

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 2 月 13 日 (13.02.2020)



(10) 国际公布号
WO 2020/029344 A1

- (51) 国际专利分类号:
H02S 50/10 (2014.01)
- (21) 国际申请号: PCT/CN2018/102907
- (22) 国际申请日: 2018 年 8 月 29 日 (29.08.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201810907589.1 2018年8月10日 (10.08.2018) CN
- (71) 申请人: 江南大学(JIANGNAN UNIVERSITY) [CN/CN]; 中国江苏省无锡市滨湖区蠡湖大道1800号物联网工程学院, Jiangsu 214122 (CN)。
- (72) 发明人: 周超超(ZHOU, Chaochao); 中国江苏省无锡市滨湖区蠡湖大道1800号物联网工程学院,

Jiangsu 214122 (CN)。陶洪峰(TAO, Hongfeng); 中国江苏省无锡市滨湖区蠡湖大道1800号物联网工程学院, Jiangsu 214122 (CN)。沈建强(SHEN, Jianqiang); 中国江苏省无锡市滨湖区蠡湖大道1800号物联网工程学院, Jiangsu 214122 (CN)。魏强(WEI, Qiang); 中国江苏省无锡市滨湖区蠡湖大道1800号物联网工程学院, Jiangsu 214122 (CN)。刘巍(LIU, Wei); 中国江苏省无锡市滨湖区蠡湖大道1800号物联网工程学院, Jiangsu 214122 (CN)。周龙辉(ZHOU, Longhui); 中国江苏省无锡市滨湖区蠡湖大道1800号物联网工程学院, Jiangsu 214122 (CN)。

(74) 代理人: 无锡华源专利商标事务所(普通合伙)(WUXI HUAYUAN PATENT AND TRADEMARK AGENCY (GENERAL PARTNERSHIP)); 中国江苏省无锡市梁溪区

(54) Title: RANDOM FOREST ALGORITHM-BASED PHOTOVOLTAIC ARRAY FAULT DIAGNOSIS METHOD

(54) 发明名称: 一种基于随机森林算法的光伏阵列故障诊断方法

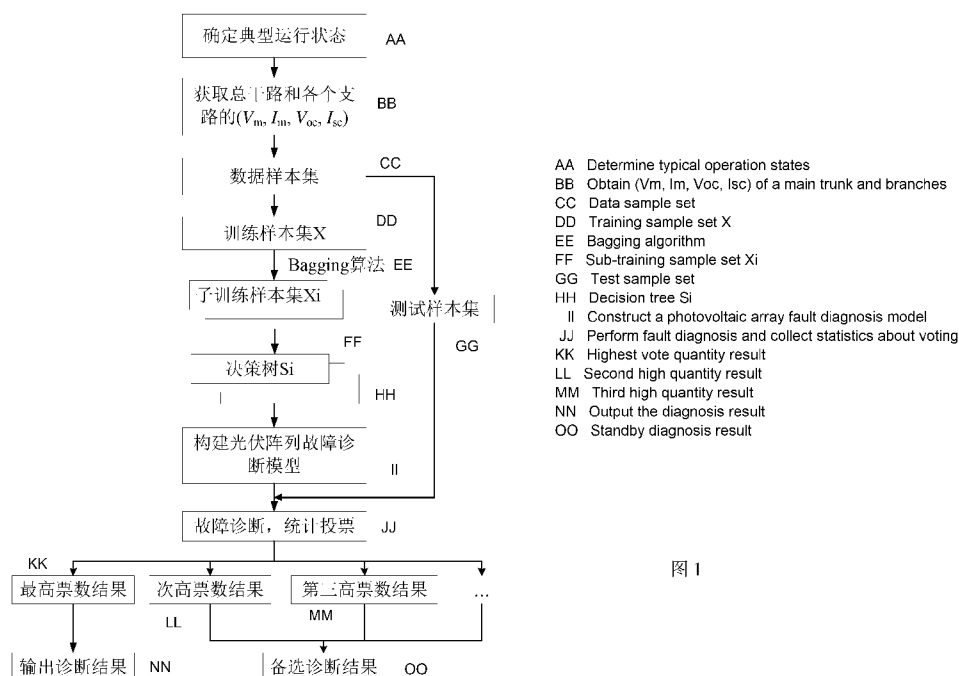


图 1

(57) Abstract: The present invention relates to the technical field of photovoltaics. Disclosed is a random forest algorithm-based photovoltaic array fault diagnosis method. The method comprises: acquiring circuit parameter groups corresponding to branches and a main trunk when a photovoltaic array is in typical operation states; constructing a fault feature vector according to the acquired circuit parameter groups to construct a data sample set; constructing a photovoltaic array fault diagnosis model on the basis of a random forest algorithm by using the data sample set; performing diagnosis on a photovoltaic array to be diagnosed using the model to obtain voting

清扬路 228 号地铁大厦 1108 室（联系人：
聂启新），Jiangsu 214023 (CN)。

- (81) 指定国(除另有指明，要求每一种可提供的国家保护)：AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

根据细则 4.17 的声明：

- 关于申请人有权申请并被授予专利(细则 4.17(ii))

本国际公布：

- 包括国际检索报告(条约第 21 条(3))。

results corresponding to the typical operation states; and obtaining a fault diagnosis result of the photovoltaic array to be diagnosed according to the voting results. The method constructs a photovoltaic array fault diagnosis model by using a random forest algorithm on the basis of a data drive idea, is adapted to actual characteristics of a photovoltaic array, overcomes the defect that a conventional neural network algorithm requires a large data volume, a long training time, and the like, and can easily and quickly complete a diagnosis task.

(57) 摘要：本发明公开了一种基于随机森林算法的光伏阵列故障诊断方法，涉及光伏技术领域，该方法包括：采集光伏阵列处于各个典型运行状态时，各个支路和总干路对应的电路参数组，根据采集到的电路参数组构建得到故障特征向量，从而构建数据样本集，利用数据样本集基于随机森林算法构建得到光伏阵列故障诊断模型，利用该模型对待诊断光伏阵列进行诊断，得到各个典型运行状态对应的投票结果，根据各个投票结果得到待诊断光伏阵列的故障诊断结果，该方法基于数据驱动的思想，利用随机森林算法构建光伏阵列故障诊断模型，适用于实际光伏阵列的特点，克服了传统神经网络算法需要数据量大、训练时间长等问题，能够简单快速地完成诊断任务。

一种基于随机森林算法的光伏阵列故障诊断方法

技术领域

5 本发明涉及光伏技术领域，尤其是一种基于随机森林算法的光伏阵列故障诊断方法。

背景技术

随着能源需求的增长、化石能源的日益枯竭与成本上涨，以及全球气候变暖等诸多因素的影响，可再生能源技术得到了飞速发展，其中，太阳能具有易
10 获取、无噪声、清洁、无穷无尽等优点，成为了可再生能源的重要部分。目前利用太阳能发电主要有两种形式：光热发电和光伏发电。光热发电与火力发电类似，但热能主要来自大规模的镜面收集，将水加热后，利用蒸汽推动传统发电机工作，从而实现发电，这种发电方式虽然规避了硅晶的光电转换，但对光照强度的要求高且发电成本偏高。而光伏发电基于光生伏特效应，直接将太阳
15 能转化为电能，建设周期短，在运行时不会产生废渣、废水和其他污染物，尤其是对于交通不发达的山区、海岛和偏远地区，光伏发电具有更重要的价值。

一般情况下，光伏发电都是利用光伏电池组件串、并联组成光伏阵列进行运行的，而光伏阵列使用的光伏电池组件较多，光伏电池组件本身发生故障的概率就较高，并且光伏阵列长期运行在户外，运行环境恶劣，易出现老化和损坏等情
20 况，从而造成光伏电池组件发电效率的降低甚至停止工作。当光伏阵列中某个光伏电池组件发生故障后，会导致系统效率的下降，并对电力系统的运行调度造成不利的影响，严重时，甚至造成财产损失和人员伤亡，因此，对光伏阵列进行故障诊断具有重要意义，目前所采用的故障诊断方法主要包括直接法和间接法两类，间接法比较典型的有红外热量检测法和发电功率对比法，直接
25 法比较典型的有对地电容法、时域反射法、智能诊断算法和电特性检测法。其中，通过将智能诊断算法和电特性检测法进行结合进行故障诊断，是目前非常具有潜力的方法。目前的智能诊断算法多采用以 BP(Back Propagation)为代表的神经网络，由于光伏阵列可能出现的故障类型较多，因此需要大量训练样本和训练时间，算法结构复杂、实现难度大、诊断时间过长且诊断的精度和可靠

性也不高。

发明内容

本发明人针对上述问题及技术需求，提出了一种基于随机森林算法的光伏阵列故障诊断方法，该方法基于数据驱动的思想，利用随机森林算法构建故障诊断模型，适用于实际光伏阵列的特点，克服了传统神经网络算法需要数据量大、训练时间长等问题，能够简单快速地完成诊断任务。

本发明的技术方案如下：

一种基于随机森林算法的光伏阵列故障诊断方法，该方法包括：

确定光伏阵列在运行过程中的典型运行状态，光伏阵列包括 n 个支路，每个支路包括 m 个光伏组件， m 和 n 均为正整数且 $m \leq 4$ 、 $n \leq 4$ ；

在光伏阵列处于各个典型运行状态时，分别采集光伏阵列中的各个支路和总干路对应的电路参数组，每个电路参数组中包括 k 个电路参数， k 为正整数，根据采集到的 $n+1$ 个电路参数组构建得到一个 p 维的故障特征向量，其中， $p = k * (n+1)$ ；

根据获取到的故障特征向量构建光伏阵列的数据样本集，将数据样本集划分为训练样本集和测试样本集；

基于随机森林算法利用训练样本集构建得到光伏阵列故障诊断模型，并利用测试样本集对光伏阵列故障诊断模型进行测试，光伏阵列故障诊断模型中包括 s 个决策树， $s \geq 2$ 且 s 为正整数；

利用测试完成的光伏阵列故障诊断模型对待诊断光伏阵列进行诊断，得到 s 个决策树对各个典型运行状态的投票结果；

根据各个典型运行状态的投票结果得到待诊断光伏阵列的故障诊断结果，故障诊断结果用于指示光伏阵列中的各个支路的运行状态。

其进一步的技术方案为，基于随机森林算法利用训练样本集构建得到光伏阵列故障诊断模型，包括：

对训练样本集进行 s 次有放回抽样，每次抽样得到一个子训练样本集，共得到 s 个子训练样本集， $s \geq 2$ 且 s 为正整数；

利用每个子训练样本集训练得到一个决策树；

确定训练得到的 s 个决策树汇总得到的集合即为光伏阵列故障诊断模型。

其进一步的技术方案为，在利用子训练样本集训练决策树的过程中，在进

行节点分裂时，从 p 个故障特征中随机选择 q 个故障特征， q 为正整数且 $q \leq p$ ，计算 q 个故障特征中的各个故障特征对应的基尼系数，将与最小的基尼系数对应的故障特征作为最佳分裂特征来进行节点分裂。

其进一步的技术方案为， $k=4$ ，每个电路参数组中包括开路电压、短路电流、最大功率点电压和最大功率点电流。

其进一步的技术方案为， $n=2$ 、 $m=3$ ，光伏阵列的典型运行状态包括五大类共十二小类，分别为：

第一大类为正常运行状态，第一大类包括一小类，为光伏阵列中的各个支路中的各个光伏组件均正常运行时的状态；

第二大类为短路故障状态，第二大类包括三小类，分别为光伏阵列中的一个支路中有一个光伏组件短路时的状态，以及光伏阵列中的一个支路中有两个光伏组件短路时的状态，以及光伏阵列中的两个支路均有一个光伏组件短路时的状态；

第三大类为开路故障状态，第三大类包括一小类，为光伏阵列中有一个支路开路时的状态；

第四大类为阴影故障状态，第四大类包括三小类，分别为光伏阵列中的一个支路中有一个光伏组件存在阴影时的状态，以及光伏阵列中的两个支路中均有一个光伏组件存在阴影时的状态，以及光伏阵列中的一个支路中有一个光伏组件存在阴影、另一个支路中有两个光伏组件存在阴影时的状态，以及光伏阵列中的两个支路中均有两个光伏组件存在阴影时的状态；

第五大类为混合故障状态，第五大类包括四小类，分别为光伏阵列中的一个支路开路、另一个支路中有一个光伏组件存在阴影时的状态，以及光伏阵列中的一个支路中有一个光伏组件短路、另一个支路中有一个光伏组件存在阴影时的状态，以及光伏阵列中的两个支路均有一个光伏组件短路、一个光伏组件存在阴影时的状态，以及光伏阵列中的两个支路均有一个光伏组件短路、两个光伏组件存在阴影时的状态。

本发明的有益技术效果是：

1、本申请公开了一种基于随机森林算法的光伏阵列故障诊断方法，该方法基于数据驱动的思想，利用随机森林算法构建光伏阵列故障诊断模型，适用于实际光伏阵列的特点，克服了传统神经网络算法需要数据量大、训练时间长等问题，能够简单快速地完成诊断任务，快速实现小型光伏阵列，尤其是 3×2 光

伏阵列的故障诊断。

2、本发明采用随机森林的故障诊断模型结构，通过多个决策树的集成，由许多弱分类器构建了一个强分类器，诊断结果由投票产生，即使最多票数的诊断结果错误，还能调取第二、三票数的诊断结果，供维修人员参考，提高维修效率，从而缩短系统的故障时间。

附图说明

图 1 是本申请公开的光伏阵列故障诊断方法的流程图。

图 2-a 是单个太阳能电池组件在不同光照强度下的 U-I 输出特性曲线。

图 2-b 是单个太阳能电池组件在不同光照强度下的 U-P 输出特性曲线。

10 图 2-c 是单个太阳能电池组件在不同温度情况下的 U-I 输出特性曲线。

图 2-d 是单个太阳能电池组件在不同温度情况下的 U-P 输出特性曲线。

图 3 是单个太阳能电池组件在光照强度 600W/m^2 、 800W/m^2 、 1000W/m^2 情况下的输出特性曲线。

图 4-a 是光伏阵列中有一个支路发生故障时的 U-I 输出特性曲线。

15 图 4-b 是光伏阵列中有一个支路发生故障时的 U-P 输出特性曲线。

图 4-c 是光伏阵列中有两个支路发生故障时的 U-I 输出特性曲线。

图 4-d 是光伏阵列中有两个支路发生故障时的 U-P 输出特性曲线。

图 5-a 是光伏阵列处于状态标签 F1 对应的运行状态时的 U-I 输出特性曲线。

图 5-b 是光伏阵列处于状态标签 F2 对应的运行状态时的 U-I 输出特性曲线。

20 图 5-c 是光伏阵列处于状态标签 F5 对应的运行状态时的 U-I 输出特性曲线。

图 5-d 是光伏阵列处于状态标签 F8 对应的运行状态时的 U-I 输出特性曲线。

图 5-e 是光伏阵列处于状态标签 F13 对应的运行状态时的 U-I 输出特性曲线。

具体实施方式

25 下面结合附图对本发明的具体实施方式做进一步说明。

本申请公开了一种基于随机森林算法的光伏阵列故障诊断方法，请参考图 1，整个故障诊断流程包括如下步骤：

第一步，搭建 $m \times n$ 形式的光伏阵列，该光伏阵列包括并联的 n 个支路，每个支路包括串联的 m 个光伏组件， m 和 n 均为正整数。阵列中的光伏组件通常是太阳能电池组件，每个光伏组件规格相同，以保证光伏阵列的电位平衡，并

30

且旁路二极管、隔离二极管、快速熔断器、电压和电流传感器等保护、检测装置均安装正确并正常工作。在实际实现时，本申请公开的方法更适用于小型光伏阵列，因此通常有 $m \leq 4$ 、 $n \leq 4$ 。

本申请以更为典型的 $n=2$ 、 $m=3$ 的情况为例进行说明，采用 6 块无锡尚德太阳能有限公司 SunTech STP 270-24/Vd 型太阳能电池组件搭建光伏阵列，该型号的太阳能电池组件的参数规格表如下表所示：

参数类型	参数值
开路电压(V_{oc})	44.5V
最大功率点电压(V_m)	35.0V
短路电流(I_{sc})	8.20A
最大功率点电流(I_m)	7.71A
最大功率(P_{max})	270W
运行温度	-40℃ ~ +85℃
系统最大承受电压	1000V DC
串联后保险丝的额定电流	20A
功率误差	±3%

单个太阳能电池组件在温度情况相同的条件下，在不同光照强度下的 U-I（电压-电流）输出特性曲线如图 2-a 所示、U-P（电压-功率）输出特性曲线如图 2-b 所示；在光照强度相同的情况下，温度从 35℃ 下降至 15℃（图中分别为 35℃、30℃、25℃、20℃和 15℃）时的 U-I 输出特性曲线如图 2-c 所示、U-P 输出特性曲线如图 2-d 所示。可以看出，随着光照强度的增加和温度环境的降低，太阳能电池组件的输出功率增加，发电量增大。单独给出了光照强度 600W/m²、800 W/m²、1000 W/m² 情况下的输出特性曲线如图 3 所示。

第二步，确定光伏阵列在运行过程中的典型运行状态，光伏阵列的典型运行状态用于指示光伏阵列中的各个支路的运行状态。光伏阵列在实际运行时会

出现较多类型的故障，主要划分为四种：短路故障、开路故障、阴影故障以及混合故障，再加上正常运行的情况，光伏阵列的运行状态主要包括五大类：正常运行状态、短路故障状态、开路故障状态、阴影故障状态以及混合故障状态。

当光伏阵列中的一个支路或多个支路发生故障时，故障的支路以及整个光伏阵列的输出特性曲线都会发生变化，如图 4-a 至 4-d 分别示出了不同情况下的故障支路、故障阵列、正常支路和正常阵列的输出特性曲线，其中正常支路是指正常运行的支路，正常阵列是指各个支路均正常运行时的光伏阵列，故障支路是指发生故障时的支路，故障阵列是指有支路发生故障时的光伏阵列，具体的：光伏阵列中有一个支路发生故障时的 U-I 输出特性曲线如图 4-a 所示；有一个支路发生故障时的 U-P 输出特性曲线如图 4-b 所示；有两个支路发生故

障时的 U-I 输出特性曲线如图 4-c 所示；有两个支路发生故障时的 U-P 输出特性曲线如图 4-d 所示。由图中可以看出，当光伏组件发生故障导致其所在支路发生故障时，无论是该发生故障的支路还是整个光伏阵列，都会出现一个明显的功率损失，正常支路和正常阵列的输出特性曲线呈单峰特性，而故障支路和故障阵列呈多峰特性。

光伏阵列的五大类运行状态可以进一步分为多个小类，实际可以在其中选择若干最为常见的小类作为光伏阵列的典型运行状态，其具体选择可以根据实际经验进行。比如在本申请 $n=2$ 、 $m=3$ 的例子中，可以在这些运行状态大类中选择十三小类最为常见的运行状态作为典型运行状态，分别为：

(1)、第一大类为正常运行状态，第一大类包括一小类，为光伏阵列中的各个支路中的各个光伏组件均正常运行时的状态。

(2)、第二大类为短路故障状态，第二大类包括三小类，分别为光伏阵列中的一个支路中有一个光伏组件短路时的状态，以及光伏阵列中的一个支路中有两个光伏组件短路时的状态，以及光伏阵列中的两个支路均有一个光伏组件短路时的状态。

需要说明的是，短路故障可以看作是一种特殊的阴影故障：光伏组件被完全遮挡而无法工作而被旁路二极管全程短路，但与阴影故障不同的是，一条支路中最多存在两个光伏组件短路，若全部光伏组件短路，则熔断装置熔断，该支路会开路。

(3)、第三大类为开路故障状态，第三大类包括一小类，为光伏阵列中有一个支路开路时的状态。

(4)、第四大类为阴影故障状态，第四大类包括三小类，分别为光伏阵列中的一个支路中有一个光伏组件存在阴影时的状态，以及光伏阵列中的两个支路中均有一个光伏组件存在阴影时的状态，以及光伏阵列中的一个支路中有一个光伏组件存在阴影、另一个支路中有两个光伏组件存在阴影时的状态，以及光伏阵列中的两个支路中均有两个光伏组件存在阴影时的状态。需要说明的是，阴影故障与短路故障不同，可以存在所有光伏组件均存在阴影的情况。

(5)、第五大类为混合故障状态，第五大类包括四小类，分别为光伏阵列中的一个支路开路、另一个支路中有一个光伏组件存在阴影时的状态，以及光伏阵列中的一个支路中有一个光伏组件短路、另一个支路中有一个光伏组件存在阴影时的状态，以及光伏阵列中的两个支路均有一个光伏组件短路、一个光

伏组件存在阴影时的状态,以及光伏阵列中的两个支路均有一个光伏组件短路、两个光伏组件存在阴影时的状态。

为了清楚简洁地表示各个典型运行状态,可以用状态标签表示各个典型运行状态,上述十三类典型运行状态以表格的形式表示如下:

运行状态大类	状态标签	光伏阵列的典型运行状态的说明
正常运行状态	F1	各个支路中的各个光伏组件均正常运行
短路故障状态	F2	一个支路中有一个光伏组件短路
	F3	一个支路中有两个光伏组件短路
	F4	两个支路均有一个光伏组件短路
开路故障状态	F5	有一个支路开路
阴影故障状态	F6	一个支路中有一个光伏组件存在阴影
	F7	两个支路中均有一个光伏组件存在阴影
	F8	一个支路中有一个光伏组件存在阴影、另一个支路中有两个光伏组件存在阴影
	F9	两个支路中均有两个光伏组件存在阴影
混合故障状态	F10	一个支路开路、另一个支路中有一个光伏组件存在阴影
	F11	一个支路中有一个光伏组件短路、另一个支路中有一个光伏组件存在阴影
	F12	两个支路均有一个光伏组件短路、一个光伏组件存在阴影
	F13	两个支路均有一个光伏组件短路、两个光伏组件存在阴影

当光伏阵列处于不同的运行状态时,光伏阵列中的两个支路以及整个光伏阵列的 U-I 输出特性曲线和 U-P 输出特性曲线均不相同,以 U-I 输出特性曲线为例:光伏阵列处于状态标签 F1 对应的运行状态时的 U-I 输出特性曲线如图 5-a 所示,光伏阵列处于状态标签 F2 对应的运行状态时的 U-I 输出特性曲线如图 5-b 所示,光伏阵列处于状态标签 F5 对应的运行状态时的 U-I 输出特性曲线如图 5-c 所示,光伏阵列处于状态标签 F8 对应的运行状态时的 U-I 输出特性曲线如图 5-d 所示,光伏阵列处于状态标签 F13 对应的运行状态时的 U-I 输出特性曲线如图 5-e 所示。需要说明的是,上述各个图中由于曲线重合,使得曲线上的标记有所叠加,比如支路 B 和阵列的曲线重合会导致图上显示六角形图示,但本领域技术人员应该可以从附图中明白其实际表示的含义,另外本申请对光伏阵列处于其他各个典型运行状态时的 U-I 输出特性曲线不一一示例。

第三步,构建故障特征向量。主要在光伏阵列处于各个典型运行状态时,分别采集光伏阵列中的各个支路和总干路对应的电路参数组,光伏组件所处的光照强度和温度的区间大致分别在 $[800, 900]\text{W/m}^2$ 和 $[40, 50]^\circ\text{C}$ 。每个电路参数组中包括 k 个电路参数, k 为正整数,在本申请中, $k=4$, 本申请采集开路电压 V_{oc} 、短路电流 I_{sc} 、最大功率点电压 V_m 和最大功率点电流 I_m 这四个参数构成电路参数组,这是由于这四个参数能够很好的刻画光伏组件的 U-I 输出特性曲线。

在光伏阵列中，共采集得到 $n+1$ 个电路参数组，每个电路参数组中包括 k 个电路参数，因此可以根据采集到的电路参数组构建得到一个 p 维的故障特征向量，其中， $p=k*(n+1)$ 。在本申请举例的 $3*2$ 光伏阵列中，主要有两个支路和一个总干路，因此需要采集两个支路和总干路的 V_{oc} 、 I_{sc} 、 V_m 和 I_m ，因此会得到一个 12 维的故障特征向量。

第四步，得到故障特征向量后，就能汇总并构建光伏阵列的数据样本集，由故障特征向量汇总得到数据样本集的方法是本领域技术人员都知道的方法，本申请对此不进行展开叙述。对数据样本集进行划分得到训练样本集和测试样本集，通常可以按照 2:1 的比例划分成为训练样本集和测试样本集。

第五步，基于随机森林算法的光伏阵列故障诊断模型的构建，设训练样本集为 X ， p 是故障特征向量的维度，在本申请举例的 $3*2$ 光伏阵列中， $p=12$ ； s 是故障诊断中需要建立的决策树的数目， $s \geq 2$ 且 s 为正整数，则基于随机森林算法利用训练样本集构建得到光伏阵列故障诊断模型的步骤可以简述为：

1、对训练样本集 X 进行 s 次有放回抽样，每次抽样得到一个子训练样本集，共得到 s 个子训练样本集，记第 i 个子训练样本集为 X_i ， i 为参数， $i=1,2,\dots,s$ 。

2、利用每个子训练样本集训练得到一个决策树，也即利用子训练样本集 X_i 训练决策树 Q_i ，在进行节点分裂时，从 p 个故障特征中随机选择 q 个故障特征， q 为正整数且 $q \leq p$ ，计算 q 个故障特征中的各个故障特征对应的基尼系数，将与最小的基尼系数对应的故障特征作为最佳分裂特征来进行节点分裂，基尼系数的计算方法可以采用现有的计算公式，本申请对此不作详细介绍。

3、将 s 个决策树 Q_i ($i=1,2,\dots,s$) 进行汇总，汇总得到的集合即为光伏阵列故障诊断模型。

第六步，光伏阵列故障诊断模型的测试。利用测试样本集对光伏阵列故障诊断模型进行测试，可以最终得到诊断精度、平均训练时间和测试时间等指标。将本申请基于随机森林 (RF) 算法的故障诊断与基于神经网络 (BP) 和极限学习机 (ELM) 的故障诊断进行对比，实际实验结果如下表所示。从表中可以看出，基于随机森林 (RF) 的故障诊断方法的诊断精度较高，同时训练和测试时间的表现也较好。

诊断方法	训练时间/s	测试时间/s	诊断精度/%
BP	2.6664	0.10852	96.92
ELM	0.3449	0.00663	97.69
RF	0.3718	0.01052	98.46

第七步，利用测试完成的光伏阵列故障诊断模型对待诊断光伏阵列进行诊断，也即采用光伏阵列故障诊断模型中的 s 个决策树进行投票，由于随机森林的诊断结果是由投票产生的，因此会得到 s 个决策树对各个典型运行状态的投票结果。根据各个投票结果即可以得到待诊断光伏阵列的故障诊断结果，也即确定待诊断光伏阵列中各个支路的运行状态。通常情况下，票数最多的典型运行状态即为待诊断光伏阵列的实时运行状态，也即可以确定光伏阵列出现的故障类型。以 20 棵决策树为例，5 组实测结果的部分投票结果汇总，以及对应的故障诊断结果如下表，表中以状态标签进行表示：

序号	投票结果	故障诊断结果	故障真值
1	F2: 19 票、F11: 1 票	F2	F2
2	F5: 17 票、F10: 3 票	F5	F5
3	F11: 10 票、F7: 5 票、F1: 2 票、F4: 1 票、F6: 1 票	F11	F7
4	F9: 11 票、F8: 9 票	F9	F8
5	F11: 15 票、F2: 4 票、F4: 1 票	F11	F11

如上表可以看出，对于第一组数据，状态标签 F2 对应的典型运行状态对应的票数为 19 票为最多，则诊断结果为待诊断光伏阵列处于状态标签 F2 对应的典型运行状态，也即诊断得到待诊断光伏阵列的一个支路中有一个光伏组件短路，并且这一诊断结果与故障真值是相符的。

对于第三组数据，状态标签 F11 对应的典型运行状态对应的票数为 10 票为最多，则诊断结果为待诊断光伏阵列处于状态标签 F10 对应的典型运行状态，也即诊断得到待诊断光伏阵列的一个支路开路、另一个支路中有一个光伏组件存在阴影，虽然这一诊断结果与故障真值不符，但由于本申请会得到各个典型运行状态对应的投票结果，因此即使发生了这种诊断错误的情况，实际运行状态的类型也会出现在投票统计中（可以看到状态标签 F7 对应的典型运行状态对应的票数为 5 票，票数第二多），相当于有备选诊断结果，所以工作人员能对各个典型运行状态的票数进行调取并查看票数统计情况，从而找出可能发生故障的类型，即使是最高票对应的诊断发生了错误，仍可以按票数高低依次进行故障排除，也即本申请实际是提供了一个诊断路线，从而提高维修效率，缩短系统的故障时间。

以上所述的仅是本申请的优选实施方式，本发明不限于以上实施例。可以理解，本领域技术人员在不脱离本发明的精神和构思的前提下直接导出或联想到的其他改进和变化，均应认为包含在本发明的保护范围之内。

权 利 要 求 书

1. 一种基于随机森林算法的光伏阵列故障诊断方法，其特征在于，所述方法包括：

5 确定光伏阵列在运行过程中的典型运行状态，所述光伏阵列包括 n 个支路，每个支路包括 m 个光伏组件， m 和 n 均为正整数且 $m \leq 4$ 、 $n \leq 4$ ；

在所述光伏阵列处于各个所述典型运行状态时，分别采集所述光伏阵列中的各个支路和总干路对应的电路参数组，每个电路参数组中包括 k 个电路参数， k 为正整数，根据采集到的 $n+1$ 个电路参数组构建得到一个 p 维的故障特征向量，其中， $p = k * (n+1)$ ；

根据获取到的所述故障特征向量构建所述光伏阵列的数据样本集，将所述数据样本集划分为训练样本集和测试样本集；

基于随机森林算法利用所述训练样本集构建得到光伏阵列故障诊断模型，并利用所述测试样本集对所述光伏阵列故障诊断模型进行测试，所述光伏阵列故障诊断模型中包括 s 个决策树， $s \geq 2$ 且 s 为正整数；

利用测试完成的所述光伏阵列故障诊断模型对待诊断光伏阵列进行诊断，得到所述 s 个决策树对各个所述典型运行状态的投票结果；

根据各个所述典型运行状态的投票结果得到所述待诊断光伏阵列的故障诊断结果，所述故障诊断结果用于指示所述光伏阵列中的各个支路的运行状态。

20 2、根据权利要求 1 所述的方法，其特征在于，所述基于随机森林算法利用所述训练样本集构建得到光伏阵列故障诊断模型，包括：

对所述训练样本集进行 s 次有放回抽样，每次抽样得到一个子训练样本集，共得到 s 个子训练样本集；

利用每个所述子训练样本集训练得到一个决策树；

25 确定训练得到的 s 个决策树汇总得到的集合即为所述光伏阵列故障诊断模型。

3、根据权利要求 2 所述的方法，其特征在于，

在利用子训练样本集训练决策树的过程中，在进行节点分裂时，从 p 个故障特征中随机选择 q 个故障特征， q 为正整数且 $q \leq p$ ，计算所述 q 个故障特征中的各个故障特征对应的基尼系数，将与最小的基尼系数对应的故障特征作为最

佳分裂特征来进行节点分裂。

4、根据权利要求1所述的方法，其特征在于， $k=4$ ，每个所述电路参数组中包括开路电压、短路电流、最大功率点电压和最大功率点电流。

5、根据权利要求1所述的方法，其特征在于， $n=2$ 、 $m=3$ ，所述光伏阵列的典型运行状态包括五大类共十二小类，分别为：

第一大类为正常运行状态，所述第一大类包括一小类，为所述光伏阵列中的各个支路中的各个光伏组件均正常运行时的状态；

10 第二大类为短路故障状态，所述第二大类包括三小类，分别为所述光伏阵列中的一个支路中有一个光伏组件短路时的状态，以及所述光伏阵列中的一个支路中有两个光伏组件短路时的状态，以及所述光伏阵列中的两个支路均有一个光伏组件短路时的状态；

第三大类为开路故障状态，所述第三大类包括一小类，为所述光伏阵列中有一个支路开路时的状态；

15 第四大类为阴影故障状态，所述第四大类包括三小类，分别为所述光伏阵列中的一个支路中有一个光伏组件存在阴影时的状态，以及所述光伏阵列中的两个支路中均有一个光伏组件存在阴影时的状态，以及所述光伏阵列中的一个支路中有一个光伏组件存在阴影、另一个支路中有两个光伏组件存在阴影时的状态，以及所述光伏阵列中的两个支路中均有两个光伏组件存在阴影时的状态；

20 第五大类为混合故障状态，所述第五大类包括四小类，分别为所述光伏阵列中的一个支路开路、另一个支路中有一个光伏组件存在阴影时的状态，以及所述光伏阵列中的一个支路中有一个光伏组件短路、另一个支路中有一个光伏组件存在阴影时的状态，以及所述光伏阵列中的两个支路均有一个光伏组件短路、一个光伏组件存在阴影时的状态，以及所述光伏阵列中的两个支路均有一个光伏组件短路、两个光伏组件存在阴影时的状态。

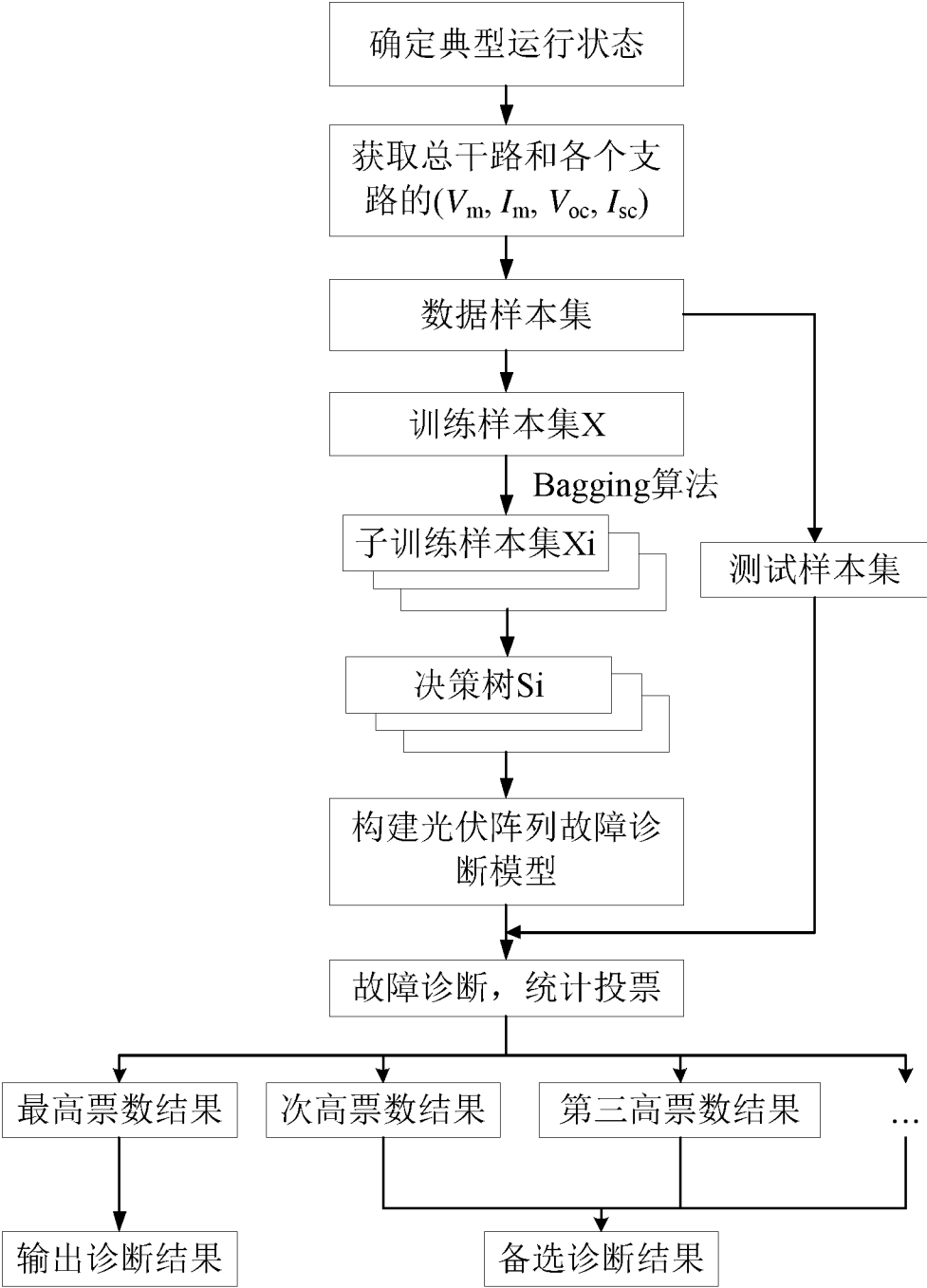


图 1

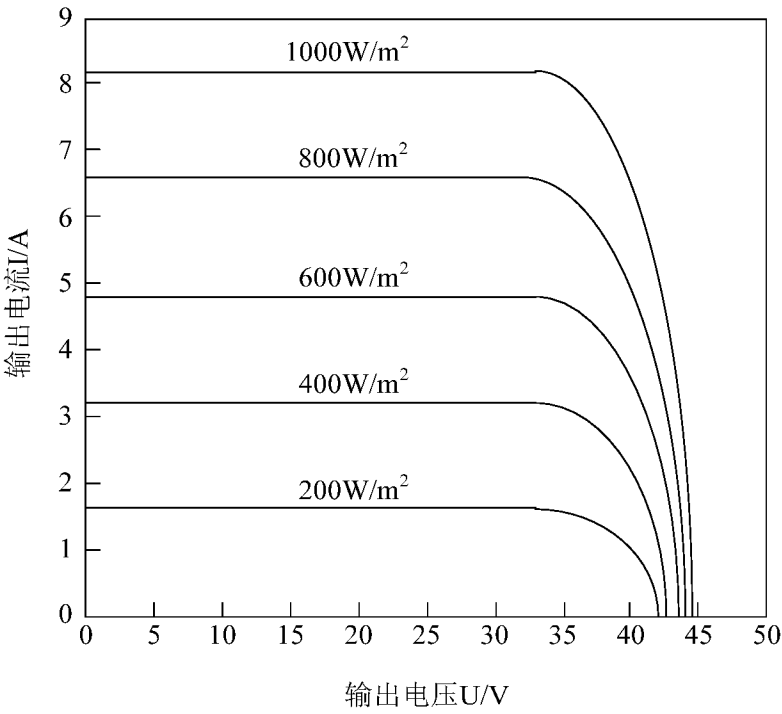


图 2-a

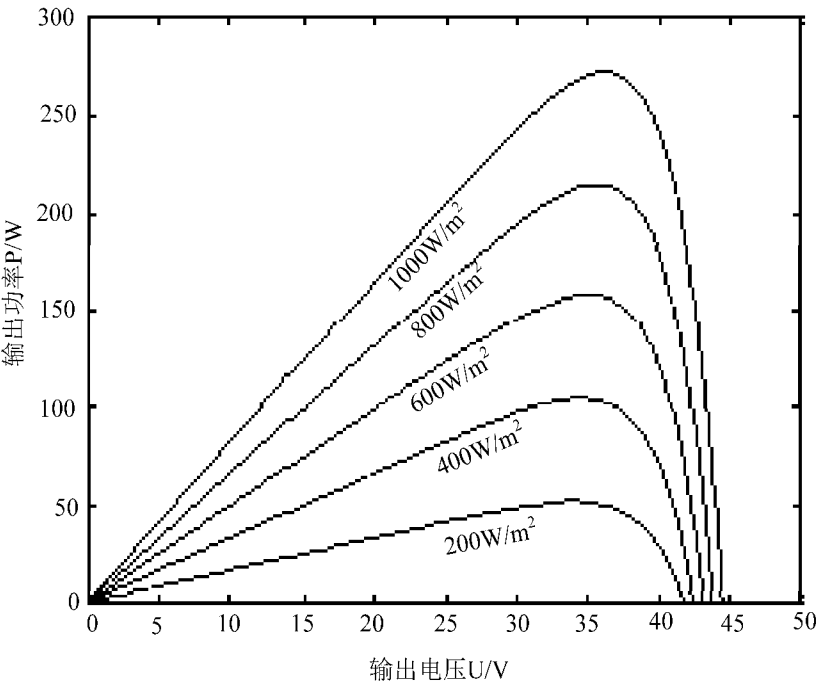


图 2-b

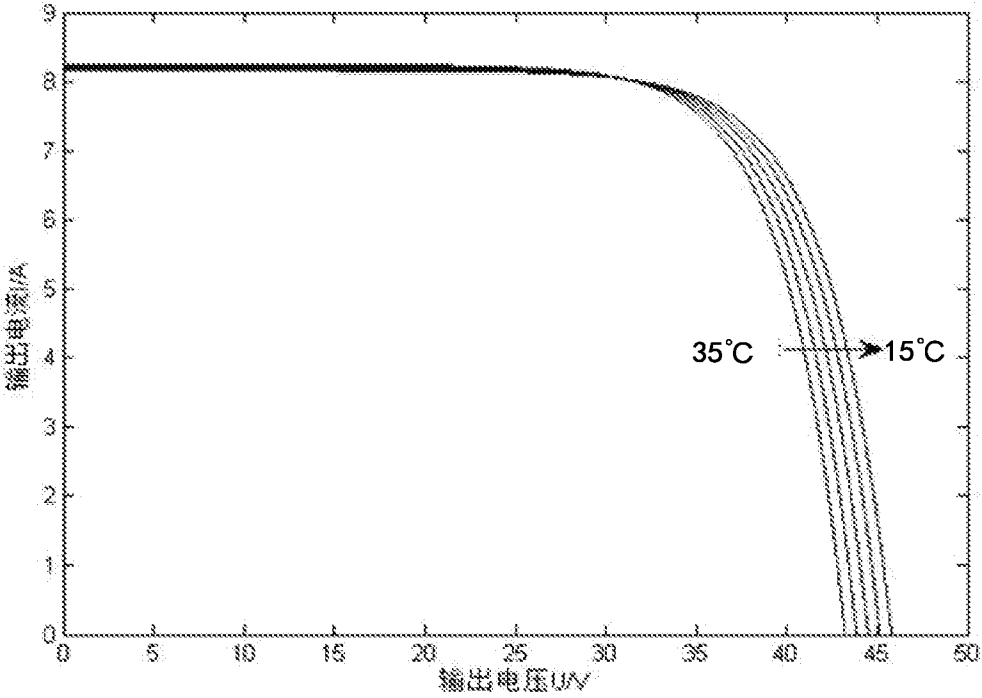


图 2-c

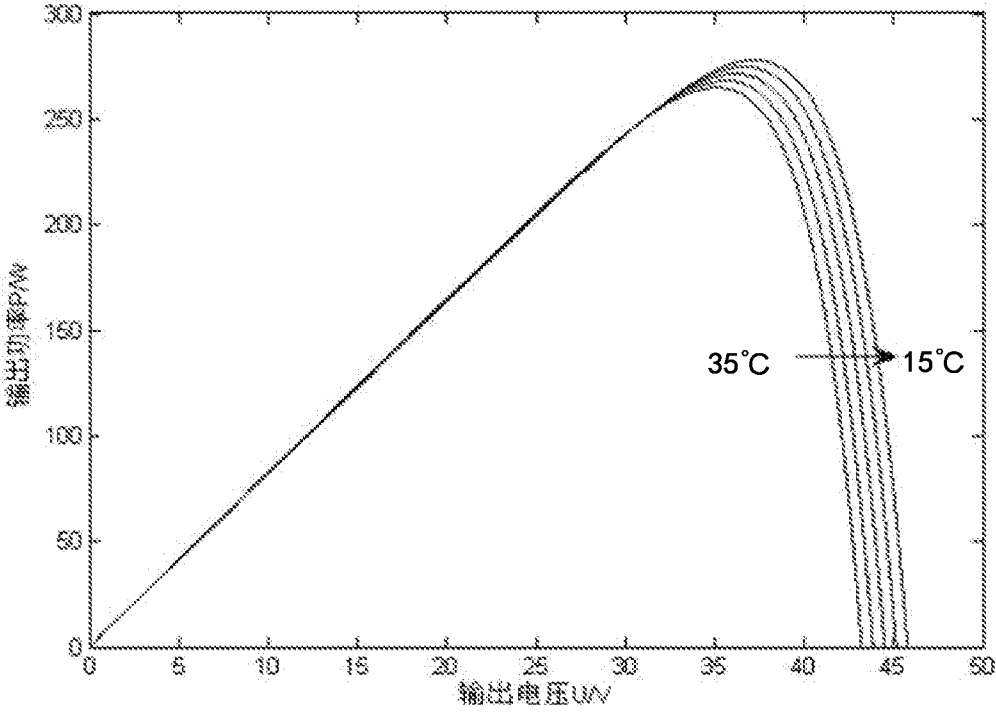


图 2-d

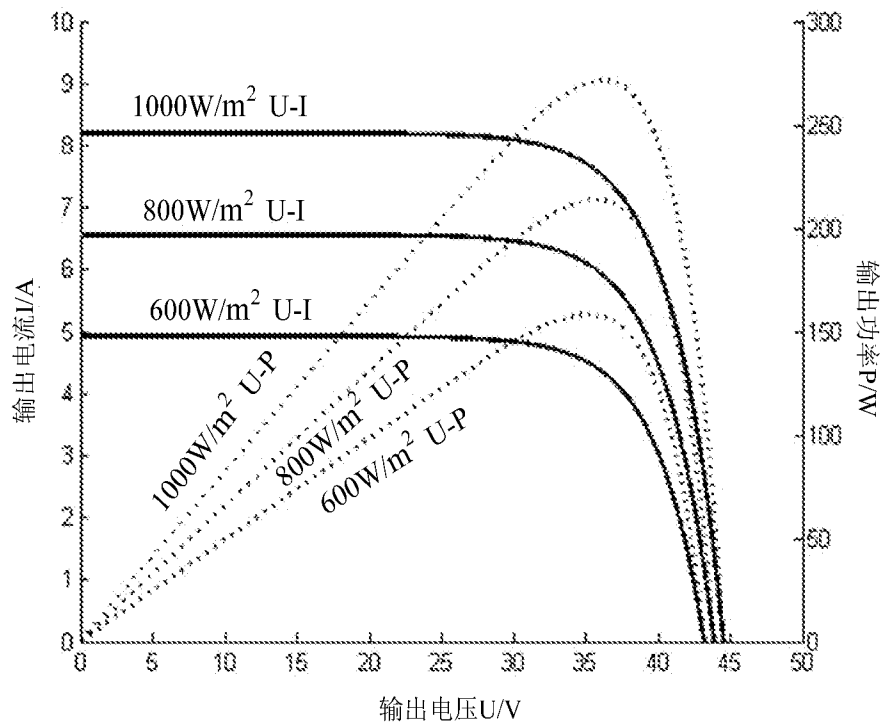


图 3

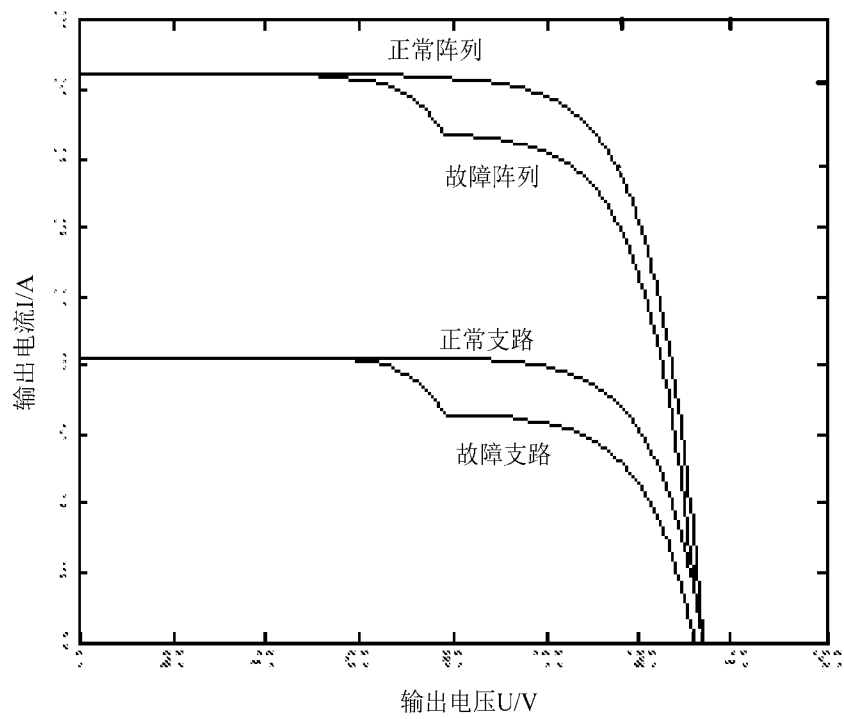


图 4-a

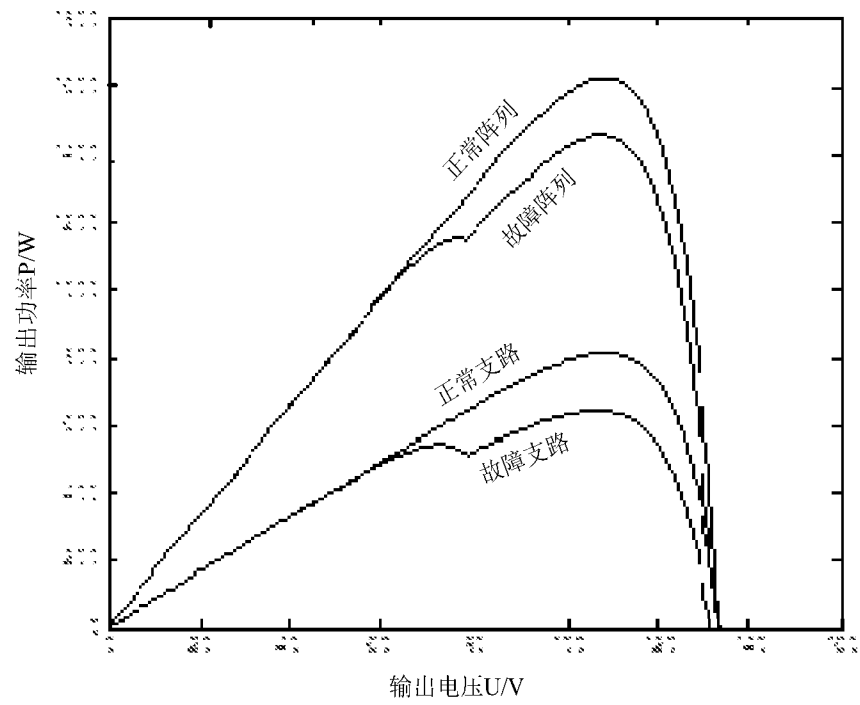


图 4-b

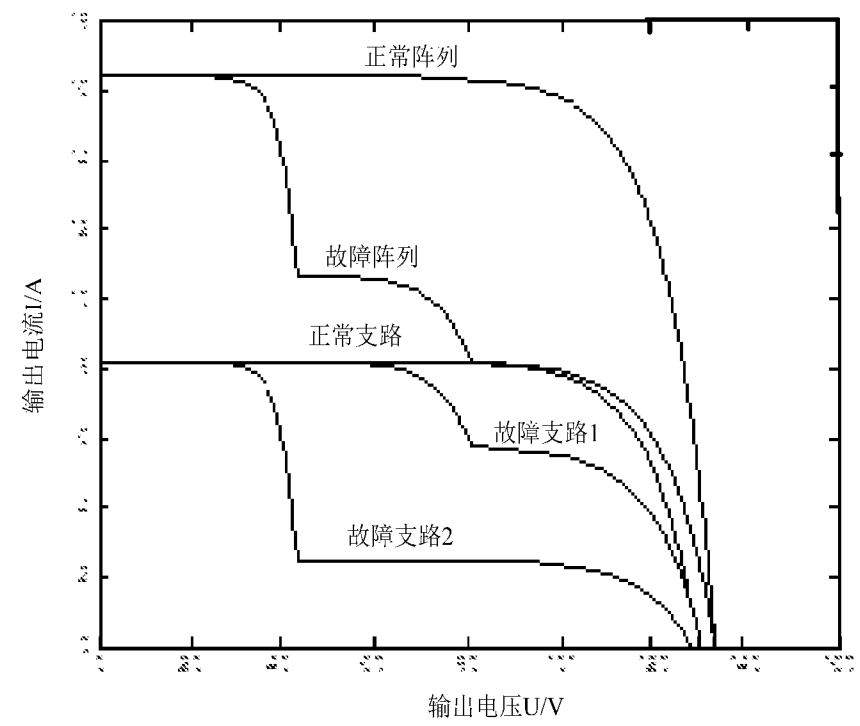


图 4-c

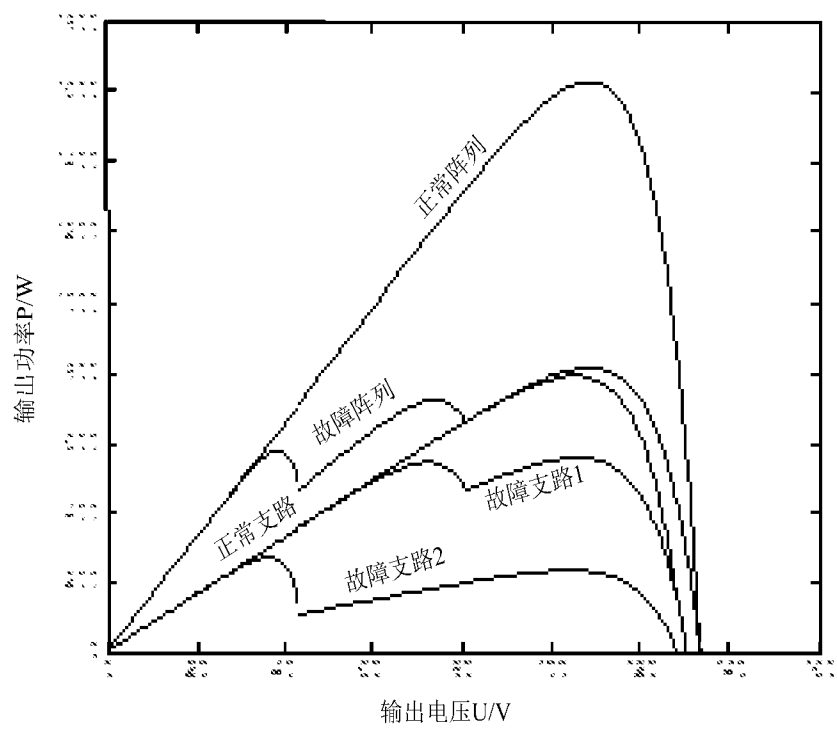


图 4-d

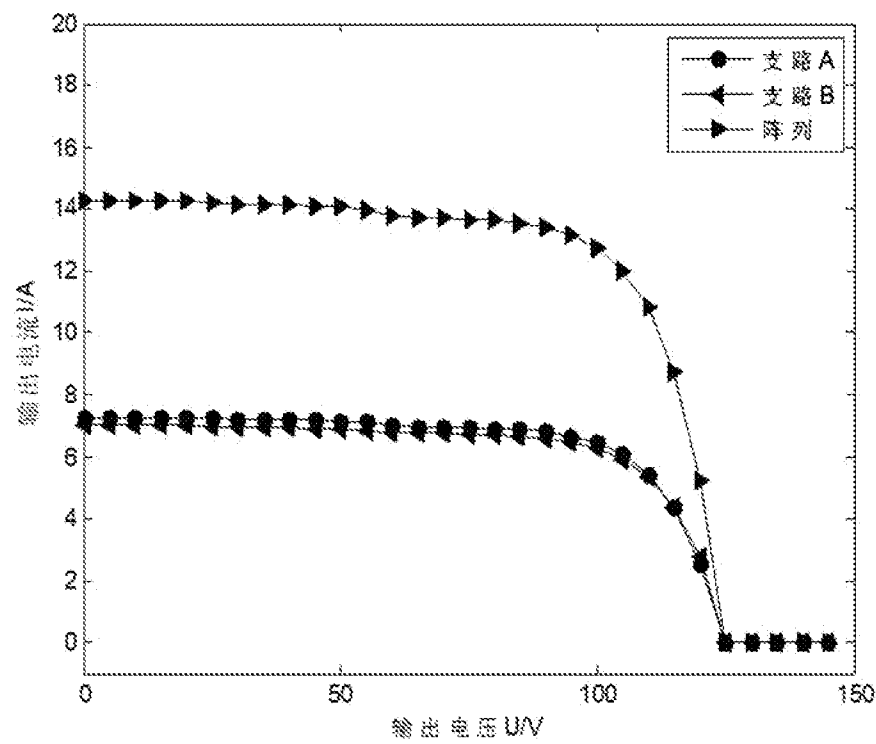


图 5-a

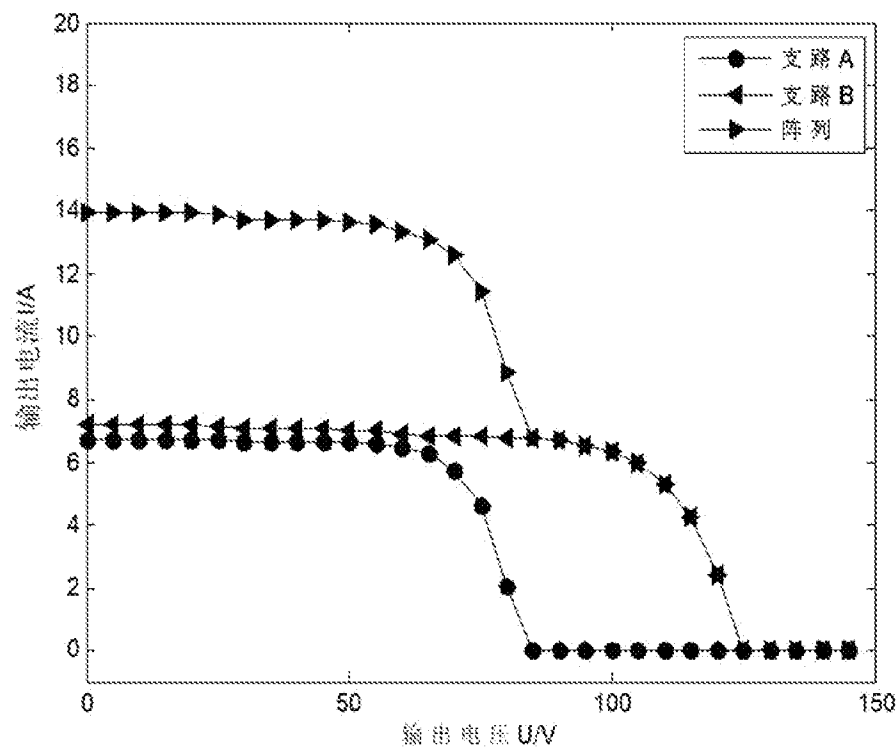


图 5-b

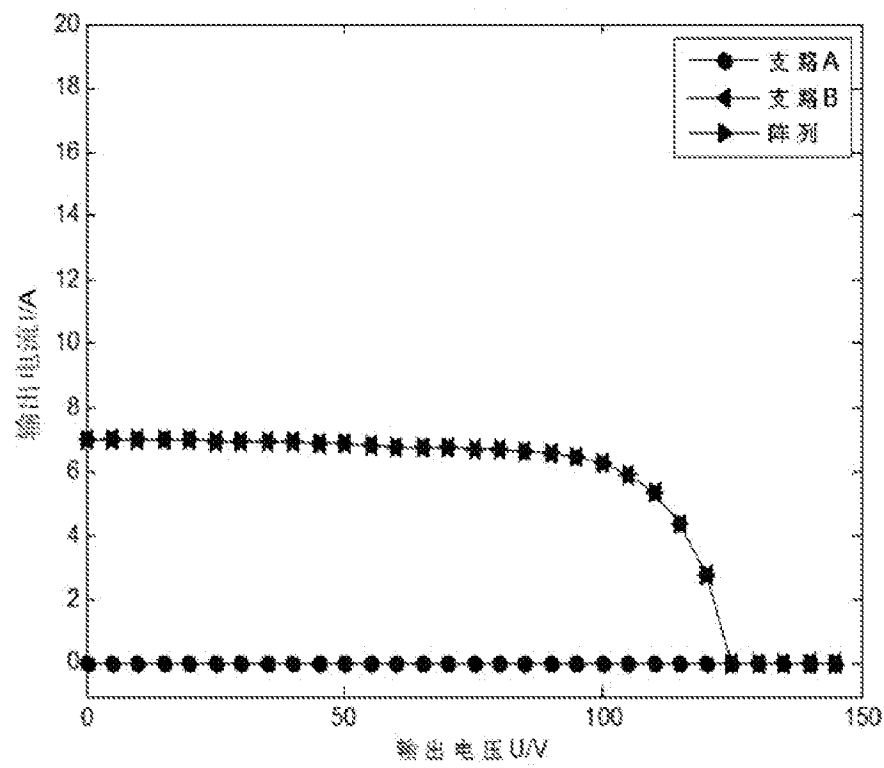


图 5-c

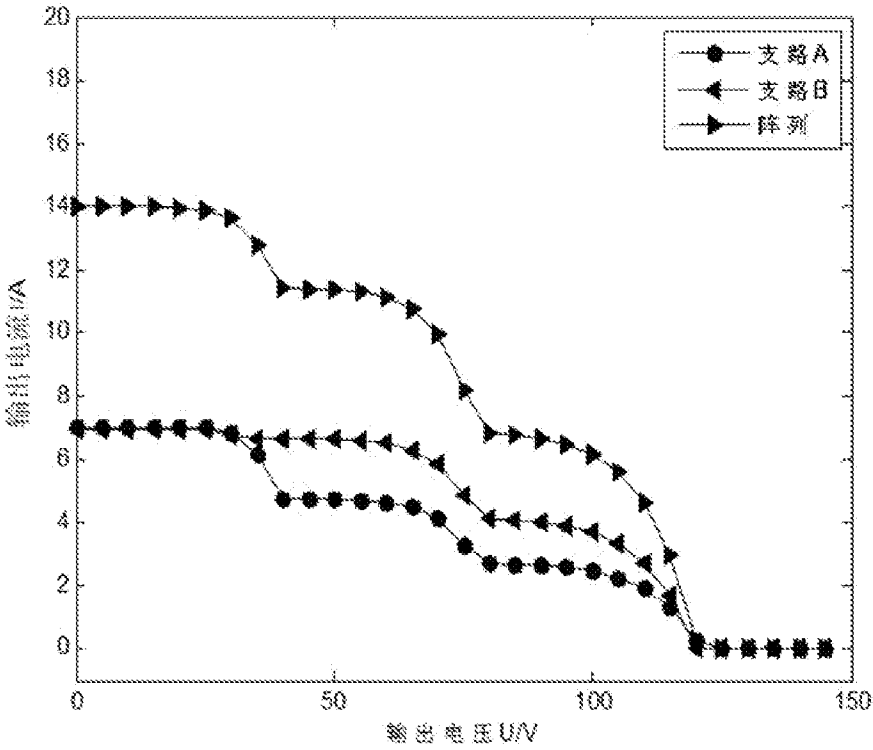


图 5-d

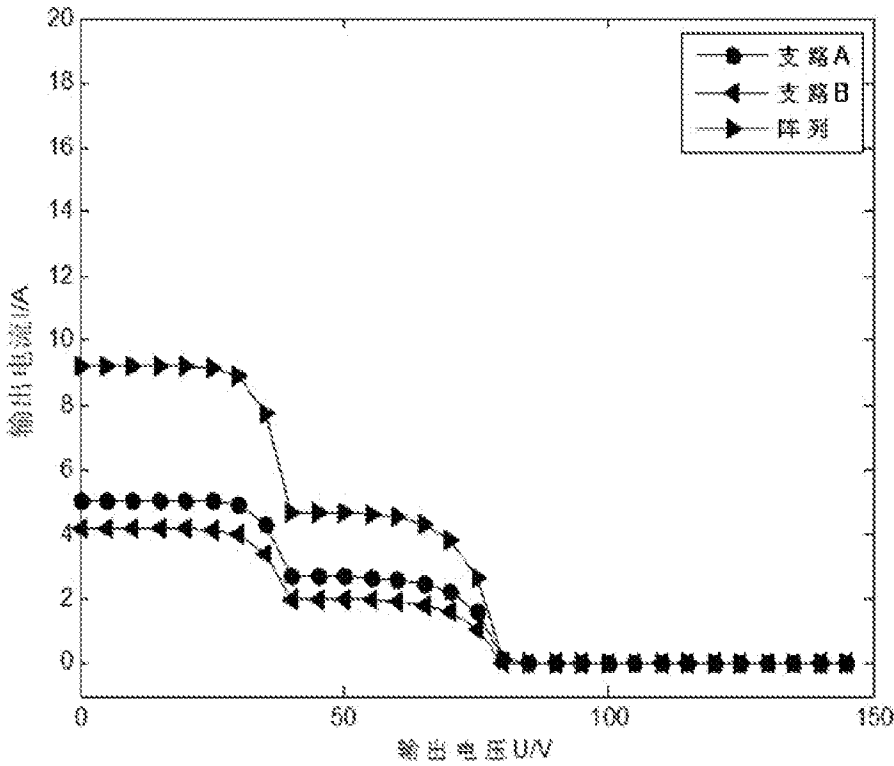


图 5-e

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/102907

A. CLASSIFICATION OF SUBJECT MATTER

H02S 50/10(2014.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H02S

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CPRSABS; CNABS; SIPOABS; CNKI: 故障, 诊断, 光伏, 阵列, 随机森林, fault, diagnosis, photovoltaic, array, random w forest

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 108062571 A (FUZHOU UNIVERSITY) 22 May 2018 (2018-05-22) entire document	1-5
A	CN 106961249 A (GUANGXI UNIVERSITY) 18 July 2017 (2017-07-18) entire document	1-5
A	CN 105141255 A (HOHAI UNIVERSITY, CHANGZHOU CAMPUS) 09 December 2015 (2015-12-09) entire document	1-5
A	CN 107579707 A (JIANGSU UNIVERSITY) 12 January 2018 (2018-01-12) entire document	1-5

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

06 March 2019

Date of mailing of the international search report

19 March 2019

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/102907

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	108062571	A	22 May 2018	None	
CN	106961249	A	18 July 2017	None	
CN	105141255	A	09 December 2015	None	
CN	107579707	A	12 January 2018	None	

国际检索报告

国际申请号

PCT/CN2018/102907

A. 主题的分类 H02S 50/10 (2014.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类		
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) H02S 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CPRSABS; CNABS; SIPOABS; CNKI: 故障, 诊断, 光伏, 阵列, 随机森林, fault, diagnosis, photovoltaic, array, random w forest		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN 108062571 A (福州大学) 2018年 5月 22日 (2018 - 05 - 22) 全文	1-5
A	CN 106961249 A (广西大学) 2017年 7月 18日 (2017 - 07 - 18) 全文	1-5
A	CN 105141255 A (河海大学常州校区) 2015年 12月 9日 (2015 - 12 - 09) 全文	1-5
A	CN 107579707 A (江苏大学) 2018年 1月 12日 (2018 - 01 - 12) 全文	1-5
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件		
国际检索实际完成的日期 2019年 3月 6日		国际检索报告邮寄日期 2019年 3月 19日
ISA/CN的名称和邮寄地址 中国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 明媚 电话号码 62411851

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/102907

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	108062571	A	2018年 5月 22日	无	
CN	106961249	A	2017年 7月 18日	无	
CN	105141255	A	2015年 12月 9日	无	
CN	107579707	A	2018年 1月 12日	无	