监控默认基于逆变器的解决方案,可检测到电网错误、交流侧错误、光伏运行不稳定、传感器错误和接地故障等。然而,这种被动监控系统经常会触发错误警报,特别是在阴天,并且无法检测导致少量电力损失的故障或光伏组件相关的故障。随着技术发展,可以利用无人机携带红外摄像机对光伏阵列图像拍摄进行故障诊断,但单靠无人机巡检,诊断效率低下^[12-13]。还有利用光伏电站输出电气特性参数来诊断故障的方法,但是会因为输出电气特性数据特征不明显造成误检^[14]。而且,光伏电站发生故障后,如果无法快速识别光伏电站的故障类型,会影响光伏电站的发电效率^[15]。

因此,若不能及时诊断出光伏阵列的健康状态以及处理光伏阵列发生的故障,光伏阵列的发电效率和使用寿命都会受到影响,并且可能会对整个光伏电站造成不可逆转的影响;如果能及时且有效的诊断出光伏阵列的健康状态和处理光伏阵列发生的故障,则能够让光伏电站稳定安全的运行。

1.2 国内外研究现状

光伏发电系统的核心部分是由光伏阵列构成的,它们发挥着至关重要的作用。光伏阵列通常工作在户外环境中,受环境变化的影响较大,导致故障发生率较高,因此对光伏阵列的健康状况进行评估尤为重要。为了精确评估光伏阵列的健康状态并采取相应措施,国内外的科研人员已经进行了许多深入的研究。总结目前关于光伏阵列健康状态评估的技术与方法,主要分为以下三类:数学模型法、物理特性诊断法和智能诊断法。

(1)数学模型法

数学模型法是通过构建光伏阵列的数学模型,根据光伏阵列模型在不同状态

下的输出特性曲线以及该阵列的特征参数与光伏阵列实际的输出特性曲线和特征参数进行比较,从而诊断其健康状态。此方法可以有效地诊断光伏阵列的健康状态,所测数据较少,因此,诊断成本较低^[16],其工作原理如图 1-3 所示。

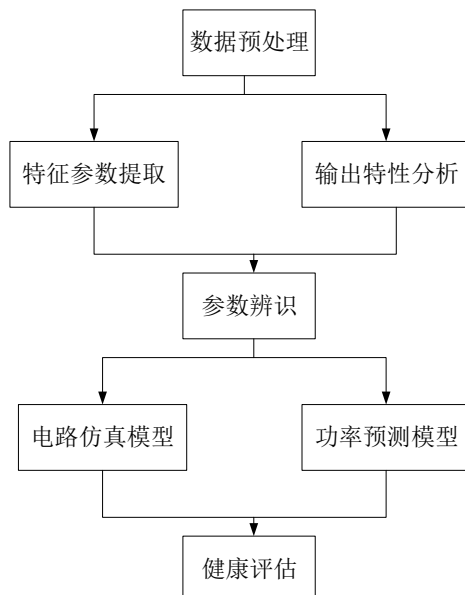


图 1-3 基于数学模型法的原理框图

Fig1-3 Block diagram based on mathematical modeling

文献[17]提出一种基于快速过采样主成分分析算法,通过提取不同工作状态下的 I-V 特性曲线,并通过分析故障时阵列工作点状态来识别故障类型。文献[18]利用光伏阵列的开路电压、短路电流和最大功率点电压和电流这四个参数,建立故障类型因子 K ,通过比较 K 与标准值之间的差异来诊断光伏阵列是否存在故障。文献[19]建立了一种简单实用的光伏阵列数学模型,比较分析了光伏阵列发生故障时的参数变化规律,同时研究了光伏阵列在正常情况和故障情况下的 I-V 特性曲线,从而诊断出光伏阵列的故障类型。文献[20]通过构建一个数学模型来模拟光伏阵列,研究在各种故障条件影响下光伏阵列的输出性能。通过对比这些性能特征,实现对光伏阵列故障的有效识别和诊断。文献[21]通过对最大功率点进行跟踪,对最大功率点的电流电压特性进行分析,再通过仿真估算内在参数并根据参数特征判断是否存在异常状态。文献[22]利用光伏阵列电压、电流、日照强度和温度测量的故障诊断与分类技术计算出光伏阵列的戴维南等效电阻,以准确识别故障。文献[23]利用光伏阵列的故障特征量来定义电压和电流指标,再根据各种故障的特征信息确定故障诊断阈值,从而实现实时电压和电流的指标与

相应的阈值进行比较，以诊断潜在的故障和故障类型。

尽管数学模型法计算简单且容易实现，但在求解过程为了计算方便，会将光伏阵列的一些特征参数取近似值，导致诊断出现错误的现象，在故障诊断中存在一定的局限性；而建立精确的光伏阵列数学模型是提高此类故障诊断方法的关键，该方法对故障组件诊断的精确度在于数学模型建立，而且诊断结果的正确性也依赖初始值的选择和关键原始数据的提取，所以使得求解可靠性差。

(2)物理特性诊断法

物理特性的健康状态评估方法主要是利用红外设备或电流、电压传感器来采集光伏阵列的温度、电流、电压等数据，进行诊断分析。通过光伏阵列的物理特性进行故障诊断和定位^[24]。

红外诊断法是通过红外热成像设备和计算机技术对故障组件进行诊断与定位。该方法可以将光伏阵列中的热辐射转换为热图像，组件温度的不同，热图像的颜色也不同^[25]。当光伏阵列的某些组件出现局部阴影或短路时，组件的温度会产生明显的变化，利用红外扫描仪可以诊断到组件温度的变化，通过热成像图中颜色分布，可以诊断出组件是否故障以及故障组件的位置^[26]。

文献[27]通过结合可见光与红外热图像的热斑诊断技术，我们能够避免出现误判的热斑虚警，提高诊断热斑的准确率。文献[28]利用正常太阳能电池的热像仪图像和 I-V 特性曲线为标准，将故障光伏阵列与正常光伏阵列的热像仪图像和电学特性对比分析，从而诊断出光伏阵列的热斑、破碎、焊带的断裂故障。文献[29]充分结合光伏阵列的红外和可见光图像进行缺陷诊断，将目标诊断算法与分割算法通过区域匹配以及二次分类模块组合起来，实现光伏组件目标框与红外异常区域的匹配，按是否存在温度异常和异物遮挡识别出存在缺陷的组件位置。文献[30]利用安装在无人机上的热像仪和电荷耦合器件 CCD 对光伏阵列进行扫描，再通过数字图像和信号处理诊断出光伏阵列中的不同故障。文献[31]提出了一种新的框架，该框架将可见光图像分割和红外图像故障定位相结合，能够适应几乎所有的亮度条件，同时引入了深度学习方法，从而完成光伏阵列的故障诊断。文献[32]采用红外热成像相机对光伏模块进行扫描，接着运用图像处理技术从热图像中提取故障相关特征，进而识别并判定光伏阵列中存在的各类故障。

红外图像分析法在检测热斑故障时具有较高的精确度，并且其优势在于无需

进行复杂的参数计算。然而，这种方法在诊断光伏阵列的短路、开路以及老化等其他类型故障时，准确性相对较低，同时增加了光伏电站的经济成本和系统复杂度^[33]，而且该方法局限于光伏阵列处于稳定的环境中，如果光伏阵列受环境温度影响较大时，红外诊断法将受到很大程度上的限制。

文献[34]通过分析直流电压的自平衡状态及其等效电路模型，并针对各类故障状况优化模式组合，以构建直流电压传感器在故障情况下的穿越运行调制策略。文献[35]针对大规模光伏阵列，采用多传感器法用于光伏阵列故障检测；并提出了一种全局分块逐步逼近法来对光伏阵列进行模块化处理，在对光伏阵列模块化处理的基础上，统计分析各阵列块的电压数据，求出每块平均值，通过评价电压波动来判断光伏组件是否存在故障。文献[36]通过在光伏阵列中安装电流电压传感器来实现故障诊断，并分析传感器安装位置的不同对光伏阵列故障诊断的影响。文献[37]提出了基于 GridSearchCV 优化的传感器优化布置故障诊断模型，利用多传感器方法测得的故障数据可以作为定位量对故障模块进行定位，从而诊断出光伏阵列的故障类型。文献[38]提出了一种故障定位策略，它不需要在每个光伏串都安装电流和电压传感器，而是通过电压传感器区分光伏阵列的字符串数量的奇偶性来部署，以准确地定位光伏阵列中的故障光伏组件。

利用在光伏阵列中广泛部署的传感器网络进行故障检测，这种方法虽然能够有效地识别并确诊光伏阵列的各类故障，但对于光伏电站的运行来说，增加了安装传感器的成本，而且传感器的故障也是一大处理难题，这就导致光伏阵列故障诊断的成本较高。

(3)智能诊断法

智能诊断法是仿真建模与智能学习算法相结合的光伏阵列健康状态评估方法，该方法对采集到的光伏阵列特征参数进行处理，通过对特征参数进行训练，构建光伏阵列故障诊断模型，以实现光伏阵列故障的准确诊断^[39-40]。智能诊断法涵盖了神经网络及多种机器学习方法，展现出了卓越的适应能力。它们能够有效处理复杂性高、数据量大以及特征呈现非线性的问题，因此在光伏阵列的故障诊断领域得到了广泛的应用。

文献[41]利用改进麻雀搜索算法优化 RBF 神经网络，用优化后的算法建立 ISSA-RBF 故障诊断模型，进而确定出能够实现组件故障诊断的模型，获得较好

的故障识别结果。文献[42]为了分析光伏组件在运行过程中出现的故障情况，建立布谷鸟搜索算法优化 BP 神经网络光伏组件故障诊断模型，并使用布谷鸟搜索算法寻找 BP 神经网络中的阈值和权值，降低网络对初始值的敏感度，避免网络陷入局部最小，从而识别光伏组件的故障类型。文献[43-45]利用模糊 C 均值聚类算法对光伏阵列的特征参数进行聚类，通过计算样本数据与聚类中心的隶属度大小关系，进而区分光伏阵列中不同故障的问题，但该算法对进行聚类的样本依赖很大，且计算复杂，故障诊断的结果易受光伏阵列工作环境因素的影响。文献[46]利用光伏阵列的特征参数构建 XGBoost 故障诊断模型，依据模型性能度量指标对 XGBoost 故障诊断模型的诊断结果进行评价，能简单、高效地识别光伏阵列的故障类型。文献[47]使用人工蜂群算法识别单二极管模型的电参数，然后实现了基于概率神经网络的光伏阵列故障诊断。文献[48]从预处理后的 I-V 特性曲线中提取新的有效特征，结合随机森林的多标记学习 k 近邻和简单多标记学习残差网络进行光伏阵列的故障诊断。文献[49]集成了自动特征工程和选择技术，处理不平衡数据集进行无偏分类，对光伏阵列的故障诊断有很高的精确度。

智能诊断方法用于光伏阵列故障诊断通常依赖于大量的光伏阵列运行数据来训练模型，这样做可以有效降低实际环境因素引起的不确定性对诊断结果的影响，进而确定出能够实现组件故障诊断的模型，获得较好的故障识别结果。然而，当需要获得必要的训练样本数据时，这种方法变得比较具有挑战性，如在训练故障模型之前，都必须依赖过往的经验对故障样本进行分类，故障样本的选择直接影响了故障模型的准确性，而且算法本身也存在着一定的限制。

1.3 论文框架

本文实现的光伏阵列健康状态评估的技术框架如图 1-4 所示。为了评估光伏阵列的健康状况，首先需要建立光伏阵列的仿真模型。然后，对不同状态下的光伏阵列进行仿真，得到光伏阵列不同状态下的仿真 I-V 和 P-V 特性曲线以及最大功率点电压和功率的数据。再对得到的光伏阵列 I-V 特性曲线进行灰色关联分析，划分出不同状态下光伏阵列的健康等级。

然后，提取不同状态下光伏阵列的最大功率点数据，对最大功率点数据进行归一化处理，针对亚健康或异常等级下的光伏阵列，采用模糊 C 均值聚类算法对

相应的光伏阵列数据进行聚类分析，从而诊断出光伏阵列的故障类型。同时，利用概率神经网络对不同状态下光伏阵列最大功率点特征参数的数据集进行训练，建立光伏阵列故障诊断模型，并与基于灰色关联分析与 FCM 算法的评估结果进行比较验证。

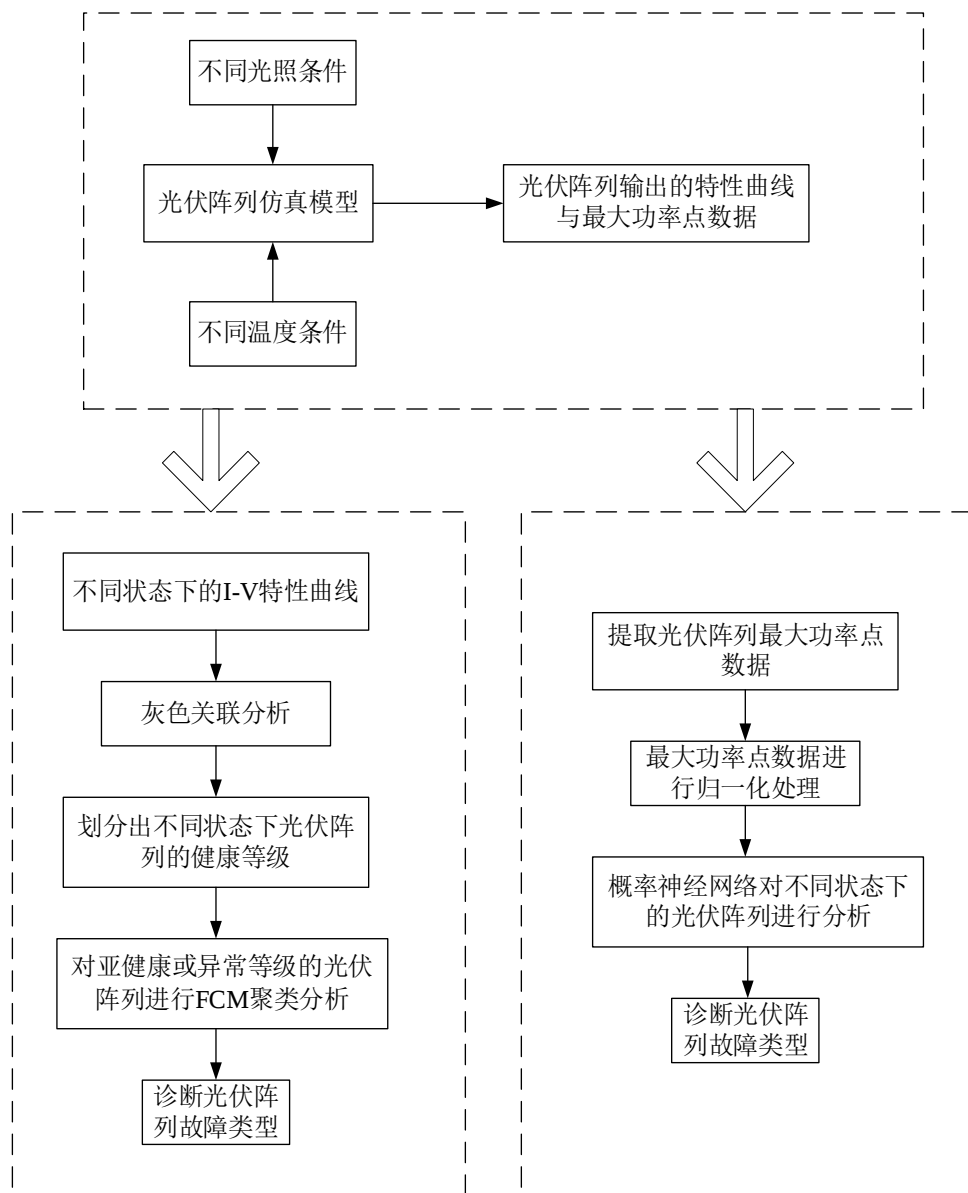


图 1-4 论文内容框架

Fig1-4 Content framework of the paper

1.4 本文结构组织

第 1 章：绪论。详细阐述了光伏阵列健康状态评估的研究背景及其重要意义，对我国新能源开发的战略目标和政策措施进行了分析，探讨了光伏发电的现状和未来潜力。并介绍了国内外相关研究的进展情况。最后，概述了本文的核心研究内容和章节安排。

第 2 章：光伏发电系统分析与光伏阵列建模。详细阐述了光伏发电的工作原理、相关的数学模型以及光伏发电系统的种类。同时，通过采用 MATLAB/SIMULINK 建立了 3×3 光伏阵列仿真模型，并对它在正常状态与故障状态进行了仿真，并分析了这些状态下光伏阵列输出特性曲线的特征。

第 3 章：基于灰色关联分析的光伏阵列健康等级划分。简单介绍灰色关联分析的原理以及对光伏阵列健康等级进行划分的原则；基于建立的仿真模型，通过变化光伏阵列仿真模型的外部激励条件对不同故障下光伏阵列进行仿真，并提取光伏阵列不同工作状态下的 I-V 特性曲线；最后，对得到的 I-V 特性曲线与正常运行下的光伏阵列 I-V 特性曲线进行灰色关联分析，得到不同工作状态下光伏阵列的健康等级。

第 4 章：基于 FCM 的光伏阵列故障诊断。引入模糊 C 聚类算法，对不同状态下的光伏阵列进行分类，并讨论光伏阵列特征参数的选取，再分析划分出健康等级的光伏阵列，对于亚健康或异常等级下的光伏阵列，利用模糊 C 均值聚类算法进行聚类分析，得到聚类中心，利用聚类中心与测试数据代入高斯隶属度函数，从而诊断出光伏阵列的故障类型。

第 5 章：基于概率神经网络的故障诊断。详细阐述了概率神经网络的相关理论。同时，通过利用光伏阵列的仿真模型，模拟了光伏阵列正常状态以及各种故障状态，并提取不同状态下光伏阵列的特征参数数据集。随后，对这些数据进行归一化处理，建立了概率神经网络故障诊断模型，用以诊断不同状态下光伏阵列的故障类型。并对利用概率神经网络诊断不同状态下光伏阵列所得到的结果与基于灰色关联分析与 FCM 算法对光伏阵列健康状态评估的结果进行比较验证。

第 6 章：总结与展望。概括了本文的研究要点，并指出所采用方法的局限性与改进空间，并对未来的研究进行了展望。

第 2 章 光伏发电系统分析与光伏阵列建模

在光伏发电系统中，光伏阵列是其核心组成部件。实际工作中的光伏阵列的性能受很多条件的影响，有外界环境的影响，如日照强度和组件工作温度；有光伏阵列系统部件异常状况等因素的影响，如 EVA 老化、钢化玻璃和电池之间的粘合剂失效等，光伏阵列输出特性随着不同因素的变化而呈现出较大差异。

对光伏阵列健康状态的研究，首先要充分学习了解光伏发电系统的组成部分和工作原理，分析是由于什么原因而导致光伏阵列处于不同的健康状态，同时分析光伏阵列在不同健康状态下的差别，依据光伏阵列不同健康状态下的特征对其进行评估。

2.1 光伏发电理论

光伏发电技术基于太阳能电池的光电效应，通过这一原理将太阳能转换为电能，只要太阳能电池得到充足的阳光照射，便能够产生电压和电流^[50]。太阳能电池主要由硅组成，硅并不是导电体，也不是绝缘体，只有在特定的条件下导电。为改善导电性，可在其中掺入部分杂质，如磷、硼等，因含有其他元素致使半导体内部电荷不平衡，从而形成 PN 结。在太阳光照射下，这类半导体内部电子获得能量，从而激发出自由电荷，这些自由电荷不断的定向移动而形成电压差；在通过将太阳能电池两端接通形成回路，电路中就会形成电流。与此同时，光伏电池面板的尺寸、数量都会直接影响到光伏电池的输出电流和输出电压；因此，为了满足实际应用的需求，一般的光伏组件都是通过对若干数量的光伏电池进行串并联而构成。光伏电池原理构造图如图 2-1 所示。

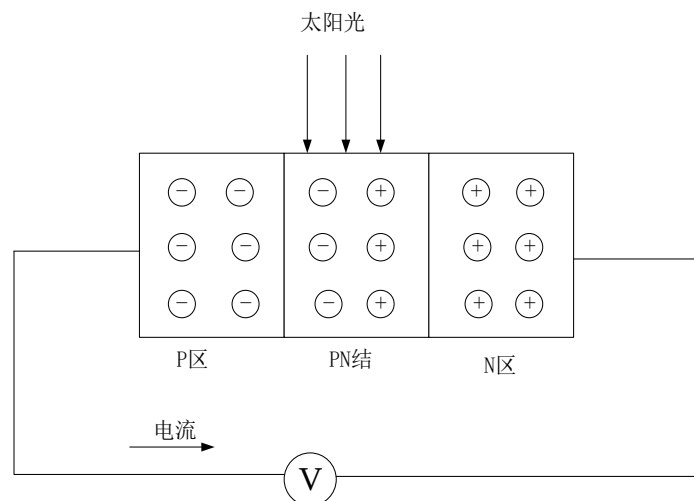


图 2-1 光伏发电原理图

Fig2-1 Schematic diagram of PV power generation

2.2 光伏发电系统类型

光伏发电系统主要有两种类型：离网光伏发电系统和并网光伏发电系统。

2.2.1 离网光伏发电系统

离网光伏发电系统是独立运行的发电系统，不与电网相连接，其系统结构如图 2-2 所示，主要包含光伏阵列、控制器、逆变器和储能元件等部件^[51]。离网光伏发电系统主要用于偏僻山区、海岛等离电网较远的区域。

对于离网光伏发电系统，光伏阵列发出的电可以直接储存在储能元件中或者直接对负荷供电。该系统的安装条件很容易满足，安装地点只需有充足的日照即可。因此，使其成为偏远地区缺乏电力供应的理想选择；同时，它也能充当电力不稳定区域的紧急备用能源解决方案^[52]。

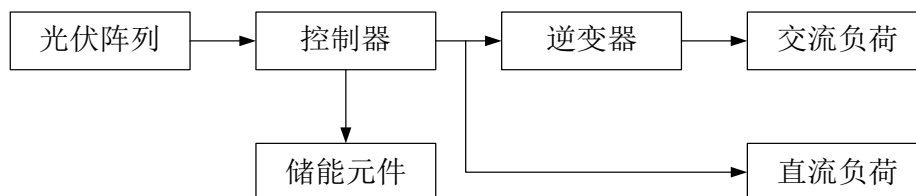


图 2-2 离网光伏发电系统

Fig2-2 Off-grid PV power generation system

2.2.2 并网光伏发电系统

并网光伏发电系统能够与电网相连接，其系统结构如图 2-3 所示。并网光伏发电系统又可分为集中式电站和分布式电站，集中式电站利用大面积的光伏阵列进行发电，通常远离负荷中心，通过输电线路与负荷以及电网相连接。分布式电站是在用户侧建设的光伏电站，能够自发自用，也可以接入电网。

并网式的光伏发电系统主要包含光伏阵列、逆变器和计量装置等电气设备。并网光伏电站只要有充足的光照都能够运行，而且能够安装在住家房顶、商业建筑或者工厂建筑房顶等这些常见地点，它还能够与电网相连实现电量互补；因此，并网光伏发电系统得到广泛应用^[53]。

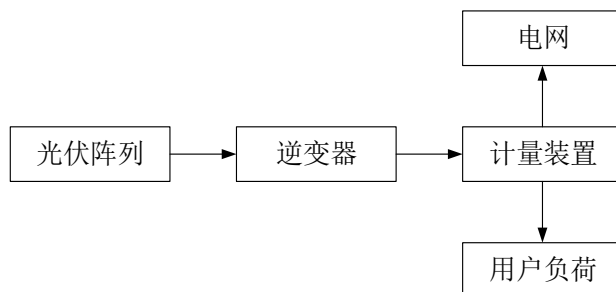


图 2-3 并网光伏发电系统

Fig2-3 Grid-connected PV power generation system

2.2.3 光伏发电系统中的设备

(1)光伏阵列

由于单个光伏组件产生的电流、电压偏小，为了产生逆变器所需要的额定电压和额定电流，需要将一定数量的光伏组件串并联构成相应的光伏阵列。实际的光伏电站就是由不同数量的光伏阵列来构成，所以光伏阵列是光伏电站的核心。

(2)控制器

控制器是离网光伏发电系统中的一个关键组件，可以实现对整个系统的监测与控制，也是连接光伏阵列、储能元件和负荷的重要组件。它主要负责管理储能元件的充放电问题，能够确保光伏发电系统的安全运行以及延长储能元件的使用年限。

(3)储能元件

储能元件是用于储存光伏发电产生电能的装置，一般用于离网光伏发电系统

中。储能元件能够存储由光伏发电系统产生的过量电能，当日照不足或在夜间或处于应急状态时，储能元件可以提供电能，确保能源供应的连续和稳定。

(4)逆变器

逆变器的主要功能是将光伏阵列所产生直流电转换为交流电，以满足交流负载的用电需求，因此在光伏发电系统中起着至关重要的作用。

(5)计量装置

电力计量装置主要用于并网光伏发电系统中，一般为双向电表。双向电表主要为准确计量光伏电站输出的电力供应至电网时的电能和当光伏发电无法满足用户需求时，计量从电网输送到用户端的电能。常规的单向电表无法胜任这两种计量需求，因此必须采用具备双向计量功能的智能电表，以实现电能的双向计量。

2.3 光伏阵列建模

2.3.1 光伏电池的数学模型

为了研究光伏阵列的电气输出特性，并评估其健康状况，需要建立光伏电池和光伏阵列的仿真模型。首先对光伏电池进行建模，光伏电池的模型如图 2-4 所示。光伏电池由 P 型半导体和 N 型半导体构成，能通过光照产生电动势，再将电池视为一个理想的电流源，再与电阻串联和并联，形成仿真光伏电池模型^[54]。理想的光伏电池输出电流与光伏电池所接收到的日照强度成比例，其中光生电流是指光伏电池在光照作用下产生的电流。而在光伏电池的内部，总会有一些能量损失发生，由此会导致光伏电池实际输出的电流电压与理想的光伏电池输出的电流电压存在一定的偏差。为了表征这些偏差，引入了一个二极管和两个电阻到模型中，二极管的作用是模拟 PN 结的特性，它具有仅允许电流单向流动的特性；串联电阻表示电池内部的电阻，一般为电池材料产生的电阻；并联电阻表示电流泄漏引起的等效电阻^[55]。

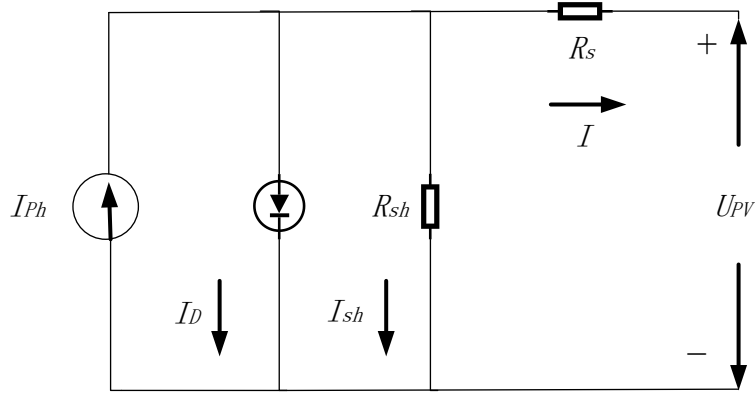


图 2-4 光伏电池单二极管等效模型

Fig2-4 Single diode equivalent model for PV cell

根据基尔霍夫电流定律，可得光伏电池的输出电流 I 如式(2-1)所示：

$$I = I_{ph} - I_D - I_{sh} \quad (2-1)$$

式中： I 为输出电流； I_{ph} 为光生电流； I_D 为二极管电流； I_{sh} 为并联电阻支路电流。其中并联电阻支路电流 I_{sh} 如式(2-2)所示：

$$I_{sh} = \frac{U_{PV} + IR_s}{R_{sh}} \quad (2-2)$$

式中： U_{PV} 为输出电压； R_s 为内部等效串联电阻； R_{sh} 为内部等效并联电阻。

根据肖克利二极管方程计算得出， I_D 具体如式(2-3)所示：

$$I_D = I_0 \left\{ \exp \left[\frac{q(U_{PV} + IR_s)}{AKT} \right] - 1 \right\} \quad (2-3)$$

式中： I_0 为二极管反向饱和电流； q 为电子电荷量，其数值为 $1.6 \times 10^{-19} \text{ C}$ ； A 为二极管理想因子，其数值一般在 1 和 2 之间； K 为波尔兹曼常数，其数值为 $1.38 \times 10^{-23} \text{ J/K}$ ； T 为光伏电池内部温度。

对式(2-1)-(2-3)进行联立计算，得到光伏电池的输出电压与输出电流关系如式(2-4)所示：

$$I = I_{ph} - I_0 \left\{ \exp \left[\frac{q(U_{PV} + IR_s)}{AKT} \right] - 1 \right\} - \frac{U_{PV} + IR_s}{R_{sh}} \quad (2-4)$$

建立的数学模型详细描绘了光伏电池的输出特性，计算结果也较为准确，但是不可避免的也存在一些不足之处。在运用光伏电池数学模型时，需要知道光生电流，二极管反向饱和电流，并联和串联电阻等参数的数值，但是这些参数与光

伏电池实时所处的环境有关，若要获得准确的数值则有一定困难^[56]。光伏电池无法实时的反应出光伏电池输出特性的变化，只能先由已知温度和日照强度等条件先计算出光生电流，二极管反向饱和电流，并联和串联电阻等数值后，再根据这些数据求出输出电流，其过程冗杂且准确度低。为了能更好更快的求出输出特性，现对光伏电池的数学模型及公式进行优化处理，如式(2-5)所示：

$$I = I_{sh} \left[1 - C_1 \left(\exp(U/C_2 U_{PV}) - 1 \right) \right] \quad (2-5)$$

另一方面在最大功率点处， $U_{PV}=U_m$ ， $I=I_m$ ，则式(2-5)可转换为式(2-6)。

$$I_m = I_{sc} \left[1 - C_1 \left(\exp\left(\frac{U_m}{C_2 U_{oc}}\right) - 1 \right) \right] \quad (2-6)$$

考虑到开路状态下， $I=0$ ， $U_{PV}=U_{oc}$ ，则式(2-5)可转换为式(2-7)。

$$0 = I_{sc} \left[1 - \left(1 - \frac{I_m}{I_{sc}} \right) \exp\left(-\frac{U_m}{C_2 U_{oc}}\right) \times \left(\exp\left(\frac{1}{C_2}\right) - 1 \right) \right] \quad (2-7)$$

式(2-6)与式(2-7)联立求解，可解出 C_1 与 C_2 如式(2-8)与(2-9)所示：

$$C_1 = (1 - I_m/I_{sc}) \exp(-U_m/C_2 U_{oc}) \quad (2-8)$$

$$C_2 = (U_m/U_{sc} - 1) / \ln(1 - I_m/I_{sc}) \quad (2-9)$$

光伏组件的最大功率点电压 U_m 、开路电压 U_{oc} 、最大功率点电流 I_m 、短路电流 I_{sc} 四个参数的取值可以从光伏组件的铭牌参数表得到，进而可以计算得到 C_1 和 C_2 的值。在实际应用中，上式没有考虑到日照强度和工作温度变化的影响，因此通过补偿系数 $a=0.0008/^\circ\text{C}$ ， $b=0.2/(\text{W}/\text{m}^2)$ ， $c=0.002/^\circ\text{C}$ 可以获得在光伏电池仿真中考虑日照强度和温度变化时的公式如式(2-10)-(2-13)所示：

$$I_{sc} = I_{scref} \left[1 + a(T - T_{ref}) \right] \cdot \frac{S}{S_{ref}} \quad (2-10)$$

$$U_{oc} = U_{ocref} \left[1 - c(T - T_{ref}) \right] \ln \left[e + b \left(\frac{S}{S_{ref}} - 1 \right) \right] \quad (2-11)$$

$$I_m = I_{mref} \left[1 + a(T - T_{ref}) \right] \cdot \frac{S}{S_{ref}} \quad (2-12)$$

$$U_m = \left[1 - c(T - T_{ref}) \right] \ln \left[e + b \left(\frac{S}{S_{ref}} - 1 \right) \right] \quad (2-13)$$

式中： S 为光伏电池接收到的实际日照强度； S_{ref} 为标准条件下的日照强度； T 为光伏电池实际工作温度； T_{ref} 为标准条件下的工作温度。

2.3.2 基于单二极管的光伏阵列数学模型

通常情况下，单个光伏电池产生的电流和电压都非常有限，在实际光伏电站中无法直接利用。因此，需要对光伏电池进行合理的串并联连接，从而构成光伏阵列，使得总的输出功率可以满足不同的需求^[57]。

为了构建一个光伏阵列，本文采取了将 N_s 个光伏电池串联在一起，并与 N_p 个光伏电池并联的方式；得到的光伏阵列单二极管数学模型如图 2-5 所示。

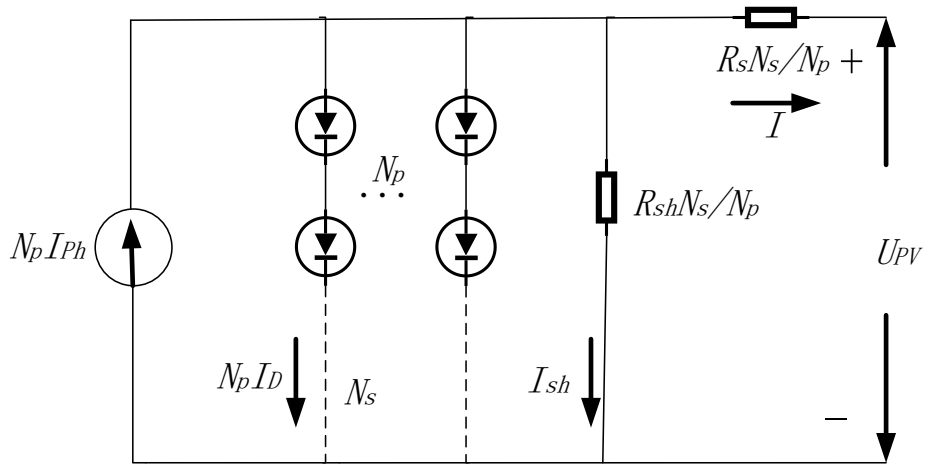


图 2-5 光伏阵列单二极管数学模型

Fig2-5 Mathematical modelling of single diodes for PV array

根据光伏阵列单二极管数学模型可以得到光伏阵列输出电压与输出电流的关系如式(2-14)所示：

$$I = N_p I_{ph} - N_p I_0 \left\{ \exp \left[\frac{q(U_{PV} + I R_s N_s / N_p)}{N_s A T K} \right] - 1 \right\} - \frac{U_{PV} + I R_s N_s / N_p}{R_{sh} N_s / N_p} \quad (2-14)$$

式中： I 为输出电流； $N_p I_{ph}$ 为光生电流； I_0 为二极管反向饱和电流； U_{PV} 为输出电压； $R_s N_s / N_p$ 为串联电阻； $R_{sh} N_s / N_p$ 为并联电阻； q 为电子电荷量； $N_s A$ 为二极管理想因子； K 为波尔兹曼常数； T 为光伏电池内部温度。

由式(2-4)与式(2-14)可以看出，光伏电池与光伏阵列的输出电流和输出电压之间关系的区别仅仅在于参数大小的不同，而它们的表达方式是完全一样的。

2.3.3 光伏阵列建模与输出特性分析

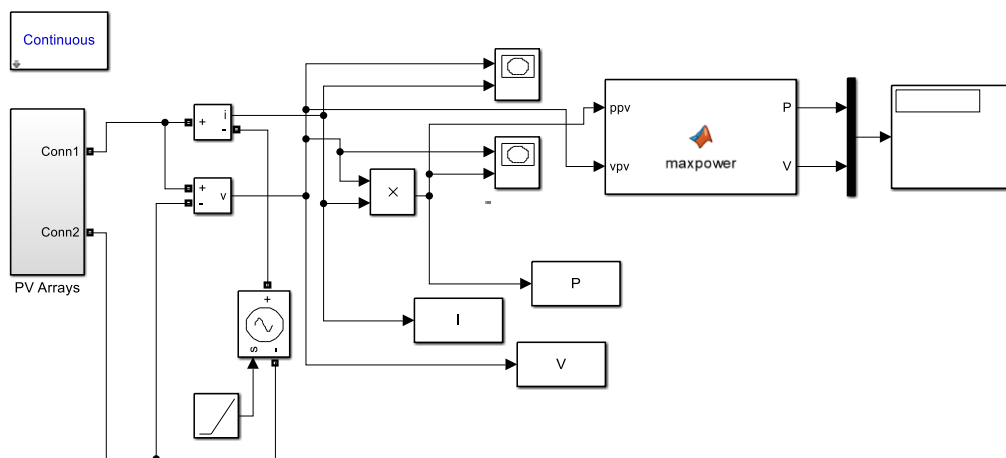


图 2-6 光伏阵列仿真模型图

为模拟光伏阵列仿真模型,将 9 个光伏组件串并联为光伏阵列之后进行封装后接入电路,其中 3×3 光伏阵列仿真模型如图 2-7 所示。

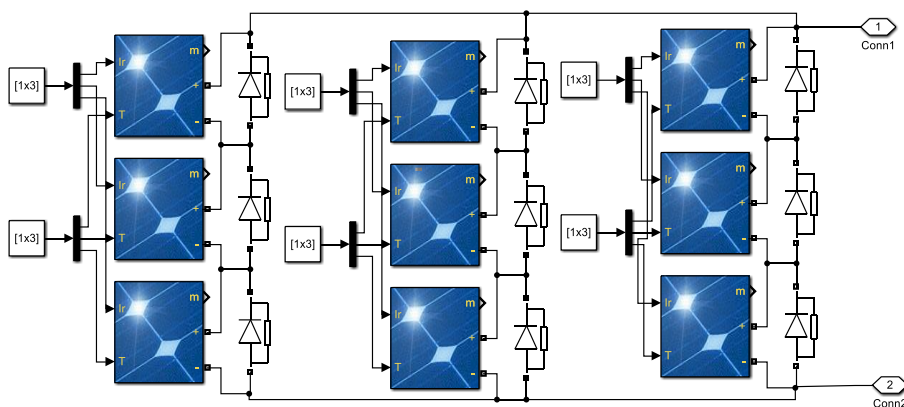


图 2-7 3×3 光伏阵列仿真模型

光伏阵列仿真模型中每个光伏组件都能单独设置日照强度和组件工作温度的数值。其中，所使用的 AavidSolarASMS-165P 光伏组件参数如表 2-1 所示。

表 2-1 光伏组件 ASMS-165P 在标准条件下的电气参数

Table2-1 Electrical parameters of the PV module ASMS-165P under standard conditions

参数	数值
最大功率 P_m/W	164.85
开路电压 U_{oc}/V	43.5
短路电流 I_{sc}/A	5.25
最大功率电压 U_m/V	35
最大功率电流 I_m/A	4.71
短路电流温度系数 $\alpha/(%/^{\circ}C)$	0.03
开路电压温度系数 $\beta/(%/^{\circ}C)$	-0.39301

为模拟光伏阵列不同的状态，可将不同组件设置不同日照强度来模拟局部阴影状态或者串联电阻来模拟光伏阵列老化状态；或将不同组件短路和不同组串开路，分别模拟出光伏阵列的短路与开路状态。而且该模型还可以直接获得光伏阵列在不同状态时的最大功率以及最大功率点电压。

当该光伏阵列运行在日照强度为 $1000W/m^2$ ，工作温度为 $25^{\circ}C$ 的标准运行情况下时，运行仿真模型后，可以获得光伏阵列的 I-V 特性曲线和 P-V 特性曲线，如图 2-8 与图 2-9 所示。可以看出光伏阵列在标准运行状态下，随着电压上升电流缓慢降低，当电压上升到某个临界点时光伏阵列对外输出的电流迅速减小到零，而光伏阵列输出功率随着电压上升先增加然后到最高点后下降，且光伏阵列只有一个最大功率点^[58]。

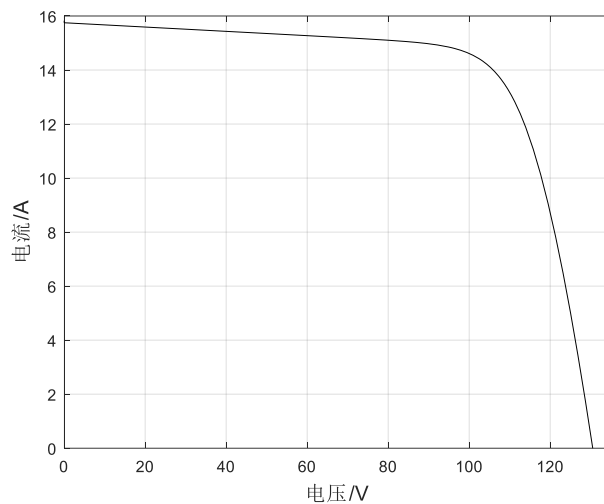


图 2-8 光伏阵列标准运行情况下的 I-V 特性曲线

Fig2-8 I-V characteristic curves of PV arrays under standard operating conditions

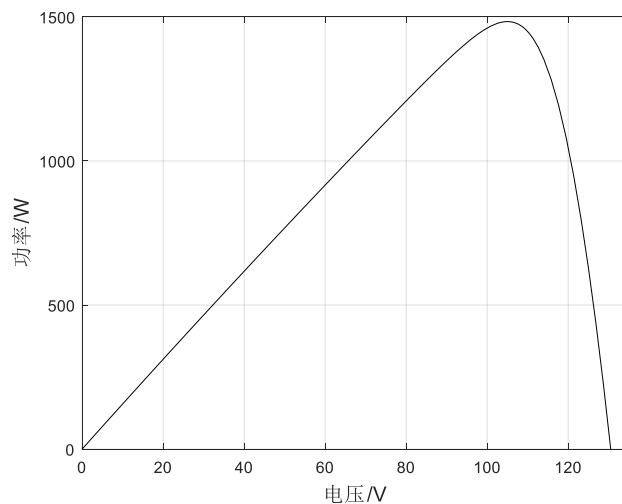


图 2-9 光伏阵列标准运行情况下的 P-V 特性曲线

Fig2-9 P-V characteristic curves of PV arrays under standard operating conditions

在此标准运行情况下，该仿真还可以得到光伏阵列的最大功率点电压 U_m 和最大功率 P 等基本电气参数，如图 2-10 所示最大功率点电压 U_m 和最大功率 P 可以直接读取出来。通过对比分析制造商提供的光伏组件参数并与仿真结果进行对照，结果完全匹配，这一过程验证了所建立的光伏阵列仿真模型的正确性。

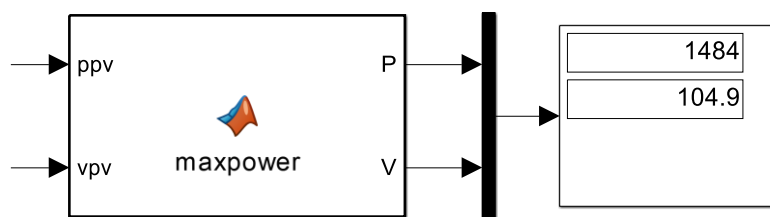


图 2-10 光伏阵列标准运行情况下的最大功率点数据

Fig2-10 Maximum power point data for standard operation of PV arrays

2.4 光伏阵列故障输出特性分析

(1) 局部阴影状态

在光伏发电系统中，光伏阵列发生局部阴影是最普遍的现象。造成光伏阵列局部阴影的主要原因是光伏阵列受到遮挡，如光伏阵列使用过程中的部分灰尘遮挡，光伏阵列表面的有树叶和垃圾的遮挡等，使部分组件较长时间处于阴影状态。对于一些小型光伏发电系统，大多建设于建筑楼顶，建筑的女儿墙会使光伏阵列产生局部阴影；对于大型的光伏发电系统，不同的季节与天气也有可能使光伏阵列发生局部阴影的情况。

光伏组件若存在遮挡现象，则输出电流会显著低于未遮挡部分的组件。这种情况下，被遮挡的光伏组件具有的电阻上会有电流通过；如果不对其进行处理，被遮挡的光伏组件会产生热量，长时间积累下来会导致阴影组件温度急剧上升，极大地影响光伏阵列的正常运行，甚至可能引发火灾，导致整个阵列被烧毁。

本节对光伏阵列局部阴影仿真如图 2-11 所示，其输入数据如图 2-12 所示。光伏阵列中一个组件的日照强度降为 500 W/m^2 ，同时根据需要设置其他的值，其他光伏组件不变为 1000 W/m^2 ，且所有组件的工作温度设置为 25°C 。

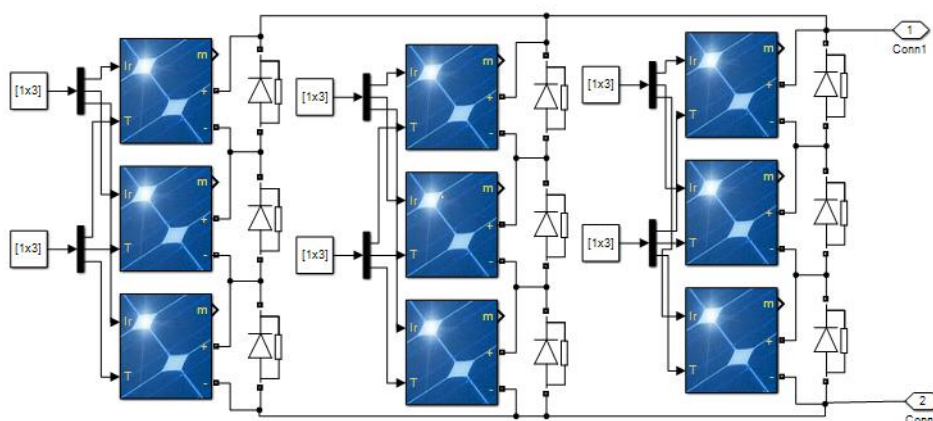


图 2-11 光伏阵列局部阴影仿真模型

Fig2-11 Local shadow simulation model of PV array

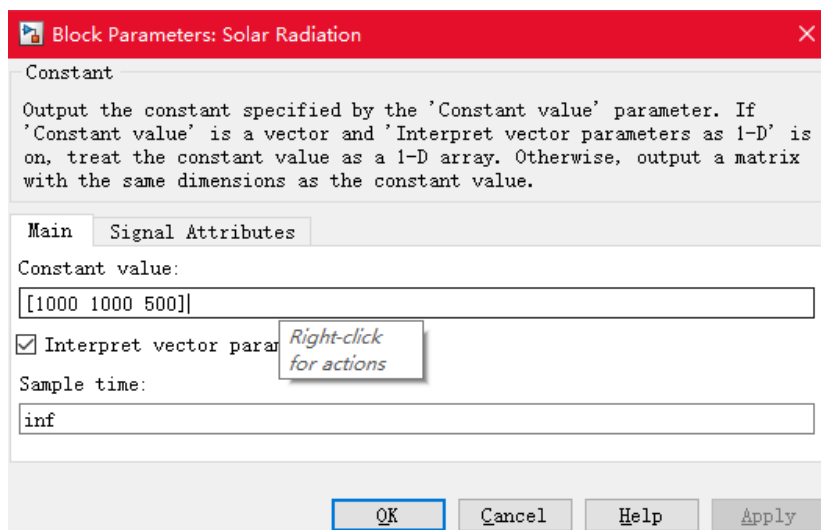


图 2-12 光伏阵列局部阴影仿真输入数据

Fig2-12 Input data of PV array local shadow simulation

局部阴影状态下光伏阵列的 I-V 特性曲线和 P-V 特性曲线如图 2-13 与图 2-14 所示。当光伏阵列在正常运行的情况下，其 I-V 特性曲线和 P-V 特性曲线均只存在一个极值点。当光伏阵列处于局部阴影状态时，输出的 I-V 特性曲线出现阶梯现象和 P-V 特性曲线出现多峰现象。光伏阵列的 I-V 特性曲线出现阶梯状态是因为个别组件受到的日照强度降低，导致输出的电流减小，从而引起曲线阶梯现象。光伏阵列 P-V 特性曲线出现多峰现象是因为光伏组件工作于不同日照强度下，不同组件的最大功率点不同。

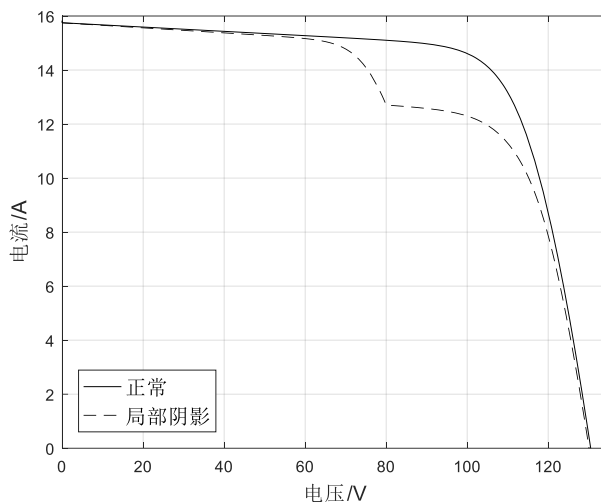


图 2-13 局部阴影状态下光伏阵列 I-V 特性曲线

Fig2-13 I-V characteristic curve of PV array under local shadowing condition

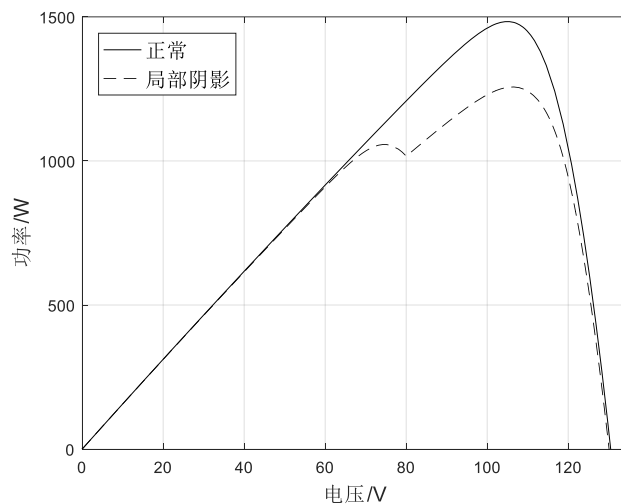


图 2-14 局部阴影状态下光伏阵列 P-V 特性曲线

Fig2-13 P-V characteristic curve of PV array under local shadowing condition

(2)老化状态

光伏阵列一般会有很长的服务寿命，大约在 25 到 30 年左右。由于光伏阵列建设在户外环境下，在光伏阵列长期的使用下，光伏阵列中各组件的封装材料的老化；光伏组件钢化玻璃出现分层现象，导致潮湿气体进入组件内部，使组件内阻增加；酸雨、盐雾等外部恶劣环境对光伏组件的材料造成腐蚀，使光伏组件加速老化。光伏阵列发生老化使得光伏阵列输出功率降低，且没有明显的故障特征，长期不维护会造成光伏阵列的整体寿命降低，严重老化甚至会造成组件内部或组件间的短路。

光伏阵列的老化状态一般表现为串联电阻 R_s 值增大。本节通过在光伏阵列支路上串联一个 $4\ \Omega$ 电阻模拟光伏阵列老化, 该仿真模型如图 2-15 所示, 其输入的数据如图 2-16 所示。仿真模型中光伏组件受到的日照强度为 $1000\ \text{W/m}^2$, 工作温度为 $25\ ^\circ\text{C}$ 。

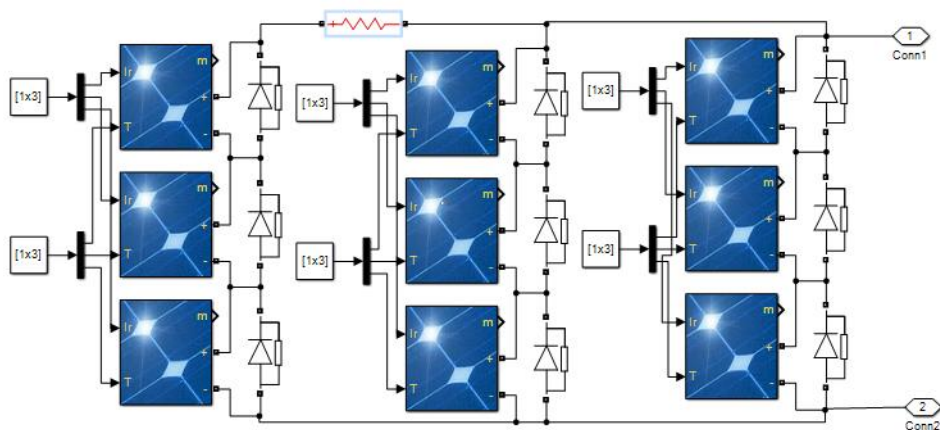


图 2-15 光伏阵列老化仿真模型

Fig2-15 PV array aging simulation model

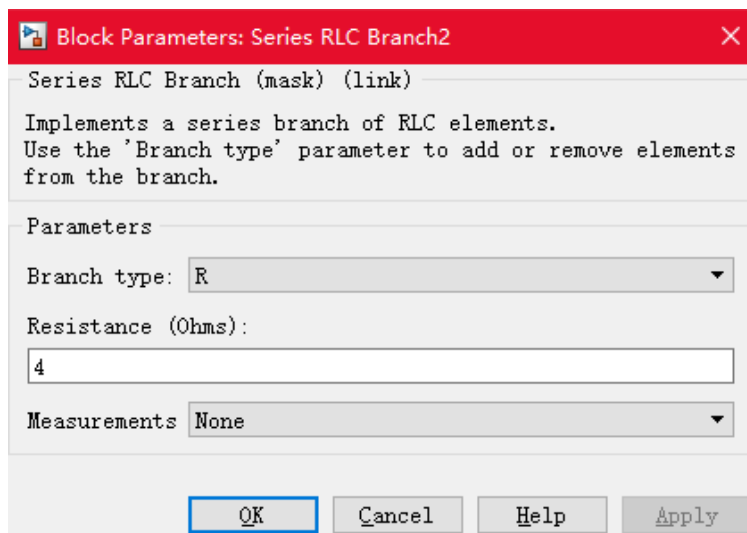


图 2-16 光伏阵列老化仿真输入数据

Fig2-16 PV array aging simulation input data

老化状态下光伏阵列的 I-V 特性曲线和 P-V 特性曲线如图 2-17 与图 2-18 所示。从仿真结果可以看出, 当光伏阵列发生老化时, I-V 特性曲线整体下移, 在相同电压下, 输出电流降低; P-V 特性曲线最大功率点功率值降低, 这表明光伏阵列在老化的状态下, 它能够提供的最大输出功率减少了。

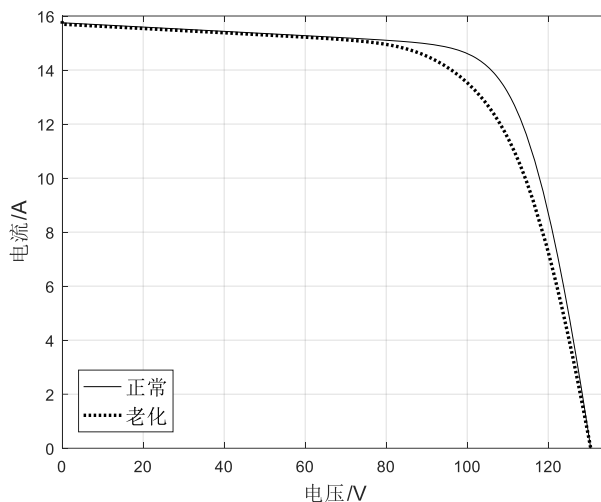


图 2-17 老化状态下光伏阵列 I-V 特性曲线

Fig2-17 I-V characteristic curve of PV array under aging condition

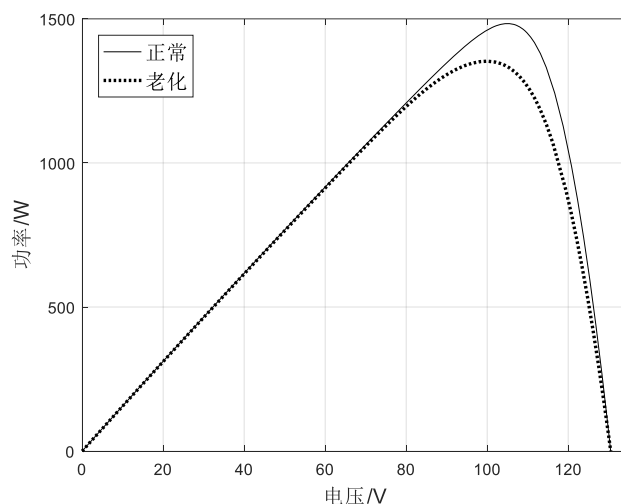


图 2-18 老化状态下光伏阵列 P-V 特性曲线

Fig2-18 P-V characteristic curve of PV array under aging condition

(3)短路故障

光伏阵列发生短路故障是由于光伏组件在电路中发生短路引起的。光伏阵列发生短路故障主要分为外部原因和内部原因，外部原因主要为环境的影响，如受恶劣环境引起的光伏组件绝缘失效和动物咬伤线路或人为破坏线路等情况，进而引发短路；内部原因则是生产操作影响，如生产或安装过程中接线错误导致电流无法正常流通或组件缺乏维护导致损坏而发生短路等情况。若长时间短路，会减小光伏阵列输出效率，甚至造成组件过热引发火灾，对整个光伏阵列产生影响。

当光伏阵列出现短路故障时，对应光伏组件无法正常输出功率，因此在仿真

模型中将一个或多个组件短路来模拟光伏阵列中的短路故障。本节首先设置光伏组件受到的日照强度为 1000 W/m^2 ，工作温度为 25°C ，再通过短路一个光伏组件仿真模型光伏阵列的短路状态，该仿真模型如图 2-19 所示。

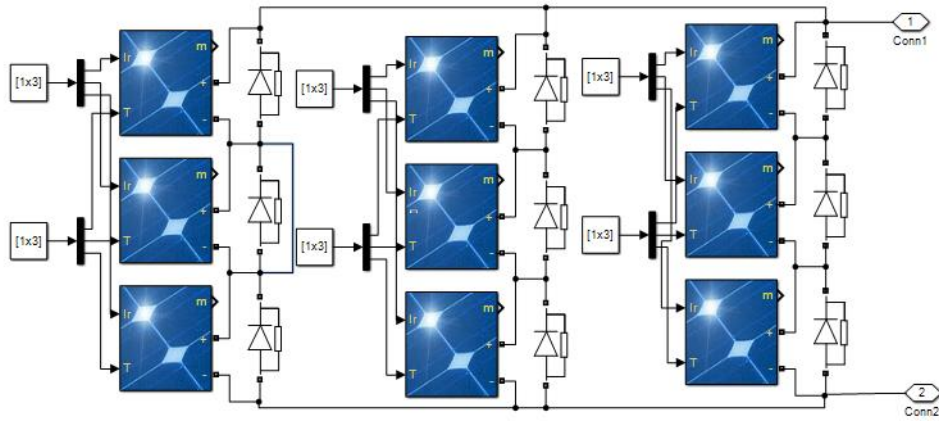


图 2-19 光伏阵列短路故障仿真模型

Fig2-19 Simulation model of short circuit fault of PV array

短路故障状态下光伏阵列的 I-V 特性曲线和 P-V 特性曲线如图 2-20 与图 2-21 所示。如果在光伏阵列中有一个组件发生短路，那么该组件的输出电流将会降至零。在这种情况下，短路光伏阵列的 I-V 特性曲线的短路电流基本不变，而开路电压则有所降低。同时，在 P-V 特性曲线上，最大功率点会有显著下降，这将导致整体的输出功率减少。

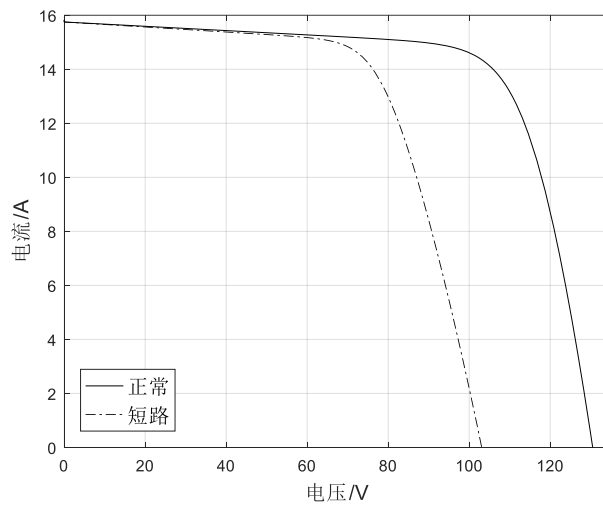


图 2-20 短路故障状态下光伏阵列 I-V 特性曲线

Fig2-20 I-V characteristic curve of PV array under short-circuit fault condition

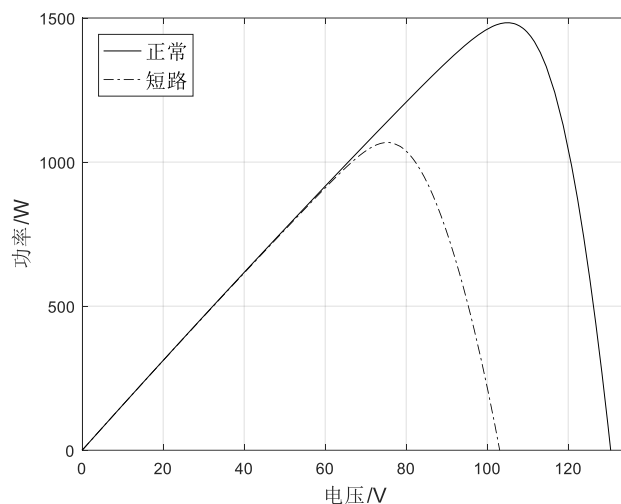


图 2-21 短路故障状态下光伏阵列 P-V 曲线

Fig2-21 P-V characteristic curve of PV array under short-circuit fault condition

(4)开路故障

光伏阵列开路一般是由于操作失误或者年久老化等原因造成光伏电池之间的连接线断开，光伏阵列开路会使得故障所在的整个光伏组串没有输出电流，从而造成光伏阵列的输出电流减小和输出功率减小。光伏阵列产生开路故障的原因也有很多，如自然环境的侵蚀影响，操作不当以及组件内部的断线等。为模拟光伏阵列出现开路故障，首先设定各组件受到的日照强度为 1000 W/m^2 ，工作温度为 25°C 。在通过开路一个光伏组串来仿真模型光伏阵列的开路状态，该仿真模型如图 2-22 所示。

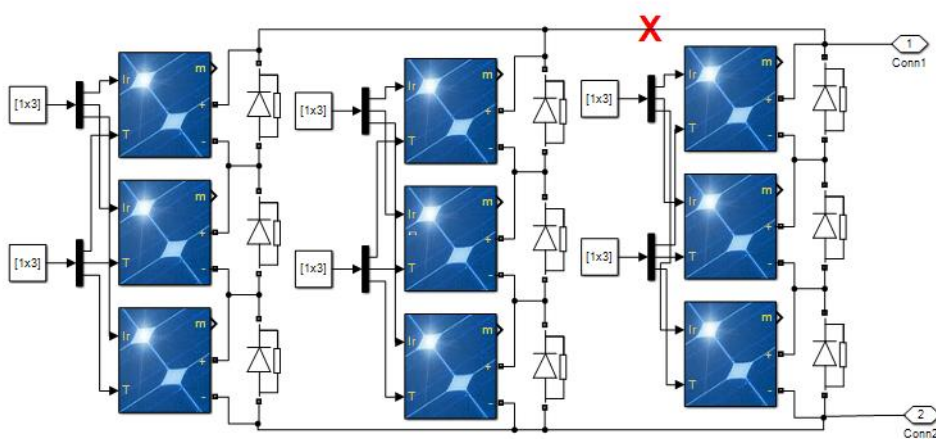


图 2-22 光伏阵列开路故障仿真模型

Fig2-22 Simulation model of open-circuit faults in PV arrays

开路故障状态下光伏阵列的 I - V 特性曲线和 P - V 特性曲线如图 2-23 与图 2-24 所示。光伏阵列发生开路故障时， I - V 特性曲线的短路电流会明显减少，因为开路组件中没有电流流过，而开路电压基本保持不变； P - V 特性曲线中的最大功率下降，而且最大功率点的位置也发生变化。

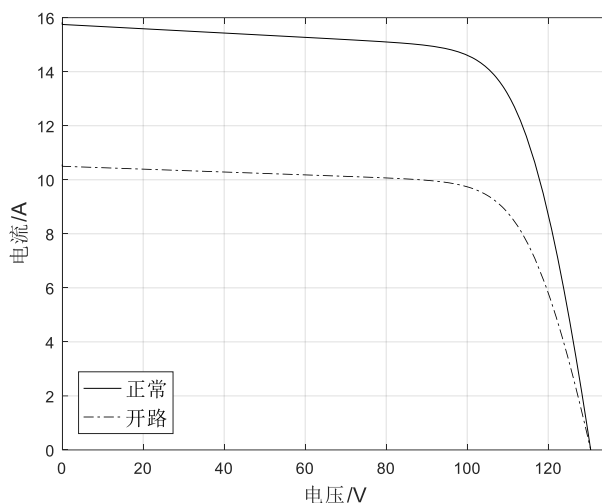


图 2-23 开路故障状态下光伏阵列 I - V 特性曲线

Fig2-23 I - V characteristic curve of PV array under open-circuit fault condition

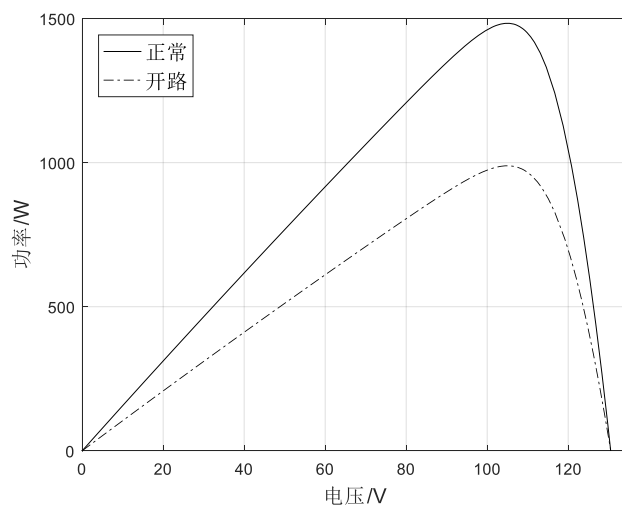


图 2-24 开路故障状态下光伏阵列 P - V 特性曲线

Fig2-24 P - V characteristic curve of PV array under open-circuit fault condition

2.5 本章小结

本章主要对光伏系统的基本理论知识进行了介绍,首先介绍了光伏发电的原理、光伏发电系统的类型以及光伏发电系统中各组成结构的介绍。然后,建立光伏阵列仿真模型。通过 MATLAB/SIMULINK 搭建 3×3 的光伏阵列模型,并得到该模型在标准运行状态与各种故障状态的输出特性。并将各种故障状态与标准运行状态的输出特性曲线作对比,分析了各种故障下基本电气参数的变化特征对其输出特性的影响。同时,通过光伏阵列的仿真,为后续的光伏阵列健康状态评估的基础。

第3章 基于灰色关联分析的光伏阵列健康等级划分

3.1 灰色关联分析原理

灰色关联分析是根据参考数据与比较数据构成曲线的相似程度大小，从而判断两曲线间关联度的方法。该方法用于衡量两条曲线之间的相互关联性，完成系统内有关数据曲线的对比分析，求出参考数据与各比较数据之间的灰色关联度，以此来确定它们之间的关联强度。与参考数据关联度越大的比较数据，其变化趋势与参考数据越接近；与参考数据关联度越小的比较数据，其变化趋势与参考数据越远^[59]。灰色关联度分析对数据要求较低，且计算量小。灰色关联分析的适用性不受数据规律限制，避免了定量分析结果与定性分析结果之间的不一致。针对实际运行的光伏阵列，采用标准运行状态下的仿真计算结果作为参考数据，故障状态下的数据作为比较数据。灰色关联分析的基本步骤如下：

(1)根据光伏阵列健康等级划分的目的来确定参考序列和比较序列。

其中参考序列设置为：

$$X_0 = (x_{01}, x_{02}, \dots, x_{0j}, \dots, x_{0n}) \quad (3-1)$$

原始评价矩阵设置为：

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_m \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix} \quad (3-2)$$

式中： m 为样本个数； n 为特征个数， x_{ij} 为第 i 个样本时刻第 j 个特征的值。

(2)为了处理不同数据量纲带来的数值大小差异，我们需要对数据进行归一化处理。这一步骤将不同范围和单位的数据调整到一个共同的数值区间，通常是 $[0,1]$ ，以便减少数值大小的直接影响，确保数据比较和分析的准确性。

(3)对已经进行归一化处理的数据进行分析，首先计算每个数据点与参考序

列中相应数据点之间的绝对误差，并从这些误差值中识别出最大误差和最小误差。

(4)通过式(3-3)计算出每个参考数据和比较数据对应的相关系数：

$$\xi_{ij} = \frac{\min_i \min_j |x_{ij} - x_{0j}| + \rho \max_i \max_j |x_{ij} - x_{0j}|}{|x_{ij} - x_{0j}| + \rho \max_i \max_j |x_{ij} - x_{0j}|} \quad (3-3)$$

式中： ξ_{ij} 为相关系数，其值为小于 1 的正数； x_{0j} 和 x_{ij} 为参考数据和测量数据； ρ 为分辨系数，取值区间为[0,1]，通常设置为 0.5。

(5)将所得相关系数通过式(3-4)进行平均，得到的平均数便是灰色关联度。

$$r = \frac{1}{m} \sum_{j=1}^m \xi_{ij} \quad (3-4)$$

3.2 光伏阵列健康状态分析

本文将光伏阵列的健康状态分为四个等级，分别为健康、亚健康、异常和故障。各健康状态等级的定义如下：

健康：光伏阵列的运行状态稳定且各参数都处于正常范围。

亚健康：光伏阵列部分参数存在轻微波动，但仍处于可接受范围内，对光伏阵列正常运行无明显影响。若情况持续恶化，可能会影响光伏阵列的正常运行，可按计划对光伏阵列进行监测和维护。

异常：光伏阵列的健康状态处于亚健康与故障之间，仍能正常工作。如果对其不进行处理则会演变为故障状态，如能及时处理则可恢复至健康状态。

故障：光伏阵列出现了短路与开路故障或者使光伏阵列输出性能极度下降的不可恢复的故障，对光伏阵列需要进行维修或更换器件的状态。

所提出的基于 I-V 特性曲线的健康等级划分方法如图 3-1 所示。首先，对光伏阵列进行仿真得到标准运行状态下和故障状态下的光伏阵列的 I-V 特性曲线，并将标准运行状态下的仿真 I-V 特性曲线作为参考 I-V 特性曲线，故障状态下的 I-V 特性曲线作为比较 I-V 特性曲线。然后，对标准运行状态下的仿真 I-V 特性曲线与故障状态下的仿真 I-V 特性曲线进行数据处理，再利用灰色关联分析法量化光伏阵列标准运行状态下的仿真 I-V 特性曲线与故障状态下的仿真 I-V 特性曲线之间的差异。最后，应用隶属匹配划分出光伏阵列的健康等级。

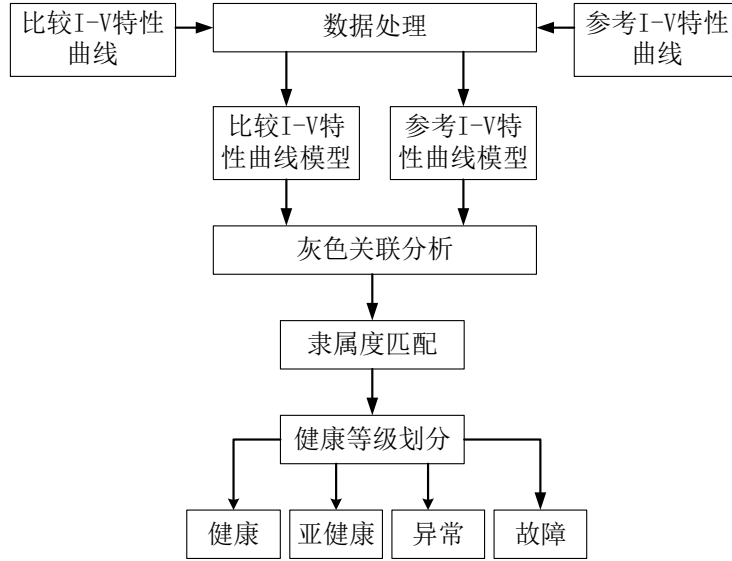


图 3-1 基于灰色关联分析的健康等级划分流程图

Fig3-1 Flow chart of health classification based on grey relational analysis

采用上述灰色关联分析方法计算参考 I-V 特性曲线与比较 I-V 特性曲线之间的灰色关联度，为了使灰色关联度与健康等级划分匹配，本文选取的健康等级隶属度函数为三角隶属度函数，如式(3-5)-(3-8)所示：

$$u_1(GRD) = \begin{cases} 0 & (GRD \leq c) \\ \frac{GRD - c}{d - c} & (c < GRD < d) \\ 1 & (GRD \geq d) \end{cases} \quad (3-5)$$

$$u_2(GRD) = \begin{cases} 0 & (0 < GRD < b) \\ \frac{GRD - b}{c - b} & (b \leq GRD < c) \\ \frac{d - GRD}{d - c} & (c \leq GRD < d) \\ 0 & (GRD \geq d) \end{cases} \quad (3-6)$$

$$u_3(GRD) = \begin{cases} 0 & (0 < GRD < a) \\ \frac{GRD - a}{b - a} & (a \leq GRD < b) \\ \frac{c - GRD}{c - b} & (b \leq GRD < c) \\ 0 & (GRD \geq c) \end{cases} \quad (3-7)$$

$$u_4(GRD) = \begin{cases} 1 & (GRD \leq a) \\ \frac{b - GRD}{b - a} & (a < GRD < b) \\ 0 & (GRD \geq b) \end{cases} \quad (3-8)$$

式中： u_1 、 u_2 、 u_3 和 u_4 分别为健康状态、亚健康状态、异常状态以及故障状态的隶属度； a 、 b 、 c 和 d 分别为三角隶属度函数的边界，取 $a=0.5$ 、 $b=0.6$ 、 $c=0.8$ 、

$d=0.9$ 。

通过三角隶属度函数的应用，计算出了光伏阵列对应于不同健康等级的隶属程度，其中隶属度数值越高，意味着系统当前状态与相应健康等级的匹配程度越高。为了更准确地表示光伏阵列的健康等级，引入健康指数。健康指数计算公式如式(3-9)所示：

$$HI = U_H \cdot V = (u_1, u_2, u_3, u_4) \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} = u_1 v_1 + u_2 v_2 + u_3 v_3 + u_4 v_4 \quad (3-9)$$

式中： U_H 为健康、亚健康、异常和故障等级下的隶属度值； V 为系数集合，取 $(v_1, v_2, v_3, v_4)^T = (0.9, 0.5, 0.3, 0.1)^T$ 。

HI 的范围为 $[0,1]$ ，当光伏阵列的健康指数大于 0.8 时健康等级为健康，此时光伏阵列能够正常工作，无需对光伏阵列进行处理；当健康指数在 0.6-0.8 之间为亚健康，当健康指数在 0.5-0.6 之间为异常，需对亚健康或异常的光伏阵列进行相应的诊断和处理，让其能够正常工作，避免光伏阵列发展为故障状态；当健康指数小于 0.5 时为故障，光伏阵列已经处于不可恢复的状态，必须对故障的光伏阵列进行维修。

3.3 不同状态下光伏阵列健康等级划分

3.3.1 光伏阵列的局部阴影状态

局部阴影本质来说就是光伏阵列中部分组件接收到的日照强度降低或者是进光量减少，当光伏阵列发生局部阴影时，光伏阵列内部一个或者多个光伏组件的日照强度明显低于其他光伏组件。本文通过降低光伏阵列中一个组件受到的日照强度来模拟光伏阵列的局部阴影状态，使用 MATLAB/SIMULINK 搭建的局部阴影状态下的仿真模型如图 2-11 所示。其中以在标准运行状态下的光伏阵列为参考，同时此状态也为正常状态。局部阴影 1、局部阴影 2 和局部阴影 3 的光伏阵列中一个组件受到的日照强度分别为正常状态下日照强度的 80%、50%和 30%，而其他组件受到的日照强度与正常状态下一致，由仿真得到上述四种情况的 I-V 特性曲线如图 3-2 所示。

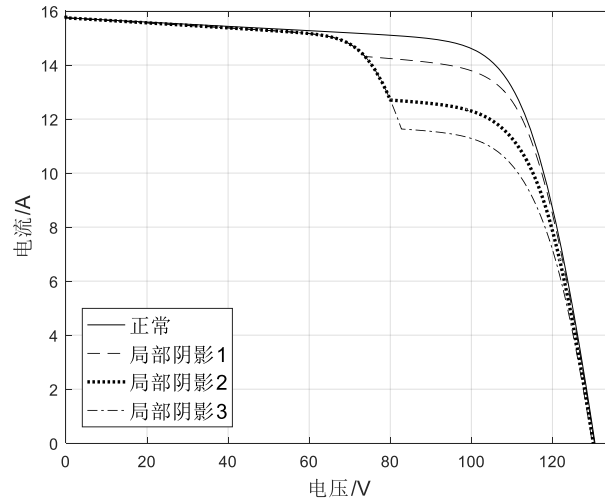


图 3-2 不同局部阴影状态下光伏阵列 I-V 特性曲线

Fig3-2 I-V characteristic curves of PV arrays under different local shadowing states

采用灰色关联分析法对仿真得到的 I-V 特性曲线进行分析，再根据式(3-5)-(3-9)进行健康等级划分，计算结果如表 3-1 所示。

表 3-1 局部阴影状态光伏阵列健康等级划分结果

Table3-1 The results of the health classification of the PV array in the local shadow state

工作状态	GRD	u_1	u_2	u_3	u_4	健康指数	健康等级
正常	1	1	0	0	0	0.9000	健康
局部阴影 1	0.8771	0.7710	0.2290	0	0	0.8084	健康
局部阴影 2	0.8382	0.3820	0.6180	0	0	0.6528	亚健康
局部阴影 3	0.8112	0.1120	0.8880	0	0	0.5448	异常

由表 3-1 可知，当光伏阵列中的一个组件受到的日照强度为正常状态下日照强度的 80%时，光伏阵列的健康指数大于 0.8，此时光伏阵列仍处于健康状态，因为光伏阵列中一个组件受到的日照强度减少的程度很低时，对光伏阵列整体的输出功率影响较小。随着阴影的加重，光伏阵列中一个组件受到的光照逐渐减少，光伏阵列的健康等级逐渐降低，当光伏阵列中的一个组件受到的日照强度为正常状态下日照强度的 50%时，光伏阵列的健康指数降到 0.6528，光伏阵列处于亚健康状态；当光伏阵列中的一组件受到的日照强度为正常状态下日照强度的 30%时，光伏阵列的健康指数降到 0.5448，光伏阵列处于异常状态。可见，在局部阴影的情况下，对光伏阵列的 I-V 特性曲线进行灰色关联分析可以有效划分光伏阵列的健康等级。

3.3.2 光伏阵列的老化状态

本文采用串联外部电阻的方法模拟光伏阵列的老化状态。在仿真光伏阵列的组串之间串联不同大小的电阻来模拟老化的严重程度。使用 MATLAB/SIMULINK 搭建的老化状态下的仿真模型如图 2-15 所示。其中老化状态 1，老化状态 2，老化状态 3 分别在光伏阵列的组串间串联 $2\ \Omega$ ， $4\ \Omega$ ， $6\ \Omega$ 的电阻，而光伏组件受到的日照强度与工作温度分别为 $1000\ \text{W/m}^2$ 与 $25\ ^\circ\text{C}$ ；再以光伏阵列在标准运行状态下为参考，由仿真模拟得到上述三种情况的 I-V 特性曲线如图 3-3 所示。

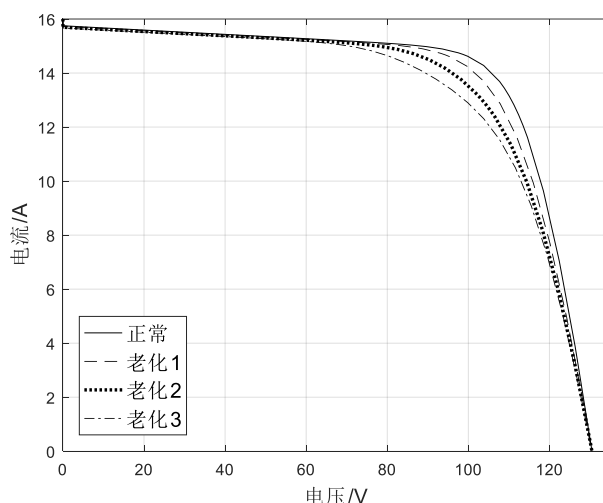


图 3-3 不同老化状态下光伏阵列 I-V 特性曲线

Fig3-3 I-V characteristic curves of PV arrays under different aging states

根据灰色关联分析以及隶属度函数计算出的光伏阵列健康等级如表 3-2 所示。当光伏阵列的组串间串联 $2\ \Omega$ 电阻时，光伏阵列的健康指数为 0.8240，处于健康状态。因此光伏阵列轻微的老化，对光伏阵列的发电效率影响不大。当光伏阵列的组串间串联 $4\ \Omega$ 和 $6\ \Omega$ 的电阻时，光伏阵列的健康指数分别为 0.6448 和 0.5160，随着光伏阵列串联阻值的增加光伏阵列的健康等级从亚健康变为异常，光伏阵列的老化状态变得更加严重，由此可以看出严重的老化对光伏阵列的发电效率影响很大。

表 3-2 老化状态光伏阵列健康等级划分结果

Table3-2 The results of the health classification of the PV array in the aging state

工作状态	GRD	u_1	u_2	u_3	u_4	健康指数	健康等级
老化 1	0.8810	0.8100	0.1900	0	0	0.8240	健康
老化 2	0.8362	0.3620	0.6380	0	0	0.6448	亚健康
老化 3	0.8040	0.0400	0.9600	0	0	0.5160	异常

在老化状态下，灰色关联分析的方法可以有效划分光伏阵列的健康等级。但老化 2 与局部阴影 2 状态下的健康等级都为亚健康，且健康指数只相差 0.008，如在未知光伏阵列状态的情况下并不能区分它的具体故障类型，因此，需要进一步诊断出光伏阵列的状态。

3.3.3 光伏阵列的短路与开路状态

一般情况，光伏阵列发生短路故障，就是由于某一个组件或某几个组件被短接了。因此，本文将光伏阵列中的一个光伏组件或两个光伏组件进行人为短路来模拟短路故障，其光伏组件受到的日照强度与工作温度仍为 1000 W/m^2 与 25°C ；使用 MATLAB/SIMULINK 搭建的短路状态下的仿真模型如图 2-19 所示，得到的光伏阵列短路 I-V 特性曲线如图 3-4 所示。

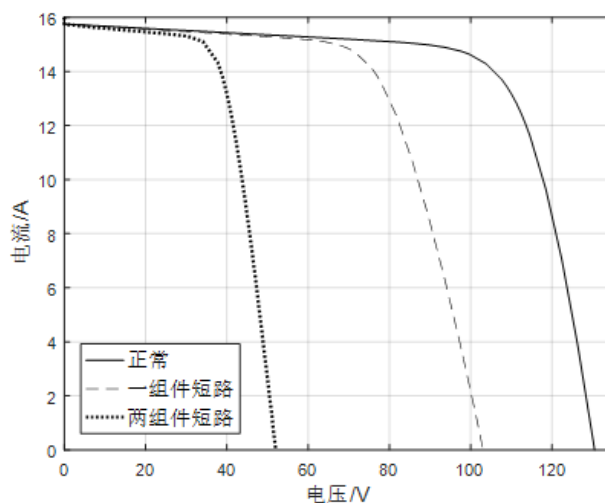


图 3-4 不同短路状态下光伏阵列 I-V 特性曲线

Fig3-4 I-V characteristic curves of PV arrays under different short-circuit states

为模拟光伏阵列的开路状态，将光伏阵列中的一个组串或两个组串进行开路来模拟开路故障，其光伏组件受到的日照强度与工作温度仍为 1000 W/m^2 与 25°C 。使用 MATLAB/SIMULINK 搭建的开路状态下的仿真模型如图 2-22 所示，得到的光伏阵列开路 I-V 特性曲线如图 3-5 所示。

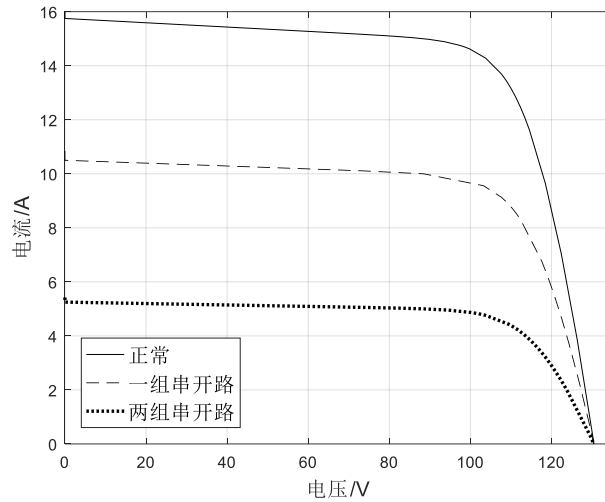


图 3-5 不同开路状态下光伏阵列 I-V 曲线

Fig3-13 I-V characteristic curves of PV arrays in different open-circuit states

根据灰色关联分析以及隶属度函数计算出的光伏阵列短路与开路状态的健康等级如表 3-3 所示。

表 3-3 短路与开路状态光伏阵列健康等级划分结果

Table3-3 Short-circuit and open-circuit state PV array health classification results

工作状态	GRD	u_1	u_2	u_3	u_4	健康指数	健康等级
一组件短路	0.7882	0	0.9410	0.0590	0	0.4882	故障
两组件短路	0.6214	0	0.1070	0.8930	0	0.3214	故障
一组串开路	0.6932	0	0.4660	0.5340	0	0.3932	故障
两组串开路	0.5993	0	0	0.9930	0.0070	0.2986	故障

由表 3-3 可知，光伏阵列的短路与开路状态对光伏阵列的健康状态影响很大，从表中能看出光伏阵列中有组件开路或短路时，该光伏阵列的健康指数都低于 0.5000，健康等级均为故障。对于短路与开路状态，它们的电流与电压相差很大，通过灰色关联分析便能很好的分析出光伏阵列的具体故障类型，因此无需进一步详细诊断。

3.4 本章小结

本章首先介绍了灰色关联分析算法的原理和该算法的计算步骤,并对光伏阵列的健康等级进行定义。同时,对不同状态下的光伏阵列进行仿真,得到不同状态下光伏阵列的 $I-V$ 特性曲线。然后,利用灰色关联分析算法对不同状态下光伏阵列的 $I-V$ 特性曲线进行分析,得到故障状态下光伏阵列与标准运行状态下光伏阵列的灰色关联度,再通过三角隶属度函数以及健康指数对光伏阵列的健康等级进行划分。

第 4 章 基于 FCM 的光伏阵列故障诊断

第 3 章已经通过 I-V 特性曲线的灰色关联分析划分出光伏阵列的健康等级，通过对健康等级的判断，能够诊断出光伏阵列健康与否。但是对于处于亚健康或异常等级下的光伏阵列且具有相同健康等级和健康指数也相近时，并不知道此时光伏阵列的具体状态。因此，本章通过使用模糊 C 均值聚类算法识别不同状态下光伏阵列的特征参数，并对特征参数进行聚类分析，并利用高斯隶属度函数对测试数据进行诊断。

4.1 模糊 C 均值聚类算法原理

传统的聚类算法属于硬聚类，这种算法的结果是固定的，每个数据只会被分到一个聚类簇中，数据不存在中间状态^[61]。硬聚类不需要计算隶属度，对数据划分计算都很快，每种数据的分类也很明确；但是对于数据的重叠或者异常数据很难处理，而本文的光伏阵列特征参数具有很大的重叠性，所以硬聚类不适合对光伏阵列特征参数进行聚类。模糊聚类算法为数据与聚类簇之间赋予了一个隶属度，隶属度的取值范围在 0 到 1 之间，而且每一个数据都以不同程度的隶属度属于多个聚类簇中，但是数据与聚类簇之间的隶属度大小决定了数据与此聚类簇之间的关联程度，如隶属度越是靠近 0，那么数据与此聚类簇几乎没有关联，隶属度越是靠近 1，那么数据基本属于此聚类簇；所以模糊聚类适合对光伏阵列特征参数进行聚类。

FCM 聚类算法是一种经典的软聚类技术，其核心思想是基于数据样本点和目标函数值的计算来实现聚类划分，并能确保不同类别之间的差异最大化。该算法通过迭代计算来更新聚类中心的位置，得到数据集的最优分类；然后，通过计算每个样本点与聚类中心之间的欧式距离，我们可以评估样本点与聚类中心的接近程度。这种距离度量基于样本点在各维度上的差值平方和的平方根。通过这些距离值，我们可以确定样本数据点与它们所属的聚类中心之间的关联性强弱，从而实现对光伏阵列特征参数的自动聚类^[43]。

假设 n 个样本点可表示为 $X=x_i, i=1,2,3\dots n$ ，把 X 划为 C 组，从而获得每一组

的聚类中心 $C=[C_1, C_2, C_3 \dots C_k]$ ，在通过计算测试数据点与不同聚类中心之间的隶属度大小，进行对比分析，即可实现测试数据的归类^[63]。隶属度的取值范围在 $[0,1]$ 之间，并且每一样本点对于所有聚类中心的隶属度之和为 1，即隶属度满足如式(4-1)所示的两个条件^[45]：

$$\begin{cases} u_{ij} \in [0,1] \\ \sum_{i=1}^c u_{ij} = 1, j = 1, 2, 3 \dots n \end{cases} \quad (4-1)$$

式中： u_{ij} 为样本 x_i 属于 j 类的隶属度； c 表示预划分的类别数； n 表示数据总量。

模糊 C 均值聚类是一种基于目标函数的模糊聚类方法^[45]，它的目标函数 J_m 如式(4-2)所示：

$$J_m = \sum_{i=1}^N \sum_{j=1}^c u_{ij}^m \|x_i - c_j\|^2 \quad (4-2)$$

式中： u_{ij} 为样本 x_i 属于 j 类的隶属度； m 为聚类的簇数； c_j 为 j 类的聚类中心； $\|x_i - c_j\|$ 为数据到聚类中心的欧式距离。

聚类是一个不断迭代计算隶属度 U_{ij} 和聚类中心 c_j 的过程。在式(4-2)的每条约束条件上添加拉格朗日乘子 $\lambda = \{\lambda_1, \lambda_2, \lambda_3, \dots \lambda_n\}$ ，将带约束条件的极值问题转换为无约束条件问题，得到的算式如式(4-3)所示：

$$J(U, C, \lambda) = J(U, C) + \sum_{j=1}^n \lambda_j \left(\sum_{i=1}^k u_{ij} - 1 \right) = \sum_{i=1}^N \sum_{j=1}^c u_{ij}^m \|x_i - c_j\|^2 + \sum_{j=1}^n \lambda_j \left(\sum_{i=1}^k u_{ij} - 1 \right) \quad (4-3)$$

对所有输入参数求导使式(4-2)达到最小化的必要条件为式(4-4)和(4-5)。

$$u_{ij} = 1 / \sum_{k=1}^c \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}} \quad (4-4)$$

$$c_j = \sum_{i=1}^N u_{ij}^m x_i / \sum_{i=1}^N u_{ij}^m \quad (4-5)$$

根据式(4-4)和(4-5)反复迭代计算聚类中心和隶属度；当迭代终止时，便可完成模糊聚类。分析表明，FCM 聚类算法是一个不断迭代计算的过程，FCM 算法的聚类流程如图 4-1 所示。

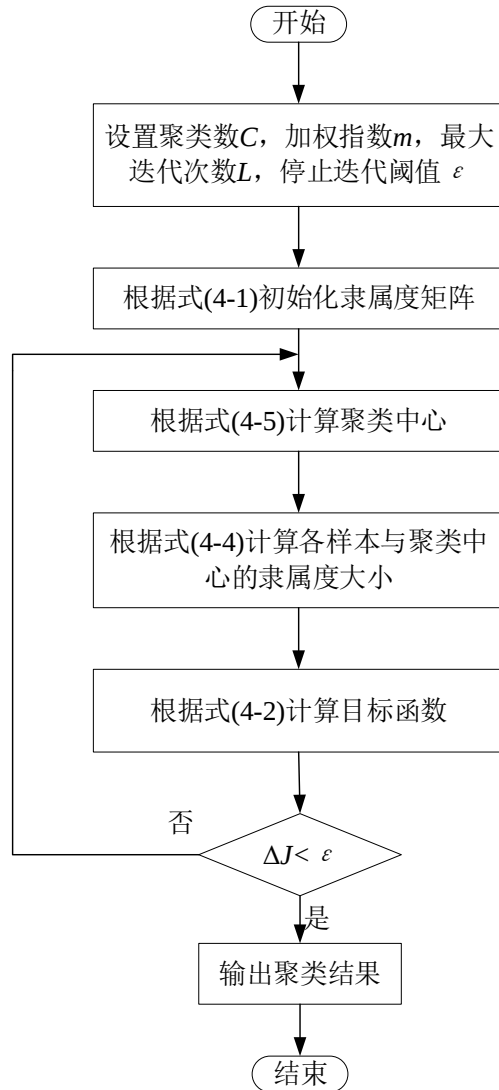


图 4-1 FCM 聚类流程图

Fig4-1 FCM clustering flowchart

基于 FCM 算法确定聚类中心和隶属度矩阵的具体步骤如下:

- (1)对光伏阵列的故障状态进行分析,获得不同故障状态的类别,从而设置聚类数 C 、加权指数 m 、迭代次数 L 以及停止迭代阈值 ε 。
- (2)用 $[0,1]$ 之间的随机数初始化由隶属度函数确定的矩阵 U_{ij} ;
- (3)用式(4-5)计算聚类中心值 C ,并根据式(4-4)计算出测试样本到聚类中心的隶属度大小。
- (4)若每一类别的目标函数值的差值 $\Delta J_m = |J_m - J_{m-1}|$ 在预设的阈值 ε 以内,则满足迭代条件,输出聚类结果;如果不满足 $\Delta J_m < \varepsilon$ 则转向步骤(2)。

对于 FCM 算法可能算法参数的选择而对聚类结果产生影响,因此通常会多

次运行算法，并评估结果的稳定性，以确定最终的聚类方案。FCM 算法的结果通常包括聚类中心和隶属度矩阵，算法结束运行时，会得到一组聚类中心，聚类中心为多维空间的一个点，代表了聚类簇的典型值，在本章中代表光伏阵列不同故障状态输出特征参数的典型值；隶属度矩阵是一个矩阵，对于每个测试数据，他的所有隶属度都为 1。

4.2 选取不同样本特征量的对比与分析

为了描述光伏阵列不同工作状态的特性，需要挑选一些特征参数来概括光伏阵列的运行情况。这些选取的特征参数应当能够充分展现光伏阵列的工作状态，并且尽量减少测量工作量。

根据光伏阵列仿真模型将光伏阵列的工作状态分为如表 4-1 所示的 11 类。F1 为光伏阵列的正常工作状态，F2-F4 为光伏阵列的局部阴影状态，F5-F7 为光伏阵列的老化状态，F8-F9 为光伏阵列的短路状态，F10-F11 为光伏阵列的开路状态。

表 4-1 光伏阵列工作状态描述

Table4-1 Description of the working status of the PV array	
工作状态	描述
F1	日照强度 1000 W/m ² ，温度 25 °C，光伏阵列正常工作
F2	一个组件 800 W/m ² ，温度 25 °C，其余组件正常工作
F3	一个组件 500 W/m ² ，温度 25 °C，其余组件正常工作
F4	一个组件 300 W/m ² ，温度 25 °C，其余组件正常工作
F5	光伏阵列串联 2 Ω 的电阻
F6	光伏阵列串联 4 Ω 的电阻
F7	光伏阵列串联 6 Ω 的电阻
F8	一个组件短路
F9	两个组件短路
F10	一组串开路
F11	两组串开路

本文以光伏阵列为基础诊断单元，将其工作状态分为正常工作、局部阴影、老化、短路与开路共 5 种工作状态。上述每一类工作状态中又可根据光伏阵列仿真模型中设置的不同参数进行细分，并将细分出的不同工作状态下的光伏阵列依次编号为 F1, F2...F11 共 11 种子故障类型，如果对这些子工作状态的光伏阵列进行 FCM 聚类，则获得的模糊聚类簇数目为 C ，其中 C 为 11。

根据上述 11 中光伏阵列的工作状态，使用最大功率点电压 U_m 和最大功率点电流 I_m 作为特征参数进行聚类分析，如图 4-2 所示，可以看出短路与开路状态的光伏阵列很容易区分，而正常状态、老化状态和局部阴影状态的光伏阵列特征参数聚类混叠在一起，难以区分光伏阵列的具体工作状态。

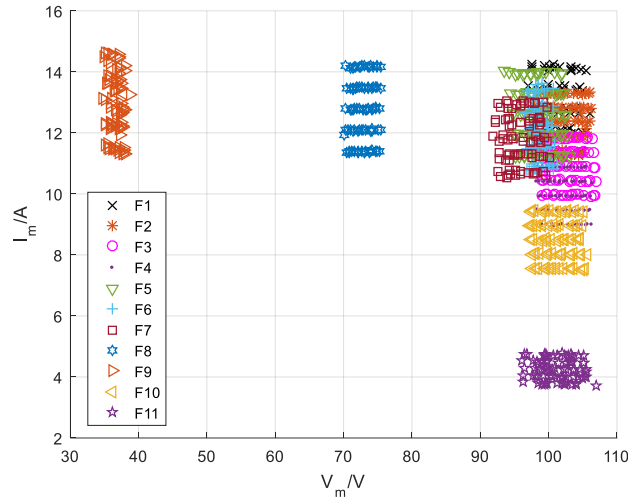


图 4-2 11 种工作状态的特征量(U_m, I_m)分布

Fig4-2 Distribution of characteristic quantities (U_m, I_m) for 11 operating states

为了能够减少聚类时光伏阵列特征参数的重叠数量，在最大功率点电压 U_m 和最大功率点电流 I_m 作为特征参数时进行聚类分析的基础上，加入最大功率 P 作为特征参数进行聚类，如图 4-3 所示。观察结果显示，加入最大功率 P 进行聚类后，样本点之间的重叠程度有所降低。但仍有一部分处于不同工作状态的光伏阵列特征参数点重叠在一起，这也说明了对光伏阵列常规的特征参数进行聚类分析，并不能很好的诊断出光伏阵列的故障类型。

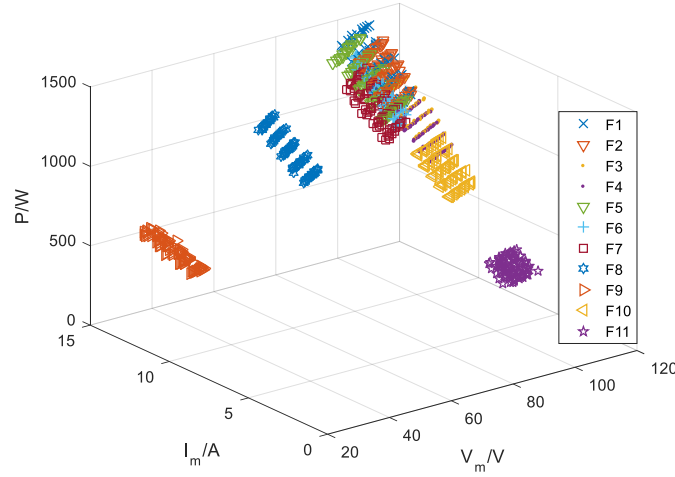

 图 4-3 11 种工作状态的特征量(U_m, I_m, P)分布

 Fig4-3 Distribution of characteristic quantities (U_m, I_m, P) for 11 operating states

为了诊断光伏阵列的状态类型, 需要选择最能反应光伏阵列状态类型的特征参数, 即将如式(4-6)-(4-8)的 3 个变量作为特征参数^[64]:

(1)归一化最大功率点电压 U_{NORM}

$$U_{\text{NORM}} = \frac{U_m}{n_s \times U_{\text{oc-ref}}} \quad (4-6)$$

式中: n_s 为光伏阵列的组件串联数; $U_{\text{oc-ref}}$ 为参考光伏组件的开路电压。

(2)归一化最大功率点电流 I_{NORM}

$$I_{\text{NORM}} = \frac{I_m}{n_p \times I_{\text{sc-ref}}} \quad (4-7)$$

式中: n_p 为光伏阵列的组件并联数; $I_{\text{sc-ref}}$ 为参考光伏组件的短路电流。

(3)填充因子 FF

$$FF = \frac{U_m I_m}{U_{\text{oc}} I_{\text{sc}}} \quad (4-8)$$

式中: U_{oc} 为光伏阵列的开路电压; I_{sc} 为光伏阵列的短路电流。

为了收集光伏阵列在不同运行条件下的特征参数数据集, 进行了针对光伏阵列 C 类状态的仿真实验, 从而获得了在多种工作状态下光伏阵列的特征参数数据样本, 按所需特征参数形成有限样本数据集 X 。其中数据集 X 如式(4-9)所示:

$$X = \begin{bmatrix} U_{\text{NORM}}^1 & I_{\text{NORM}}^1 & FF^1 \\ \vdots & \vdots & \vdots \\ U_{\text{NORM}}^i & I_{\text{NORM}}^i & FF^i \\ \vdots & \vdots & \vdots \\ U_{\text{NORM}}^n & I_{\text{NORM}}^n & FF^n \end{bmatrix} \quad (4-9)$$

式中：X 为样本的特征参数数据集；n 为每个光伏阵列样本的编号。

4.3 高斯隶属度原理与故障诊断流程

FCM 算法的核心思想在于通过不断的迭代计算，形成样本数据的最佳聚类簇。根据公式(4-2)可知，FCM 以欧氏距离来衡量样本数据所处点与该样本所处簇的聚类中心的隶属程度，但这种方法在处理该聚类簇的边界样本数据点时，会受到邻近聚类簇的聚类中心干扰，导致这些样本点的隶属度无法准确反映其与本聚类中心的实际关系。在光伏阵列输出的特征参数数据中，由于日照强度与组件工作温度的影响，光伏阵列输出的特征参数没有明显的规律，且易出现异常数据，这些因素对形成理想的聚类簇有一定的影响，从而降低了聚类效率与聚类的准确度。因此，传统的模糊 C 均值算法存在一些严重问题：一是在处理大量数据时计算复杂，效率低下；二是可能陷入局部最小值而无法收敛；三是仅局限于处理球形数据集，缺乏普适性。所以要引入隶属度函数对这些问题进行解决。

由仿真得到的光伏阵列特征参数在不同的状态下会处于不同的范围内，本章运用高斯隶属度函数来描述这种特点。隶属度函数是模糊集合中的一个重要概念，他可以描述光伏阵列某一测试样本数据对光伏阵列不同状态下的数据形成的聚类簇的隶属程度。高斯隶属度函数曲线如图 4-4 所示。

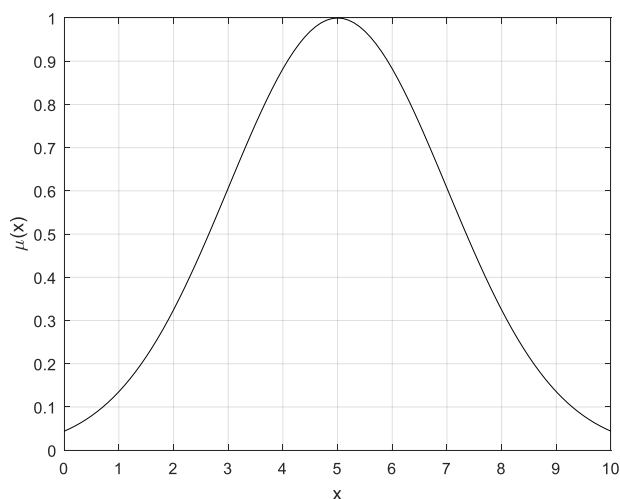


图 4-4 高斯隶属度函数曲线

Fig4-4 Gaussian membership function curve

高斯隶属度函数为：

$$\mu(x) = e^{-\frac{(x-\mu)^2}{2\sigma^2}} \times 100\% \quad (4-10)$$

式中： $\mu(x)$ 为测试数据 x 的隶属度； μ 为高斯分布的均值，也是隶属度函数最高点的位置； σ 为标准差，也是高斯分布的宽度。

根据高斯函数的特性 σ 的值可由如下公式求得：

$$\sigma = \frac{\mu_{\max} - \mu_{\min}}{6} \quad (4-11)$$

式中： $\mu_{\max}$ 和 $\mu_{\min}$ 为光伏阵列特征参数的最大值和最小值。

光伏阵列故障诊断的流程如图 4-5 所示。

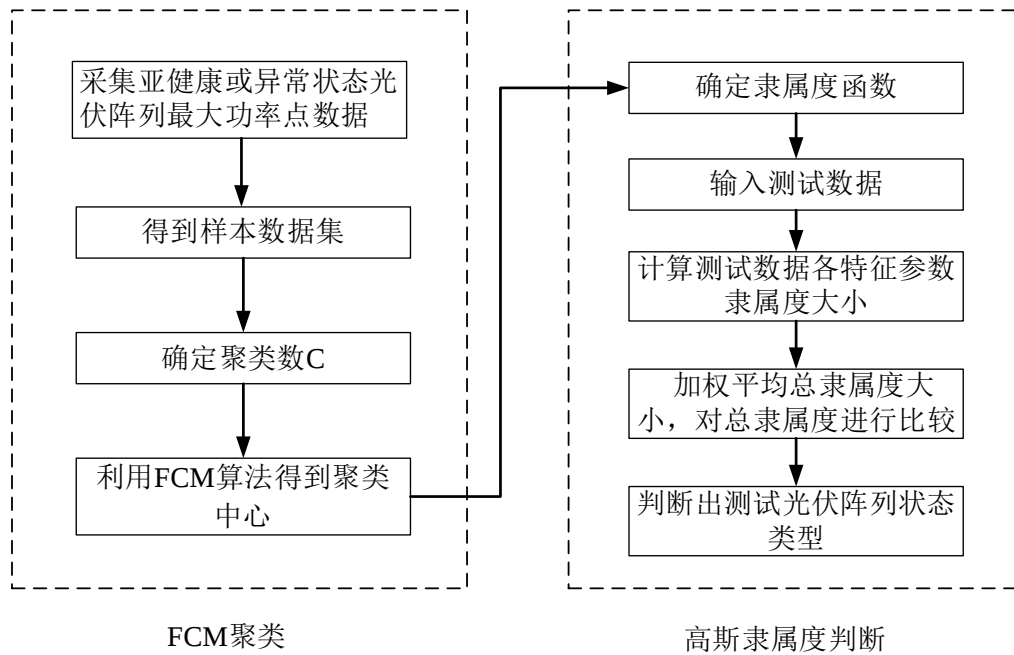


图 4-5 基于 FCM 的光伏阵列故障诊断流程图

Fig4-5 Fault diagnosis flow chart of PV array based on FCM

基于 FCM 的光伏阵列故障诊断分为可以分为两个部分，分别为 FCM 聚类和高斯隶属度函数对光伏阵列的故障进行诊断。详细的诊断过程如下：

(1)采集处于亚健康或异常状态光伏阵列最大功率点样本数据，为了消除环境参数对这些数据的影响，需要对得到的样本数据进行归一化处理，从而得到样本数据集；再对光伏阵列进行故障分析确定聚类数 C ，并将上述样本数据集和故障分类数 C 输入算法中。

(2)利用模糊 C 均值聚类算法对样本数据集进行聚类得到样本数据的聚类簇，进而得到每个聚类簇的聚类中心点。

(3)确定高斯隶属度函数参数，输入步骤(2)获得的聚类中心与需要测试的数据进行光伏阵列故障诊断，计算出测试数据与各个聚类簇的聚类中心之间的隶属度大小。

(4) 加权平均计算得到该测试数据的总隶属度，并比较该测试数据与各种故障状态聚类簇之间的的总隶属度，进而根据测试数据中最高的隶属度来判断并确定样本的故障状态类别，以此作为样本的最终诊断结果。

4.4 FCM 聚类结果分析

4.4.1 FCM 聚类仿真分析

由 4.2 节可以看出使用传统的 FCM 对光伏阵列最大功率点的数据进行聚类, 光伏阵列的各种状态之间存在互相重叠的现象, 难以区分光伏阵列的每一种状态; 而且需要对光伏阵列的每一种状态都进行数据的提取, 而后进行计算, 在使用这种方法时, 不仅耗费大量时间整理数据, 而且计算量很大。本文第 3 章利用 I-V 曲线的灰色关联分析划分出光伏阵列的健康等级, 通过对健康等级的判断, 对于处于亚健康或异常等级下的光伏阵列且具有相同健康等级和健康指数也相近时, 再对相应的光伏阵列进行最大功率点数据的提取, 对其进行 FCM 聚类; 因为灰色关联分析仅需要 I-V 曲线的仿真数据和测量数据, 同时算法的计算量小; FCM 算法相对复杂, 但是通过对 I-V 曲线进行灰色关联分析后, FCM 所需的数据量和计算减少很多, 从而能精确地诊断出光伏阵列的故障类型。

由第 3 章可知老化 2 与局部阴影 2 状态下的健康等级都为亚健康, 且健康指数只相差 0.008, 若在未知光伏阵列状态的情况下需要进一步诊断出光伏阵列的状态。为了诊断出光伏阵列的故障类型, 用图 2-5 所示的仿真模型模拟工作于不同日照强度与工作温度下的光伏阵列, 其中日照强度以 50 W/m^2 为步长从 800-1000 W/m^2 变化, 工作温度以 1°C 为步长从 25-40 $^\circ\text{C}$ 变化。

选取处于亚健康等级下的一光伏阵列作为测试光伏阵列, 该光伏阵列所处状态下得到的数据作为测试数据。利用 FCM 算法诊断测试光伏阵列的故障类型, 通过仿真获取正常, 局部阴影 2 和老化 2 状态的最大功率点数据, 再以 U_{NORM} , I_{NORM} 和 FF 作为特征参数对上述 3 种状态进行聚类。聚类结果如图 4-6 所示。

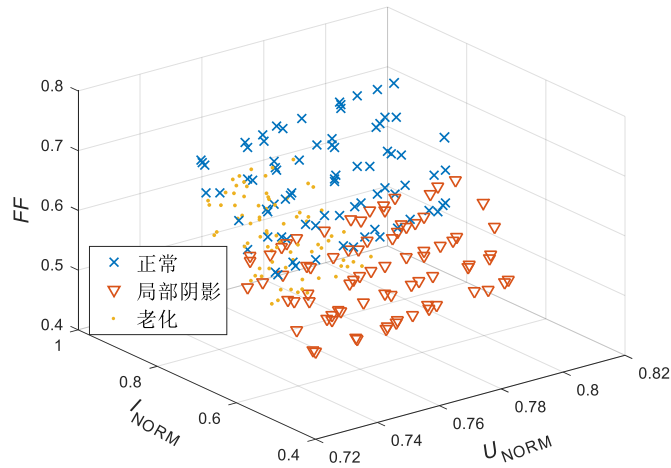


图 4-6 FCM 聚类结果

Fig4-6 FCM clustering results

基于上述 3 种状态, 可设置聚类数目 $C=3$, 加权指数 $m=2$, 最大迭代次数 $L=1000$, 停止迭代阈值 $\varepsilon=10^{-5}$ 。FCM 聚类算法的输入数据为上述 3 种状态的仿真数据, 利用 FCM 聚类算法获得聚类中心, 获得的聚类中心结果如表 4-2 所示。

表 4-2 不同状态光伏阵列的聚类中心

Table4-2 Clustering centers of PV arrays in different states

健康等级	工作状态	聚类中心
健康	正常	(0.7680, 0.8494, 0.6524)
亚健康	局部阴影	(0.7728, 0.6424, 0.4963)
亚健康	老化	(0.7674, 0.7422, 0.5695)

4.4.2 高斯隶属度诊断

将表 5 中不同状态光伏阵列的聚类中心与高斯隶属度算法结合, 计算测试数据与各聚类中心的隶属度, 从而诊断出光伏阵列的状态类型。按照式(4-11)的标准差 σ 计算方法, 计算 3 种状态特征参数的标准差为 $\sigma_{U_{\text{NORM}}} = 0.0124$, $\sigma_{I_{\text{NORM}}} = 0.0515$ 和 $\sigma_{FF} = 0.0456$ 。由此得到高斯隶属度函数如式(4-12)所示:

$$\left\{ \begin{array}{l} \mu(U_{\text{NORM}}^i) = e^{-\frac{(U_{\text{NORM}}^i - \mu U_{\text{NORM}}^Q)^2}{2 \times 0.0124^2}} \times 100\% \\ \mu(I_{\text{NORM}}^i) = e^{-\frac{(I_{\text{NORM}}^i - \mu I_{\text{NORM}}^Q)^2}{2 \times 0.0515^2}} \times 100\% \\ \mu(FF^i) = e^{-\frac{(FF^i - \mu FF^Q)^2}{2 \times 0.0456^2}} \times 100\% \\ \mu_T = \frac{1}{3} [\mu(U_{\text{NORM}}^i) + \mu(I_{\text{NORM}}^i) + \mu(FF^i)] \end{array} \right. \quad (4-12)$$

式中： U_{NORM}^i 、 I_{NORM}^i 和 FF^i 为第 i 个测试数据的归一化最大功率点电压，归一化最大功率点电流和填充因子； U_{NORM}^Q 、 I_{NORM}^Q 和 μFF^Q 为 FCM 算法得到的聚类中心值， μ_T 为前述各隶属度的加权平均值。

将测试光伏阵列的最大功率点参数代入式(4-6)-(4-8)中得到归一化的特征参数 $U_{\text{NORM}}=0.7572$ 、 $I_{\text{NORM}}=0.7795$ 、 $FF=0.5867$ 再将归一化的特征量代入式(4-12)中得到与各状态聚类中心的隶属度，如表 4-3 所示。

表 4-3 测试数据与各聚类中心之间的隶属度

Table4-3 Affiliation between the test data and the clusters

工作状态	$\mu(U_{\text{NORM}}^i)$	$\mu(I_{\text{NORM}}^i)$	$\mu(FF^i)$	μ_T
正常	46.71%	39.81%	35.42%	40.65%
局部阴影	26.88%	2.89%	14.01%	14.59%
老化	9.53%	76.93%	93.13%	73.20%

由表 4-3 可以看出测试数据经计算后得到的总隶属度在老化状态时是最大值，所以测试光伏阵列是由于老化而处于亚健康等级下。所取测试光伏阵列的诊断结果与预设结果一致，证明了 FCM 能够诊断出光伏阵列的故障类型。

再用图 2-5 所示的仿真模型在日照强度和工作温度条件不变的情况下对光伏阵列的正常、局部阴影、老化状态进行模拟，生成新的测试样本，并对其进行诊断。对生成的新的测试样本诊断分析后，得到的结果如表 4-4 所示。

表 4-4 3 种工作状态诊断结果

Table4-4 Diagnostic results of 3 operating states

工作状态	测试样本	正确诊断	诊断准确率
正常	10	9	90%
局部阴影	20	19	95%
老化	20	18	90%

由表 4-4 可以看出，在进行的 50 个样本的测试中，有 4 个样本的故障诊断结果不准确，准确率达到了 92%。这一结果说明，在成功获取聚类中心的基础上，高斯隶属度函数能够有效地对样本进行准确的健康状态评估，只有极少数样本被错误分类。这可能是由于所选用的光伏发电系统规模较小，导致不同类型的故障在聚类中心上的分布较为接近。在进行隶属度计算的过程中，这种接近性增加了误判的可能性。此外，由于光伏阵列中正常工作状态、局部阴影状态和老化状态的影响在表现形式上具有相似性，使得故障特征不够明显，进一步增加了分类错误的机率。

4.5 本章小结

本章主要介绍了模糊 C 均值如何诊断出光伏阵列的故障类型。模糊 C 均值算法存在数据特征难以提取，非凸形数据集中难以达到良好的聚类效果，高维数据处理困难等问题。因此，对特征数据进行归一化处理，消除了输出数据随接收日照强度和组件的工作温度的波动而呈现的非线性变化；并在灰色关联分析划分出光伏阵列健康等级的基础上，对亚健康或异常等级下的光伏阵列进行聚类，减少了对特征参数数据的需求量；同时，引入高斯隶属度函数，进一步扩大了样本特征，减少了计算复杂程度，并且诊断出光伏阵列的故障类型，同时也给出诊断结果的正确率。

第 5 章 基于概率神经网络的光伏阵列故障诊断

为检验基于灰色关联分析与 FCM 算法对光伏阵列健康状态评估的性能，利用第 2 章已有的不同状态下的光伏阵列仿真模型，设计基于概率神经网络的光伏阵列故障诊断算法。

5.1 概率神经网络原理

概率神经网络的结构简单易理解，便于算法设计，利用线性学习方法实现非线性学习，在模式分类领域有广泛应用。在 MATLAB 中，newpnn 函数能够便捷构建概率神经网络^[57]。PNN 属于前馈神经网络，源于径向基函数，基于贝叶斯决策理论和概率密度函数估计两大理论。广泛用于识别分类问题，通过线性训练实现非线性分类目标。PNN 融合了贝叶斯最优决策理论下的概率密度估计和径向基网络。在一定条件下，其分类决策面与贝叶斯最优决策面相符。概率神经网络的优点包括：

(1)结构简单，训练速度快，节省训练时间。概率神经网络不需要频繁的调整网络参数，有很好的泛化能力，对解决非线性问题非常有效。

(2)通过隐含层的数量就是训练样本的数量，而且隐含层的权重是固定的，这样不会让概率神经网络出现局部最小值问题。

(3)适用范围广泛，可根据实际情况轻松扩展或缩减网络结构，无需重新训练即可添加额外样本。

5.1.1 贝叶斯决策理论和概率密度函数

(1)贝叶斯决策理论

贝叶斯决策理论公式提供了一种在给定相关条件的的情况下，更新对某个假设概率估计的方法^[65]。贝叶斯决策理论公式如式(5-1)所示：

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (5-1)$$

式中：P(A|B)为后验概率，是在事件 B 已经发生的情况下，我们对事件 A 发

生概率的重新评估； $P(A)$ 为先验概率，是在未考虑事件 B 发生之前，仅基于现有信息对事件 A 发生概率的估计； $P(B|A)/P(B)$ 为可能性函数，是一个调整因子，它基于先验概率，通过考虑事件 B 的发生来修正我们对事件 A 发生概率的估计，使其更接近实际情况。

(2)概率密度函数

概率密度函数是描述连续的随机变量在某一特定区域内的概率分布^[66]。任意一个满足式(5-2)和(5-3)条件的函数 $f_x(x)$ 都可以被定义为其概率密度函数。

$$\int_{-\infty}^{\infty} f_x(x) dx = 1 \quad (5-2)$$

$$f_x(x) \geq 0, -\infty < x < \infty \quad (5-3)$$

随机变量 X 在区间上的概率可以由其概率密度函数的定积分表示，如式(5-4)与(5-5)所示：

$$P(a < X < b) = \int_a^b f_x(x) dx \quad (5-4)$$

$$F(x) = P(X < x) = \int_{-\infty}^x f_x(\varphi) d\varphi \quad (5-5)$$

式中： $F(x)$ 是 X 的累积分布函数，

5.1.2 概率神经网络模型结构

概率神经网络模型结构如图 5-1 所示。概率神经网络模型通常由四个层次组成，依次是输入层、隐含层、求和层、输出层^[67]。

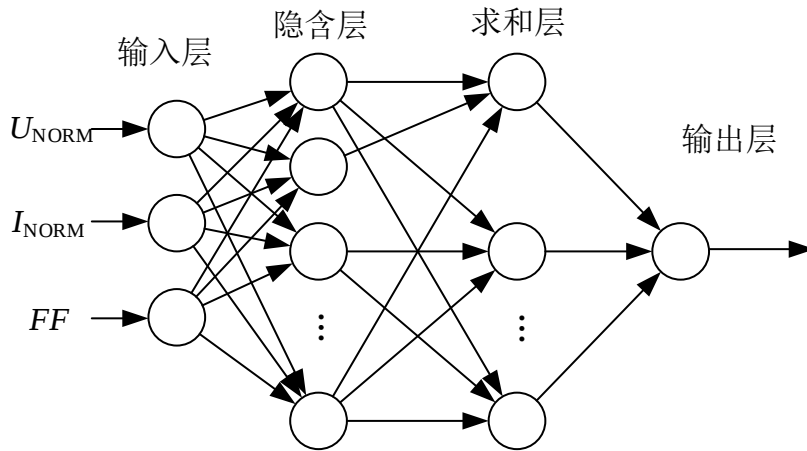


图 5-1 概率神经网络模型

Fig5-1 Probabilistic neural network models

第一层为输入层，在输入层，每个输入节点负责接收一个单独的输入值，并将其转换为一个输出值。输入层的主要作用是确定测试样本与训练样本之间的差异，这些差异实际上表示了它们在向量空间中的相对位置。这个位置差异的大小揭示了测试样本与训练样本之间的相似度。计算出的这些差异值会被送往网络的下一层，即隐含层。

第二层为隐含层，在神经网络结构中隐含层可以有一个或多个，权重 W_i 负责将输入层与隐含层相连接。隐含层包含的节点数量与训练样本的数量一一对应。隐含层使用输出函数 $g(Z_i)$ 替代了传统的 S 型激活函数。 $g(Z_i)$ 函数如式(5-6)所示：

$$g(Z_i) = \exp\left[\frac{(Z_i - 1)}{s^2}\right] \quad (5-6)$$

式中： Z_i 为第 i 个神经元的输入， s 为均方差。

每个隐含层的神经元生成输入模式向量 X 与权值向量 W_i 的标量积，标量积采用如式(5-7)所示：

$$Z_i = X \times W_i \quad (5-7)$$

再对 Z_i 进行非线性运算，当 X 和 W 均规格化为单位长度时，其输出函数如式(5-8)所示：

$$f(X, W_i) = \exp\left[\frac{-(X - W_i)^T (X - W_i)}{2\sigma^2}\right] \quad (5-8)$$

式中： W_i 为输入层到隐含层的权值， σ 为平滑因子，根据具体问题不断调整 σ 的取值，一般取 σ 为 0.1。

平滑因子的取值在概率神经网络中起决定性作用，影响着概率神经网络的分类决策。如果平滑因子的取值过小会导致概率神经网络过度训练；如果平滑因子的取值过大会导致概率神经网络更加平滑，有助于消除异常数据，但也可能降低训练的正确性。

第三层为求和层，也可称为累加层，其神经元仅与同一类模式层中的单元相连。在概率神经网络的结构里，第一层输入层与第二层隐含层负责输入数据以及对数据与类别进行关联与整合，并且能得到每个类别的概率，求和层的输出与贝叶斯理论推导出的最优估计结果相吻合，第三层的核心功能是整合第二层隐含层传递的数据，并将其汇总后传递给第四层，以便进行最终的输出。

第四层为输出层，该层主要用于从多个类别模式的概率密度函数估计中选出具有最高概率的神经元作为最终输出。所以，认为测试样本最有可能属于具有最高概率的神经元代表的类别，会输出一个 1；而所有其他代表不同类别的神经元则输出 0。

这四层网络确保了概率神经网络能够从所有可能的类别中选择一个最可能的类别作为样本的分类结果。因此，采用概率神经网络能够高效地识别光伏阵列中存在的故障类型，并对每一种故障进行精确的区分和归类。

5.2 基于概率神经网络的光伏阵列故障诊断模型

5.2.1 基于概率神经网络的光伏阵列故障诊断流程

概率神经网络故障诊断流程如图 5-2 所示。

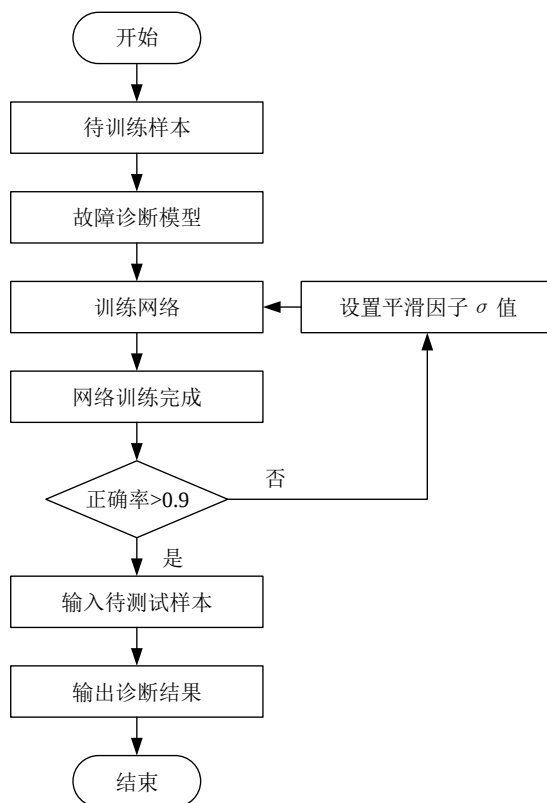


图 5-2 概率神经网络故障诊断模型

Fig5-2 Probabilistic neural network fault diagnosis model

概率神经网络故障诊断的步骤如下：

(1)模拟仿真。通过模拟仿真技术对光伏阵列进行测试，以模拟可能出现的各

种故障情况，并记录下这些故障的数据。这些数据将用于创建训练和测试样本。

(2)数据归一化。对收集到的训练样本和测试样本数据进行归一化处理，以确保数据在相同的范围内，便于神经网络的处理。

(3)建立诊断模型。建立一个针对光伏阵列故障检测的概率神经网络模型，并明确模型的核心参数设置。

(4)模型训练。将归一化后的训练样本数据输入到概率神经网络模型中进行训练。训练过程中，如果模型的准确率超过 90%，则认为训练成功，可以输出故障类型的诊断结果。如果准确率低于 90%，则需要调整模型中的平滑因子，以提高训练的准确率，直到达到满意的准确度。

(5)模型测试。将归一化处理后的测试样本数据代入训练好的模型中进行诊断，以验证模型的诊断能力，并得到最终的故障诊断结果。

5.2.2 基于概率神经网络的光伏阵列故障诊断

基于图 2-5 的光伏阵列仿真模型对第 4 章描述的 11 种光伏阵列状态类型进行仿真并提取最大功率点数据，再利用式(4-6)-(4-8)对提取出的最大功率点数据进行归一化处理，对于得到的数据用于训练与诊断。本文采集光伏阵列各种状态的数据共 550 组。其中正常状态有 50 组数据，局部阴影与老化状态各 150 组数据，其中局部阴影状态的 F2-F4 每种各 50 组数据，老化状态的 F5-F7 每种各 50 组数据，以上六种状态每种状态提取 40 组数据用于训练，剩余 10 组数据用于诊断；短路与开路状态各 100 组数据，其中短路状态的 F8-F9 每种各 50 组数据，开路状态的 F10-F11 每种各 50 组数据，以上四种状态每种状态提取 40 组数据用于训练，剩余 10 组数据用于诊断。

利用 newpnn 函数建立概率神经网络，其中 newpnn 函数使用格式如(5-9)所示：

$$\text{newpnn} = (P, T, \sigma) \quad (5-9)$$

式中： P 代表待训练样本向量， T 代表待诊断样本向量， σ 为概率神经网络的平滑因子。

σ 的值域是从 0 延伸到正无穷大。随着 σ 值趋向于 0，概率神经网络的分类效果类似于最近邻分类器，主要依据与待分类样本距离最近的样本来进行判断。而

随着 σ 值的增加,网络在做出分类决策时会考虑更多邻近的设计向量。确定 σ 为 1 后,对训练样本和待诊断样本进行规范化处理,确保数据在同一标准下进行诊断。接着,将处理后的训练样本送入已经搭建的 PNN 模型中。完成模型的训练阶段后,把待诊断样本送入已经训练好的 PNN 模型中,实现对光伏阵列故障的准确诊断。序号为 F1 代表光伏阵列处于正常状态,序号为 F2-F4 代表光伏阵列处于局部阴影状态,序号为 F5-F7 代表光伏阵列处于老化状态,序号为 F8-F9 代表光伏阵列处于短路状态,序号为 F10-F11 代表光伏阵列处于开路状态。概率神经网络训练结果如图 5-3 所示,诊断结果如图 5-4 所示。

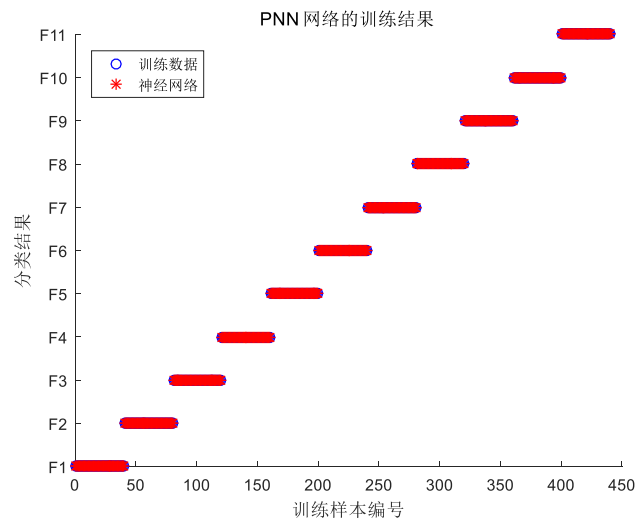


图 5-3 概率神经网络训练结果

Fig5-3 Probabilistic neural network training results

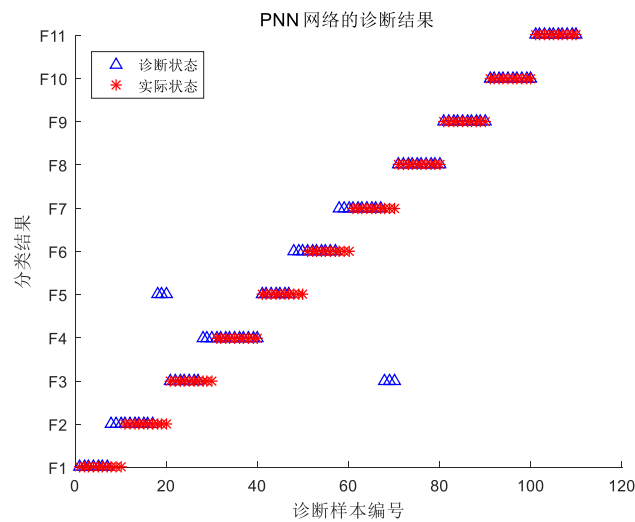


图 5-4 概率神经网络诊断结果

Fig5-4 Probabilistic neural network diagnostic results

由图 5-3 可知，训练数据与神经网络数据全部重合，所以局部阴影状态，老化状态，短路与开路状态的 11 种工作模式训练结果全部正确。由图 5-4 可知，PNN 故障诊断模型对于大部分故障都能实现有效辨识，错判主要集中于 F1-F7，其中 F1 为正常状态，F2-F4 为局部阴影状态，F5-F7 为老化状态，而对于 F8-F9 的短路状态与 F10-F11 的开路状态则能够 100%诊断正确。

对测试数据进行概率神经网络诊断，得到的精确度如表 5-1 所示。

表 5-1 概率神经网络诊断精确度分析

Table5-1 Probabilistic neural network diagnostic accuracy analysis			
工作状态	测试样本	正确诊断	诊断准确率
F1	10	8	80%
F2,F3,F4	30	25	83.33%
F5,F6,F7	30	26	86.67%
F8,F9,	20	20	100%
F10,F11	20	20	100%

由表 5-1 可知，PNN 诊断模型的整体辨识精度为 90%，对于正常状态，局部阴影状态和老化状态下的光伏阵列，诊断精度分别为 80%，83.33%和 86.67%，对这些状态下的光伏阵列进行故障诊断很容易造成误判。主要的原因在于这三种状态的光伏阵列对于 3×3 光伏阵列的影响非常微弱，而且它们对光伏阵列的影响十分相似，故障特征不明显，从而导致误判的发生。然而在短路与开路故障状态下，PNN 模型的诊断准确率达到 100%。综上所述，PNN 光伏阵列诊断模型能够有效地诊断和识别光伏阵列的故障类型。但是与本文所提灰色关联分析与 FCM 的算法相比较，正确率并无本文所提算法高，而且需要对所有 11 种光伏阵列的状态进行训练，增加了提取数据的数量以及计算的复杂度。

5.3 本章小结

本章根据第 2 章的光伏阵列仿真模型，并通过仿真获取光伏阵列不同状态下的特征参数数据集，建立概率神经网络光伏阵列故障诊断模型，并应用所得数据对不同状态的光伏阵列进行诊断。结果表明，概率神经网络光伏阵列故障诊断模型对短路与开路故障的光伏阵列进行诊断能够实现 100%正确，而正常状态、局部

阴影状态和老化状态的光伏阵列故障诊断有一定误差,总体上该路神经网络诊断模型的整体辨识精度为 90%。同时,与灰色关联分析与 FCM 的健康状态评估进行比较验证,证明了利用灰色关联分析与 FCM 的光伏阵列健康状态评估的正确性以及简便性。

第 6 章 总结与展望

6.1 研究内容总结

随着国家提出的“双碳”目标，人们日益关注可再生能源和清洁能源的发展。在世界能源结构的转变中，这些能源类型正发挥着越来越显著的作用，有助于实现能源使用的可持续性以及环境的长期健康。在此背景下，光伏发电作为高效清洁能源，日益成为电力主要来源之一，而光伏阵列大多处于环境复杂多变的情况下，容易出现各类状况。本文针对光伏阵列日常发生的各类故障诊断方法查阅了大量相关文献，对各类诊断方法的问题讨论与各类诊断技术进行了研究总结。根据现阶段各类故障诊断方法特点，本文设计了基于灰色关联分析于 FCM 的光伏阵列健康状态评估方法。本文主要研究内容如下：

(1) 查阅大量的国内外有关光伏发电的相关文献资料，分析在当今能源发展的背景下，光伏发电技术的现实意义以及未来的潜力，概述了光伏阵列常见的故障种类和相应的诊断技术。基于光伏电池的发电机制，构建了光伏电池的电路模型，并进一步搭建了光伏阵列的仿真模型。通过仿真分析，得到正常状态下和故障状态下的输出特性曲线以及特征参数。

(2) 借助 MATLAB/SIMULINK 构建了光伏阵列的仿真模型，模拟了光伏阵列在局部阴影、老化、短路和开路状态下的情况，研究了不同状态下光伏阵列的输出特性和特征参数，确立了光伏阵列在不同状态下的 I-V 特性曲线有明显差异。因此，可根据光伏阵列输出的 I-V 特性曲线对其进行健康等级划分。利用灰色关联分析算法，对光伏阵列在故障状态下输出的 I-V 特性曲线与正常状态下的 I-V 特性曲线进行灰色关联分析，得到灰色关联度，再将其代入隶属度函数与健康指数公式，从而计算出健康指数。通过计算，能够准确划分出光伏阵列的健康等级。

(3) 利用模糊 C 均值算法对光伏阵列的故障进行诊断，需要不同状态下光伏阵列的特征参数，但需要的数据量大且计算复杂。因此，在灰色关联分析划分出光伏阵列健康等级的基础上，并对特征数据进行归一化处理，消除了输出数据随接受日照强度和组件工作温度的波动而呈现的非线性变化。仅需对亚健康或异常等级下的光伏阵列进行聚类，减少了对特征参数的需求量，并且得到聚类光伏阵

列的聚类中心；再引入高斯隶属度函数，将聚类中心与测试数据带入隶属度函数，则诊断出光伏阵列的工作状态，且诊断结果的正确率达到 92%。

(4)为验证灰色关联分析与 FCM 融合的健康状态评估的正确性，根据光伏阵列仿真模型与已获取的不同状态下的特征参数的数据集，并对所有数据进行归一化处理，建立概率神经网络光伏阵列故障诊断模型，并应用所得数据对不同状态下的光伏阵列进行诊断。概率神经网络光伏阵列故障诊断模型对故障等级下的光伏阵列能实现 100%正确，而局部阴影，老化状态的光伏阵列诊断有一定误差。再与灰色关联分析与 FCM 的健康状态评估进行比较，证明了灰色关联分析与 FCM 融合的健康状态评估方法的正确性与简便性。

6.2 研究工作展望

本文利用灰色关联分析与 FCM 融合对光伏阵列健康状态进行评估，虽然能够实现对光伏阵列进行健康等级划分以及故障诊断，且减少了特征参数的需求量，但是本文还有许多不足之处仍需要做进一步改进。

(1)本文中的不同工作状态光伏阵列的仿真模型都是在 MATLAB/SIMULINK 的仿真环境下完成的，关于光伏阵列的健康状态评估目前只限于理论研究，下一步应延伸到实际生活中，对现实的光伏阵列进行健康状态评估，使之具有实际意义。

(2)实际运行的光伏阵列，其受到太阳的日照的强度、光伏组件的工作温度和光伏阵列的使用时间等都对光伏阵列的输出特性有着显著的影响。本文的研究对这些变量的处理还过于简单，未来的研究应当更加深入地探讨这些因素如何具体影响光伏阵列。

(3)在本文中，光伏阵列的不同状态被单独拿出来进行分析与诊断，而没有考虑光伏阵列可能同时存在多种状态的情况，未来研究方向应该将同时存在多种状态时，对光伏阵列进行健康状态评估作为重点。

参考文献

- [1] 辛保安,单葆国,李琼慧等.“双碳”目标下“能源三要素”再思考[J].中国电机工程学报,2022,42(09):3117-3126.
- [2] 陈国平,董昱,梁志峰.能源转型中的中国特色新能源高质量发展分析与思考[J].中国电机工程学报,2020,40(17):5493-5505.
- [3] 国家能源局.关于2019年风电、光伏发电项目建设有关事项的通知[EB/OL].[2019]49号.
https://www.gov.cn/zhengce/zhengceku/2019-09/29/content_5434799.htm
- [4] 康克佳.国家能源局发布1-11月份全国电力工业统计数据[N].中国城市报,2023-12-25(012).
- [5] Yan J,Yang Y,Campana E P,et al.City-level analysis of subsidy-free solar photovoltaic electricity price, profits and grid parity in China[J].Nature Energy,2019,4(8):709-717.
- [6] Aleksandra F K,Alessandra S,Mirjana D,et al.Trends in the integration of photovoltaic facilities into the built environment[J].Open House International,2020,45(1-2):195-207.
- [7] Dan C,Jia N Y,Feng X,et al.A market-based framework for CO2 emissions reduction in China's civil aviation industry[J].Transport Policy,2023,143:150-158.
- [8] 李继峰,郭焦锋,高世楫等.我国实现2060年前碳中和目标的路径分析[J].2021,38(04):37-47.
- [9] Alam,Khorshed M,Khan,et al.A comprehensive review of catastrophic faults in pv arrays:types,detection,and mitigation techniques[J].IEEE Journal of Photovoltaics,2015,5(3):982-997.
- [10] 周向阳.基于图像处理的光伏阵列故障检测研究[D].贵阳:贵州大学,2021.
- [11] 张珍奇,王文革.光伏发电并网的探讨[J].光源与照明,2022(4):222-224.
- [12] 李文才,彭程.分布式光伏电源并入配电网的故障分析及保护研究[J].机电工程技术,2019,48(06):186-188.
- [13] 谌守禄,吴文宝,裘名华等.基于航拍图像的无人机光伏组件巡检策略研究及运

- 用[J].江西电力,2019,43(04):18-20.
- [14] 闫天一.太阳能电站光伏组件故障监测系统的研发[D].镇江:江苏大学,2019
- [15] 顾崇寅,徐潇源,王梦圆等.基于 CatBoost 算法的光伏阵列故障诊断方法[J].电力系统自动化,2023,47(02):105-114.
- [16] 秦本乾.光伏阵列故障检测与定位研究[D].昆明:云南师范大学,2021.
- [17] 李元良,丁坤,陈富东,等.基于快速过采样主成分分析法的光伏阵列故障诊断[J].电网技术,2019,43(1):308-315.
- [18] 王元章,李智华,吴春华.一种四参数的光伏组件在线故障诊断方法[J].中国电机工程学报,2014,34(13):2078-2087.
- [19] 王瀚笠,郭珂,袁丹夫等.光伏阵列简化模型与故障分析[J].通信电源技术,2019,36(6):4-10.
- [20] 叶进,董美辰,何华光等.基于梯度下降和差分进化的光伏阵列 MPPT 方法[J].控制工程,2019,26(09):1695-1702.
- [21] Wang W,Alex C F,Zhang J,et al.Fault diagnosis of photovoltaic panels using dynamic current-voltage characteristics[J].IEEE Transactions on Power Electronics,2016,31(2):1588-1599.
- [22] Kumar B K,Kumar A P.Detection and classification of faults in solar PV array using thevenin equivalent resistance[J].IEEE Journal of Photovoltaics,2020,10(2):644-654.
- [23] Pei T,Hao X.A fault detection method for photovoltaic systems based on voltage and current observation and evaluation[J].Energies,2019,12(9):75-80
- [24] 王元章,李智华,吴春华.光伏系统故障诊断方法综述[J].电源技术,2013,37(09):1700-1705.
- [25] Roberto P,Marina P,Andrea F,et alAutomatic faults detection of photovoltaic farms:solair, a deep learning based system for thermal images[J].Energies,2020,13(24):6496-6496.
- [26] 毛峡,李亚豪.光伏阵列故障检测中的无人机红外图像拼接[J].太阳能学报,2020,41(3):262-269.
- [27] 蒋琳,苏建徽,李欣等.基于可见光和红外热图像融合的光伏阵列热斑检测方法

- [J].太阳能学报,2022,43(01):393-397.
- [28] 赵雪,杨晨,夏晨蹊等.基于热像仪的太阳能电池故障诊断[J].贵州大学学报(自然科学版),2019,36(01):60-65.
- [29] 皮宇啸.基于可见光和红外图像的光伏阵列缺陷检测[D].北京:华北电力大学,2023.
- [30] Alsafasfeh M,Abdel-Qader I,Bazuin B,et al.Unsupervised fault detection and analysis for large photovoltaic systems using drones and machine vision[J].Energies,2018,11(9):105-117
- [31] Feng H,Jie S,Hang M,et al.A novel framework on intelligent detection for module defects of PV plant combining the visible and infrared images[J].Solar Energy,2022,236:406-416.
- [32] V.S.B.Kurukuru,A.Haque,M.A.Khan.Fault classification for photovoltaic modules using thermography and image processing[C].2019 IEEE Industry Applications Society Annual Meeting.IEEE,2019:1-6.
- [33] 陈克松.基于深度学习的光伏组件故障诊断方法[D].兰州:兰州理工大学,2022.
- [34] 汪诚,彭合,华红婷等.基于有源耦合单元的级联光伏系统直流电压传感器故障穿越策略[J].中国电机工程学报,2023,43(24):9594-9608.
- [35] 张晓娜,高德东,刘海雄,等.一种新型光伏阵列多传感器故障检测定位方法[J].可再生能源,2016,34(2):166-172.
- [36] 吴延霞,李田泽,乔家振等.光伏阵列故障定位中传感器安放位置分析[J].通信电源技术,2017,34(5):50-51.
- [37] Hai W,Feng J S.Optimal sensor placement and fault diagnosis model of PV array of photovoltaic power stations based on XGBoost[J].IOP Conference Series: Earth and Environmental Science,2021,661(1):12-25.
- [38] Pei T,Zhang J,Li L,et al.A fault locating method for PV arrays based on improved voltage sensor placement[J].Solar Energy,2020,201:279-297.
- [39] Hubert S,Karolina E S.Applications of artificial intelligence algorithms in the energy sector[J].Energies,2022,16(1):347-347.
- [40] Lina A,M.A A,Ibrahim A.Solar PV power forecasting at Yarmouk University using machine learning techniques[J].Open Engineering,2022,12(1):1078-1088.

- [41] 武文栋,施保华,郑传良等.基于改进麻雀搜索算法优化 RBF 神经网络的光伏阵列故障诊断[J].智慧电力,2023,51(02):77-83.
- [42] 张杰,易辉,张霞等.基于布谷鸟算法的光伏组件故障诊断模型优化[J].电源技术,2020,44(01):76-79.
- [43] 毕锐,丁明,徐志成,等.基于模糊 C 均值聚类的光伏阵列故障诊断方法[J].太阳能学报,2016,37(3):730-736.
- [44] 刘圣洋,冬雷,王晓晓等.基于高斯核模糊 C 均值聚类的光伏阵列故障诊断方法[J].太阳能学报,2021,42(05):286-294.
- [45] 吴春华,俞薛颖,李智华等.基于 FCM 与高斯隶属度的光伏组件健康状态诊断[J].电网技术,2022,46(05):1887-1896.
- [46] 段震清,孙建民,梁凌等.基于 XGBoost 算法的光伏阵列故障诊断方法研究[J].太阳能,2023(01):62-68.
- [47] Garoudja E,Chouder A,Kara K,et al.An enhanced machine learning based approach for failures detection and diagnosis of PV systems[J].Energy Conversion and Management,2017,151:496-513.
- [48] Zeng X H,Peng P C,Chen X L,et al.Compound fault diagnosis for photovoltaic arrays based on multi-label learning considering multiple faults coupling[J].Energy Conversion and Management,2023,279-280.
- [49] Dhritiman A,Soumesh C,Kumar A C.Stacking ensemble based fault diagnosis approach for improved operation of photovoltaic arrays[J].Energy Sources,Part A:Recovery,Utilization,and Environmental Effects,2022,44(2):5421-5439.
- [50] 黄建华,向钠,齐锴亮.太阳能光伏理化基础[M].北京,化学工业出版社,2017.
- [51] 易威.光伏阵列的参数辨识及故障诊断研究[D].杭州:浙江大学,2021.
- [52] 程陈,陈堂贤,孙培胜等.基于灰狼算法的光伏组件故障诊断模型优化[J].电源技术,2022,46(06):684-687.
- [53] 陈泽理,卢箫扬,林培杰等.基于伪标签-1D DenseNet-KNN 的光伏阵列开集复合故障诊断方法[J].福州大学学报(自然科学版),2023,51(04):490-497.
- [54] Siddiqui M,Arif A,Bilton A,et al.An improved electric circuit model for photovoltaic modules based on sensitivity analysis[J].Solar Energy,2013,90:29-42.
- [55] Raja O,Sheikh I J.Estimation of irradiance and temperature of solar photovoltaic

- system operating at maximum power point[J].International Journal of Ambient Energy,2023,44(1):124-130.
- [56] 张商州,刘宝盈,徐晓龙.光伏阵列数学模型的建立与分析[J].自动化与仪器仪表,2016,(07):192-194.
- [57] 宋文海.基于概率神经网络的光伏阵列故障检测研究[D].淄博:山东理工大学,2020.
- [58] Lun S,Du C,Sang J,et al.An improved explicit I-V model of a solar cellbased on symbolic function and manufacturer's datasheet[J].Solar Energy,2014,110:603-614.
- [59] 吴云,雷建文,鲍丽山,等.基于改进灰色关联分析与蝙蝠优化神经网络的短期负荷预测[J].电力系统自动化,2018,42(20):67-74.
- [60] 郭婷婷,武永鑫.基于仿真模型的光伏组串故障分析方法[J].天津科技大学学报,2019,34(06):71-75.
- [61] 王海岚.基于密度 Canopy 的高维模糊 C 均值聚类推荐算法[D].桂林:桂林理工大学,2023.
- [62] Mahela P O,Shaik G A.Recognition of power quality disturbances using S-transform based ruled decision tree and fuzzy C-means clustering classifiers[J].Applied Soft Computing,2017,59:243-257.
- [63] 陆灵骅,朱红路,连魏魏等.模糊 C 均值聚类在光伏阵列故障样本数据识别中的应用[J].发电技术,2018,39(02):188-194.
- [64] Lin P,Lin Y,Chen Z,et al.A density peak based clustering approach for fault diagnosis of photovoltaic arrays [J].International Journal of Photoenergy,2017(9):1-14.
- [65] 杨静,陈冬,程小红.贝叶斯公式的几个应用[J].大学数学,2011,27(02):166-169.
- [66] Zhu H,Lu L,Yao J,et al.Fault diagnosis approach for photovoltaic arrays based on unsupervised sample clustering and probabilistic neural network model[J].Solar Energy,2018,176:395-405.
- [67] 史忠植.神经网络[M].北京,高等教育出版社,2009.

攻读学位期间发表的学术论文目录

- [1] Shinong Wang, Chuanzhi Wang, Yuan Ge, Shilin Liu, Jie Xu, Ragab Ahmed Amer. In-depth analysis of photovoltaic module parameter estimation[J]. Energy.
- [2] 王传志,吴宏伟,汪石农.灰色关联分析与 FCM 融合的光伏阵列状态评估[J].重庆工商大学学报(自然科学版).

致谢

在这篇论文即将画上句号之际，我心中涌动着无尽的感激与思绪。这不仅是对我研究工作的总结，更是对那些在我学术旅程中给予我支持、鼓励和启发的人的深深感谢。在此，我想用这篇致谢来记录下这段旅程中的点点滴滴，以及那些不可磨灭的记忆。

首先，我衷心感谢我的导师汪石农副教授，他的专业知识、深邃见解和悉心指导是我完成这篇论文的坚强后盾。在研究的每一个阶段，汪石农副教授都以其卓越的学术洞察力和严谨的治学态度，为我提供了宝贵的指导和建议。在论文的选题、研究和撰写过程中汪石农副教授的细致审阅和建设性建议，使我能够不断改进和完善我的研究。每一次讨论和建议都让我受益匪浅，他的严谨态度和高标准要求也激励我追求更高的学术成就。在此，我向汪石农副教授表达我最深切的敬意和感激，感谢他在这段学术旅程中的陪伴和指引。

其次，我还要感谢一教404的所有同学们，他们的专业知识、热情讨论和无私分享为我的研究提供了宝贵的资源。感谢刘俊杰师兄、万仕帅师兄、汪永旺师兄、杨海挺师弟、卓文涛师妹、吴宏伟师弟、卢兴凯师弟、王正师弟、李祥祥师弟以及同门陈德政、顾黄一在研究的每一个阶段都给予了我无私的帮助和宝贵的意见。我们无数次的讨论和交流，不仅让我在学术上受益匪浅，也让我在思考问题时更加全面和深刻。

此外，我衷心感谢我的父母，感谢他们无条件的爱与支持。他们不仅在物质上给予我支持，更在精神上给予我无尽的鼓励。在我面对挑战和困难时，父亲的智慧和坚韧，母亲的关怀和鼓励，一直是我前进的动力。他们的无私奉献塑造了我，让我有信心追求卓越。这篇论文的每一页都凝聚着他们的爱与期望，没有他们，我的成就不可能实现。感谢他们无尽的爱，我将永远珍惜。

最后，我要感谢所有给予我帮助和支持的人。是你们让我的研究之路不再孤单，是你们让我的学术之旅充满了意义和价值。这篇论文的完成，是我们共同努力的结果，是对我们所有人辛勤工作和不懈追求的最好证明。

在此，我将这篇论文献给所有支持我、鼓励我、启发我的人。感谢你们，没有你们，就没有今天的我。

尚德敏学 唯实惟新

