

DOI: 10.12158/j.2096-3203.2023.01.023

## 变电站图纸物理回路建模方法及实现

侯进<sup>1,2</sup>, 王祥宇<sup>1,2,3</sup>, 郝彦超<sup>1,2,3</sup>, 文志龙<sup>1,2,3</sup>, 李昀喆<sup>1,2,3</sup>, 康萍萍<sup>1,2</sup>

1. 西南交通大学信息科学与技术学院智能感知智慧运维实验室, 四川 成都 611756;
2. 西南交通大学综合交通大数据应用技术国家工程实验室, 四川 成都 611756;
3. 西南交通大学唐山研究院, 河北 唐山 063000)

**摘要:**目前变电站原始图纸信息保存单一、数字化解析程度不高,人工解析变电站原始图纸工作量庞大。为了解决该问题,提出基于可携带文档格式(portable document format, PDF)图纸智能化提取变电站设备及其连接关系信息、构建物理回路模型的方法,并根据智能变电站物理配置描述(substation physical configuration description, SPCD)文件对该方法进行具体实现。首先,对图纸中提取的图元信息进行处理;然后,通过字符串相似度匹配结合KR的贪心串覆盖(running Karp-Rabin greedy string tiling, RKR-GST)算法实现电气符号的识别,使用梯度提升决策树和逻辑回归(gradient boosting decision tree-logistics regression, GBDT-LR)混合算法进行变电站图像分类;最后,根据SPCD文件完成从原始图纸到物理回路模型的数字化描述。实验结果表明,在图元存在误差的情况下,电气符号匹配正确率达到93%,物理回路识别正确率超过90%。

**关键词:**智能变电站;二维图纸;配置描述文件;物理回路;图像处理;图纸活化;辅助设计

**中图分类号:** TM933; TP391

**文献标志码:** A

**文章编号:** 2096-3203(2023)01-0193-08

### 0 引言

目前变电站施工改扩和运行维护严重依赖工程图纸生成的配置文件实现二次应用<sup>[1]</sup>。但现存大量变电站暂不具备完善的二次系统物理回路模型,设计单位依然通过光缆联系原始图纸表达物理回路的设计,工程实施单位同样根据光缆联系原始图纸完成二次系统物理回路的施工<sup>[2-3]</sup>。其中描述物理回路模型设计的变电站物理配置描述(substation physical configuration description, SPCD)文件<sup>[4]</sup>所需的关键信息来自原始图纸,图纸不带有任何属性,信息保存单一,很难对其进行加工和后续处理。通常在新建变电站时,都需要设计人员根据原始二次图纸手动提取信息,保存成光缆清册 Excel 文件、设计图纸计算机辅助设计(computer aided design, CAD)文件,或通过配置工具生成模型配置可扩展标记语言(extensional marked language, XML)文件,工作量庞大,不利于智能变电站数字化技术的发展和二次应用的开发<sup>[5]</sup>。

针对上述问题,亟须研究变电站原始图纸的关键信息智能提取、解析和应用技术。已有研究中,文献[6]通过环形分割特征方法识别对象图例,并通过连通域标记方法识别拓扑关系,但不适用于变

电站对象图例。文献[7]仅研究提取图纸标注信息,文献[8]仅研究提取图纸标题栏内容,均未涉及电气符号等的识别。文献[9]基于CAD利用电气符号构建图元矩阵,通过相似矩阵匹配电气符号,属于对属性图<sup>[10]</sup>的改进,文献[11]基于图像通过YOLOv3模型识别电气符号,但仅对电气符号进行识别,未进一步实现变电站的二次应用。综上,目前不存在从变电站可携带文档格式(portable document format, PDF)原始图纸到智能变电站配置文件应用过程的相关研究,而以物理设备连接的SPCD配置系统,已经开始了重要信息的数字化研究。

基于上述问题,文中首先从SPCD进行分析,针对PDF原始图纸,提取并处理图元信息;然后基于图元实现电气符号识别、变电站图像分类和拓扑信息解析;最后设计物理回路建模算法,得到单张图纸对应的中间模型配置描述(intermediate model configuration description, IMCD)<sup>[12]</sup>文件,以更加智能化的方式完成变电站图纸物理回路建模,实现SPCD的辅助设计。

### 1 基于SPCD的物理回路

单装置物理配置描述(individual physical configuration description, IPCD)文件<sup>[4]</sup>描述了单个设备的物理信息配置。在实际工程中,分为2个阶段。

一是设备制造阶段生成IPCD,二是系统集成阶段通过配置工具导入IPCD,配置全站SPCD。IPCD

收稿日期:2022-06-13;修回日期:2022-08-27

基金项目:国家重点研发计划资助项目(2020YFB1711902);四川省科技计划资助项目(2020SYSY0016)

和 SPCD 文件结构如图 1 所示。

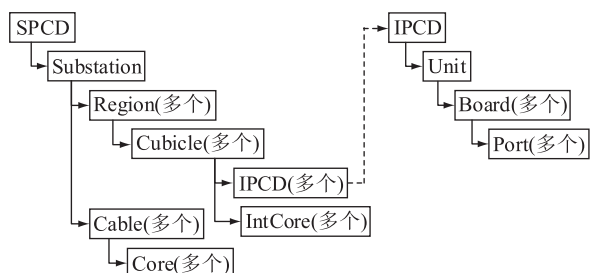


图 1 IPCD 和 SPCD 文件结构

Fig.1 IPCD and SPCD file structure

经过分析,无法从图纸中直接获取 SPCD 所需的全部信息,只能尽量直观和完整地获取表 1 的具体属性信息,研究物理回路建模算法,设计 IMCD 文件。

表 1 IMCD 文件元素属性

Table 1 IMCD file element attribute

元素	属性
Cubicle	name, desc
Unit	name, desc, iedName, manufacturer, type, class
Board	slot, desc, type
Port	no, desc, direction, plug, usage
IntCore/Core	name, portA, portB, type

变电站中,装置单元 (Unit) 的设置名称 (name) 和设备型号 (type) 从图纸中获取,设备文本描述 (desc)、设备类型 (class) 和设备厂商 (manufacturer) 须手动填写,desc 和 class 是对应的枚举类型。每个 Unit 至少包含一个板卡 (Board), Board 的编号 (slot) 和文本描述 (desc) 从图纸中获取,板卡型号 (type) 有则提取无则缺省。每个 Board 上有多个端口 (Port),端口组序号 (no) 对于同组端口 (如一对收发端口) 采用 A—Z 标明,其他属性从图纸中获取。连线是指任意设备间直接连接的一段光纤,可能是柜内纤芯 (IntCore) 也可能是芯线 (Core),屏柜 (Cubicle) 内部的是 IntCore, Cubicle 之间的是物理线缆 (Cable)。物理回路是指智能设备间从发送端口 (Tx) 到接收端口 (Rx) 的一个完整通道,是由连线和端口组成。单张图纸可能只标识收发的连接端口或装置,无法生成完整规范的回路信息,也无法区分 IntCore 和 Cable,因此统一只解析图纸本身的连接信息,包括该端口、线缆信息、用途描述等属性,有则获取就近文本信息,无则缺省。文本信息采用正则表达式根据字段标识格式解析字符串,获取具体的属性信息。

物理回路需要屏柜、装置、端口 (端子)、线缆、连接设备和标注文本等信息。这类变电站图像一

般是由多个矩形组合描述设备区域,由直线描述连线关系。根据图纸清册对图像进行分类,不通用实际生产过程中具有差异性的各类图像<sup>[13]</sup>。同一类图纸,甚至一张图纸就包含不同特征的变电站对象,因此应根据特征对变电站图像进行分类。其中电气符号是最重要的特征,文中识别组成电气符号的图形图元,根据图元拓扑构建连接关系字符串,再通过字符串相似度匹配算法模糊匹配,则电气符号识别问题转化为字符串匹配问题。

## 2 物理回路建模

### 2.1 图元信息提取

PDF 的逻辑或语义结构无法直接读取,须经过复杂的转化才能获得内容,而且难以保障其准确性<sup>[14]</sup>,故将 PDF 转化成便携式网络图形 (portable network graphics, PNG),并基于图像提取图元信息。图像分辨率对图像的检测效果不同,一般放大 4 倍或 6 倍。检测图元时,先进行图像二值化、降噪等预处理,再使用 OpenCV 软件库的核矩阵 (Kernel)、霍夫变换和形态学等算法检测图元<sup>[8]</sup>。

构成变电站图像的基本图元是直线,检测直线时,文中不同于霍夫变换,须对直线检测算法进行优化。多次实验选取合适的尺寸阈值 (Scale),根据二值图的行或列,与 Scale 进行整除操作,自适应获取 Kernel,设置垂直或水平方向上的膨胀腐蚀迭代次数 (Iterations),一般为 1 至 4,使用开闭运算检测直线。通过轮廓线检测算法依次提取直线、圆 (圆弧) 的图形图元坐标和尺寸信息。通过二进制位操作对原始图像进行减法运算,去除图形图元。为了提高泛化性,剩下可能存在的图像和文本信息一并进行处理<sup>[14]</sup>。通过阈值灰度投影算法分割字符并采用光学字符识别 (optical character recognition, OCR) 模型提取文本信息,对无关且细小的图像起到很好的过滤作用。PDF 图元信息提取流程如图 2 所示。

直线是变电站图像最广泛的图元,为了消除缩放等变换对特征提取产生的影响,规定区域左上角顶点为坐标原点,采用基于直线线段长度加权的主成分分析法进行比例归一化处理<sup>[13]</sup>。统一处理成图元,得到标题栏  $I(x_i, y_i, w_i, h_i)$ 、直线  $L(x_i, y_i, w_i, h_i)$ 、圆 (圆弧)  $O(x_i, y_i, r_i)$  和文本  $T(x_i, y_i, w_i, h_i)$ ,用于表征图元的类型及其在图纸中的位置,作为变电站图像分类和物理回路建模的基础。其中,  $x_i, y_i$  为坐标;  $w_i$  为矩形的宽;  $h_i$  为矩形的高;  $r_i$  为圆的半径;  $i = 1, 2, \dots, n$ ,  $n$  为图纸上的图元个数。

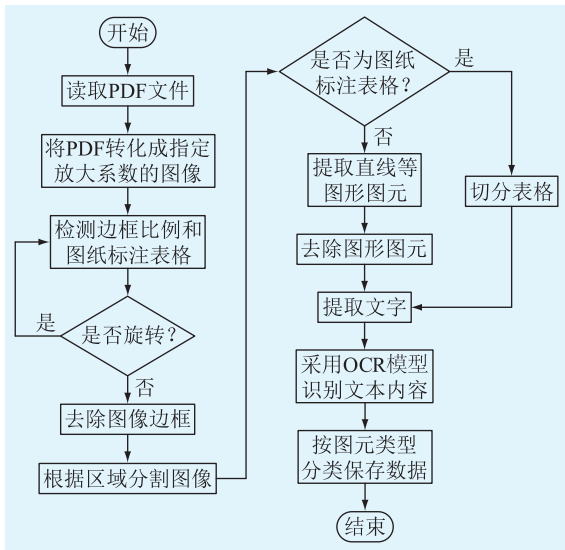


图2 PDF图元信息提取流程

Fig.2 PDF primitive information extraction process

## 2.2 电气符号识别

物理回路的电气符号由  $O$  和  $L$  组成的短直线  $L_s$  组合而成, 其中  $L_s$  的长度由  $O$  的直径比例阈值设定, 须选取最优区分度的阈值。为了解决组成电气符号图元和变电站图像图元的冲突, 保证电气符号的唯一性, 根据  $O$  和  $L$  定义电气符号最小的几何图元集合  $S$ , 基本图元要素包含短直线  $L_s$  构成的矩形以及不能构成矩形的独立短垂直线  $L_{sv}$ 、短水平线  $L_{sh}$ 、斜线  $L_{sb}$ 、圆(圆弧)  $O$ 。

集合  $S$  通过包含、相交和分离 3 种位置关系构成电气符号, 节点表示图元类型, 边表示图元位置关系<sup>[7]</sup>。根据拓扑构建拓扑连接关系字符串, 首先从图元坐标最小的节点起始, 由小到大, 先判断相交, 再判断包含, 最后判断分离, 以两图元之间的最邻近欧式距离是否为 0 (相交) 或比较指定的阈值 (包含或分离) 对集合  $S$  进行遍历。每个图元都要遍历判断与之前图元的位置关系, 直到完成相关所有图元的遍历<sup>[15]</sup>, 得到电气符号对应的拓扑关系字符串。最后根据组成电气符号图元信息, 得到电气符号矩形图元集合  $J(x_i, y_i, w_i, h_i)$ 。

常见端口的电气符号如图 3 所示, 节点 1 表示矩形, 节点 2 表示圆, 边 a 表示包含, 边 b 表示分离。则图 3 中端口符号的拓扑连接关系字符串分别是 1a2、2a1、1a2b2a1。由于存在图元提取误差和设计差异, 将常用的物理回路电气符号和多个拓扑字符串构造一对多的典型映射关系, 通过字符串相似度匹配结合 KR 的贪心串覆盖 (running Karp-Rabin greedy string tiling, RKR-GST) 算法<sup>[3]</sup>, 求出相似度最大的拓扑字符串对应的电气符号。

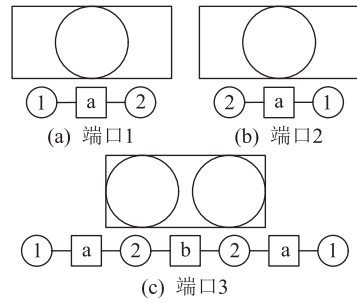


图3 电气符号拓扑字符串

Fig.3 Electrical symbol topology string

RKR-GST 算法结合了 GST 算法与 RKR 算法的优点, 模式串和比较串中每个元素不需要一一比较, 只在子串散列值相同时才进行比较和模式匹配, 效率较高。具体流程是: 首先指定搜索长度  $s$ , 并根据搜索长度  $s$  分别分割出模式字符串  $Q$  与比较字符串  $P$  的子字符串, 使用散列函数分别计算子字符串的散列值并进行比较, 如果相同则认为 2 个子字符串匹配。然后对 2 个子字符串后面的字符串进行贪婪匹配, 如果相同则继续匹配, 直到不能匹配为止, 同时记录匹配长度、 $Q$  中开始匹配的位置  $q$ 、 $P$  中开始匹配的位置  $p$ , 直到搜索长度  $s$  的子字符串全部匹配完毕。最后, 计算匹配子字符串长度之和的 2 倍除以 2 个字符串长度之和得到相似度。

## 2.3 变电站图像分类

使用统计分布方法提取具有区分能力的特征描述符, 能够较好地表达基于图元的变电站图像特点。主要提取图形图元特征向量  $F_e$  和电气符号特征向量  $F_j$ , 用于合成图像特征向量  $F$ 。

$$F = W_e F_e + W_j F_j \quad (1)$$

式中:  $W_e$ 、 $W_j$  为加权系数。

提取图元的个数和长度 2 种特征来合成式 (1) 中的图形图元特征  $F_e$ 。统计  $L$  在垂直和水平方向上的个数和累加长度, 统计  $O$  的个数和圆周累加长度, 对个数和长度做数据归一化处理, 结果值映射到  $[0, 1]$  区间, 获得个数特征向量  $F_n$  与长度特征向量  $F_l$ 。

$$F_e = W_n F_n + W_l F_l \quad (2)$$

式中:  $W_n$ 、 $W_l$  为加权系数。

将电气符号矩形图元集合  $J$  分为  $M$  类, 每类的个数为  $F_m$ ,  $m = 1, 2, \dots, M$ 。统计各类电气符号的个数, 进行数据归一化处理, 结果值映射到  $[0, 1]$  区间, 获得  $m$  维的电气符号特征向量  $F_j$ 。

$$F_j = W_m \times F_m \quad (3)$$

式中:  $W_m$  为  $W_m$  构成的加权系数向量;  $F_m$  为  $F_m$  构成的向量。

对变电站图像进行分类,得到构成物理回路的变电站图像,根据式(1)的图像特征训练分类模型。逻辑回归(logistics regression, LR)算法要求选取的训练特征各个维度独立,容易导致结果欠拟合,不具有特征组合的能力,实际应用中,变电站图像的多数特征之间具有相关性。梯度提升决策树(gradient boosting decision tree, GBDT)算法训练的特征能够有效解决 LR 特征要求高的问题。故文中提出一种基于 GBDT-LR 耦合训练的二分类算法。

训练具体步骤如下:(1) 将原始数据分为 2 个数据集,数据集 1 的训练集采用 GBDT 进行训练,构造  $K$  个决策树,得到分类器。(2) 使用训练好的 GBDT 分类模型对测试集进行预测,以预测值构造新特征,并进行独热编码。(3) 得到新的特征向量和数据集 2 的训练集,并提供给 LR 进行训练。(4) 得到 GBDT-LR 分类模型,对测试集进行二分类并输出结果。GBDT-LR 算法训练过程如图 4 所示。

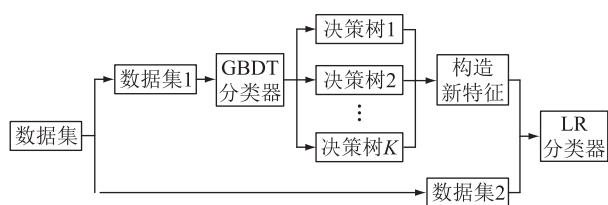


图 4 GBDT-LR 算法训练过程

Fig.4 GBDT-LR algorithm training process

### 2.4 拓扑关系构建

基于 2.1 节和 2.2 节的分析,设计物理回路建模方法,获取设备及其连接关系的数字化描述,解析成单张图纸对应的 IMCD。物理回路建模流程如图 5 所示。

先通过矩形区域获取设备信息:对直线进行分类和排序,提取成对的直线,按照逆时针方向,遍历最大长度的垂直线和水平线,可以找到若干个封闭连通矩形区域,即设备(一般是 Unit)。再以矩形区域左上角为坐标原点,坐标从小到大,依次遍历该设备区域内的直线及电气符号(一般是 Port),定位到电气符号所在的设备(一般是 Board)。依次就近匹配文本信息。然后基于图论深度优先搜索(depth first search, DFS)算法获取连接关系信息<sup>[16-17]</sup>:文中将 Port 或终端设备作为节点,连线作为边,用邻接表构建拓扑。模型的连接关系是长直线,与短直线没有交集,互为补集,不存在斜线。定位矩形区域中 Port 为起点,进行 DFS 遍历,以最邻近欧式距离寻找拓扑路径,依次遍历与之相连的长直线(去除设备的直线),并匹配节点与边的文本标注,路径终点是 Port 或标注终端设备信息的直线。重复上述

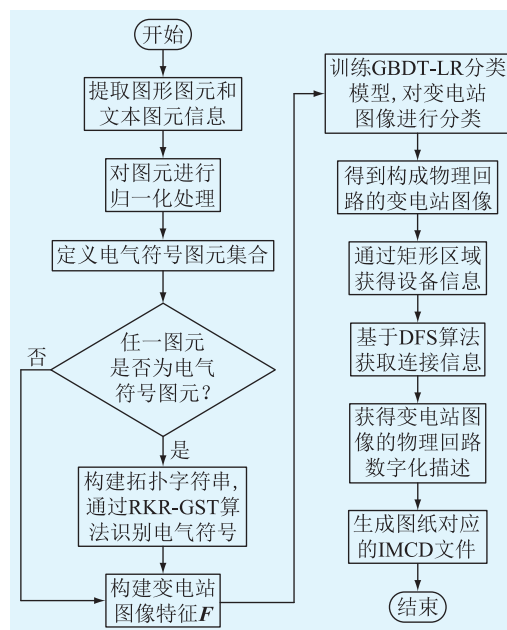


图 5 物理回路建模流程

Fig.5 Physical circuit modeling flow

过程,直到遍历完毕所有节点。

## 3 实验与应用

实验开发语言为 Python,从以下 4 个方面验证文中变电站图纸物理回路建模方法的有效性。

### 3.1 图元信息识别实验

统计提取信息正确数量与样本总数的比值即正确率为识别效果的评价标准。使用 Python 的 fitz 工具将 PDF 转化成 PNG,根据文中算法提取图形图元。OCR 模型面向印刷字体的效果比较好<sup>[8]</sup>,可以使用 Tesseract 等开源 OCR 模型提取图像文字图元信息,文中采用百度 OCR 模型。以 10 张图纸提取效果为例,PNG 图元识别效果如表 2 所示。霍夫变换不适用于对短直线、圆形、文本进行图元识别。基于图像识别图元信息的正确率随放大倍数的增大而提高,对短直线、圆形、文本进行图元识别,采用文中方法时放大 4 倍的平均正确率为 76.85%,放大 6 倍的平均正确率为 85.68%,但图像处理算法的耗时也会随放大倍数的增大呈指数增加。直线识别效果很好,采用文中方法时正确率为 98.41%,采用霍夫变换时正确率为 81.75%,故文中直线检测方法优于传统的霍夫变换直线检测算法,基本不存在断裂和漏检的情况。而其他图元存在误差是因为基于图像识别的效果不仅取决于检测算法和 OCR 模型,还取决于提取范围分辨率和是否有图元干扰。当存在少量文本和图形图元重叠时,误差是无法避免的。故须根据单张原始图纸,对其生成的 IMCD 进行人工校核。

表 2 PNG 图元识别效果

Table 2 PNG image element recognition effect

图元分类	全部数量	正确识别数量		
		放大 4 倍		放大 6 倍
		霍夫变换	文中方法	文中方法
直线	252	206	248	252
短直线	224		168	193
圆形	69		56	60
文本	126		98	106

对物理回路的电气符号进行识别。针对某省 220 kV 变电站共 140 张 PDF 物理回路图纸,将图像放大 4 倍,制作 1 680 张电气符号图像作为数据集,训练集和测试集的比例为 9:1。训练模型实验环境使用 Intel(R) Core(TM) i7-9700K 处理器、NVIDIA GeForce RTX 2080 SUPER 显卡,在 Windows 10 系统下进行实验。与文献[9]和文献[11]方法进行对比,电气符号识别准确率如表 3 所示。制图依据的具体规则以及制图者的习惯均存在差异,导致电气符号规范性差,较难统一。使用基于深度学习的目标检测算法<sup>[11]</sup>识别电气符号,必然存在图像尺寸、数据集大小、训练程度和识别场景等各种限制和误差。文献[11]中基于改进 YOLOv3 神经网络模型的电气符号识别准确率仅有 76.56%。文中将 PDF 转化为 PNG,再基于图像识别图元,必然存在误差,使用文献[9]中的矩阵匹配算法会造成误差进一步增大,准确率仅有 90.62%。文中使用 RKR-GST 算法对拓扑连接关系字符串进行模糊匹配,得到最大相似度对应的电气符号,在图元存在误差的情况下,识别正确率达到 93.75%,故文中方法优于文献[9]和文献[11]的方法,且时间复杂度和空间复杂度也优于文献[9]和文献[11]的方法,有利于后续变电站图像的自动化分类和物理回路建模。

表 3 不同方法电气符号识别准确率

Table 3 Electrical symbol recognition accuracy of different methods %

方法	电气符号识别准确率
YOLOv3	76.56
文献[11]方法	81.25
文献[9]方法	90.62
文中方法	93.75

### 3.2 变电站图像分类实验

统计 140 张 PDF 物理回路图纸,共 463 个变电站图像。实验中,基于式(1)一式(3),设置  $W_e = 0.4, W_j = 0.6, W_n = W_1 = 0.5, M = 10, W_1 = W_2 = \dots = W_{10} = 0.2$ ,得到式(1)的变电站图像特征向量  $F$ 。根据变

电站图像特征构造的分类模型以及二分类输出的结果,使用正确率、精准率、召回率和  $F_1$  值为分类效果标准<sup>[8]</sup>,对比 LR 和 GBDT-LR 的分类效果,如表 4 所示。结果表明,文中所提 GBDT-LR 耦合训练模型的分类性能优于 LR,可以更好地对变电站图像进行分类,省去手工选择变电站图像的操作。

表 4 分类模型实验对比

Table 4 Experimental comparison of classification models

模型	正确率/%	精准率/%	召回率/%	$F_1$ 值
LR	96.77	97.43	99.06	0.982 4
GBDT-LR	98.30	98.58	99.64	0.991 1

### 3.3 物理回路建模实验

以图 6 所示典型变电站图像为例,基于文中变电站图纸物理回路建模方法,构建物理回路模型的数字化描述如图 7 所示。

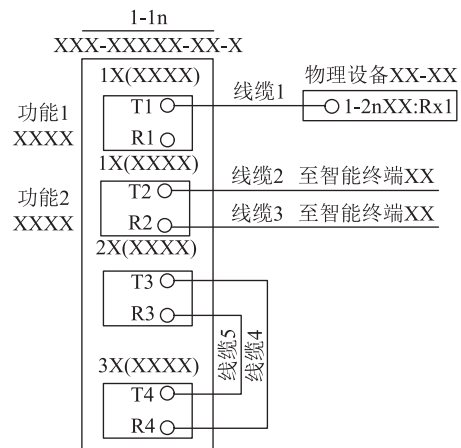


图 6 变电站图像

Fig.6 Substation image

```

-<Cubicle name="220kVXX" desc="220kVXX">
  -<Unit name="1-1N" desc="XX" type="XX...-X"
    manufacturer="XX" iedName="XX" class="XX">
    -<Board desc="(XX)" type="None"
      slot="1X">
      <Port desc="T1" usage="功能1XX"
        plug="LC" no="A" line_num="线缆1"
        direction="Tx"/>
      <Port desc="R1" usage="功能1XX"
        plug="None" no="A"
        line_num="None" direction="Rx"/>
      .....
    </Board>
  </Unit>
  <IntCore name="TX001" type="Tx"
    PortB="1-2nXX:Rx1" PortA="1X_T1"/>
  <IntCore name="TX002" type="Tx" PortB="至
    智能终端XX" PortA="1X_T2"/>
  .....
</Cubicle>
    
```

图 7 物理回路模型数字化描述

Fig.7 Digital description of physical circuit model

获取变电站图纸构建物理回路模型所需的关键信息,包括二次设备及对应物理线缆的个数与文本。以 10 张图纸解析效果为例,物理回路信息识别效果如表 5 所示。由表 5 可知,物理回路信息识别正确率均超过 90%。设备和线缆的对象识别正确率较高,均超过 97%,可以很好地构建物理回路模型。设备和线缆的文本信息识别误差较大,识别正确率达到 90%,应对物理回路模型的数字化描述进行人工校核。

表 5 物理回路信息识别效果

Table 5 Physical loop information identification effect

物理回路	全部数量	正确识别数量	正确率/%
物理设备对象	96	95	98.95
物理线缆对象	148	144	97.30
物理设备文本	97	88	90.72
物理线缆文本	148	135	91.22

根据物理回路模型数字化描述标准<sup>[4]</sup>,生成单张图纸对应的 IMCD 文件。IMCD 文件类似于 IPCD 文件,又包含图纸中关于 SPCD 文件的具体信息,是介于 IPCD 和 SPCD 之间的配置描述文件,相当于构建了变电站原始图纸到配置文件的桥梁,完成从变电站图形图像到数字化物理回路模型的转化。IMCD 文件既可以和 PDF 格式的变电站图纸一样用于资料的保存和传播,又可以和 XML 格式的配置文件的解析和应用,有利于智能变电站数字化技术的发展和二次应用的开发。最后再通过人工校核和配置工具进一步完成 SPCD 文件,符合电力行业现在对智能变电站配置方面的研究及应用需求。

### 3.4 辅助设计的实现与应用

为检验文中建模的应用效果,将文中模型集成到配置工具中,SPCD 辅助设计流程如图 8 所示。

变电站图纸主要识别包括设计院名称、工程名称、图纸名称、图号等标题栏信息,设备、端口、连线等变电站图像信息以及设备厂商等第三方信息。根据这些关键信息解析成单张图纸对应的 IMCD 文件,即对变电站图纸物理回路模型进行数字化描述。文件可能存在误差,尤其是变电站图像的文本信息,须进一步人工校核,即对照变电站图纸图像和 IMCD 文件进行关键信息的检查核实,可以直接在文件上修改。二次设备的关联可能涉及多张图纸,遍历每张图纸的 IMCD 文件,匹配设备唯一字段标识<sup>[18-19]</sup>,最后通过配置工具完成 SPCD 辅助设计。

以 10 张图纸为例,使用传统手工<sup>[20]</sup>和文中方法提取图纸信息,完成物理回路的建模,并记录耗

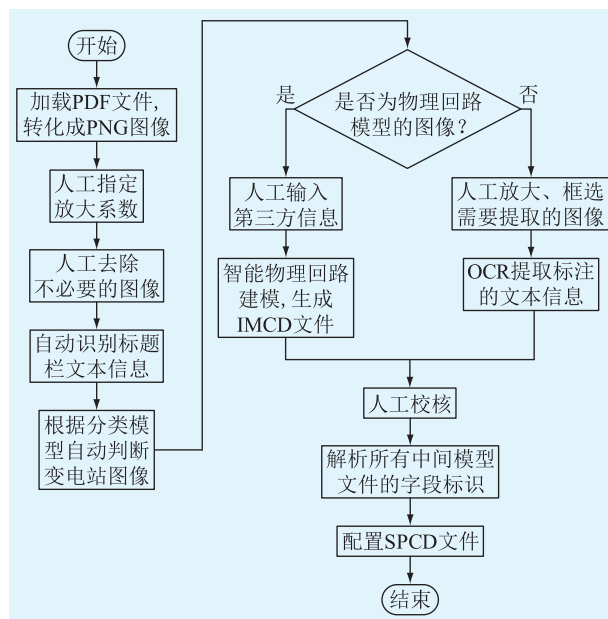


图 8 SPCD 辅助设计流程

Fig.8 SPCD auxiliary design process

时,如表 6 所示。

表 6 传统人工和文中方法耗时对比

Table 6 Time consuming comparison between traditional manual work and the proposed method

图纸序号	变电站图像数量	耗时/min		
		传统手工	文中方法	
			算法处理	人工校核
1	2	13.2	1.0	3.0
2	1	4.5	0.5	1.0
3	9	32.0	5.0	15.0
4	1	8.6	1.0	4.0
5	10	20.6	5.0	5.0
6	2	14.0	1.0	4.0
7	3	7.3	1.0	3.0
8	2	15.3	1.0	4.0
9	2	11.2	1.5	2.0
10	3	15.6	2.0	6.0
合计	35	142.3	19.0	47.0

表 6 中统计文中方法的总耗时包括算法处理耗时和处理后的人工校核耗时,总耗时为传统手工方法的 46.38%。文中方法的算法处理耗时较少,人工校核耗时较多,这取决于变电站图像的复杂程度,对复杂且密集的图像,由于错误较多,人工校核的耗时较长,对整齐且规律的图像,人工校核的耗时较短,效率更高。经过文中方法对图纸的关键信息智能提取、解析和应用后,人员进行校核的关注点主要为图纸中文字信息比较模糊且关键的地方,不用完整读取全部的线路图,可以大幅节省人工核对时间。

## 4 结语

文中针对智能变电站配置过程中,人工读图工作量庞大、效率低下等问题,综合电气符号识别、变电站图像分类和拓扑信息解析三方面研究,提出PDF图纸物理回路建模方法,并应用到变电站原始图纸到变电站SPCD文件的转化过程中。实践证明,文中方法大幅节省人工核对时间,有效解决了现阶段原始图纸数字化薄弱的问题,为变电站二次高级应用业务的开展奠定了基础。在后续工作中,应进一步优化图元提取算法,提高图元识别正确率,完善物理回路模型的构建。

### 致 谢

本文得到福建永福数字能源技术有限公司科技项目(R110319H01006)资助,谨此致谢!

### 参考文献:

- [1] AZAR R. Substations: transformations and improvements[J]. IEEE Power and Energy Magazine, 2019, 17(4): 108-105.
- [2] 李磊,杨毅,苏麟,等. 智能变电站二次系统物理回路建模方法及设计实现[J]. 电力系统保护与控制, 2016, 44(24): 130-139.  
GAO Lei, YANG Yi, SU Lin, et al. A modeling approach and design implementation of secondary system physical circuit for smart substation[J]. Power System Protection and Control, 2016, 44(24): 130-139.
- [3] 李劲松,童晓阳,张炜,等. 大数据视角下相近SCD文件的差异分析方法[J]. 电力系统自动化, 2021, 45(4): 164-171.  
LI Jinsong, TONG Xiaoyang, ZHANG Wei, et al. Analysis method for difference of similar SCD files from the perspective of big data[J]. Automation of Electric Power Systems, 2021, 45(4): 164-171.
- [4] 智能变电站光纤回路建模及编码技术规范:GB/T 37755—2019[S]. 北京:中国标准出版社, 2019.  
Technical specifications for modeling and coding fiber loop in smart substation: GB/T 37755-2019[S]. Beijing: Standards Press of China, 2019.
- [5] 刘千宽,刘宏君,丁晓兵,等. 服务于变电站数字化设计的二次设备建模技术研究[J]. 电力系统保护与控制, 2021, 49(2): 166-172.  
LIU Qiankuan, LIU Hongjun, DING Xiaobing, et al. Research on secondary equipment modeling technology for substation digital design[J]. Power System Protection and Control, 2021, 49(2): 166-172.
- [6] 张琪,叶颖. 基于对象图例及其拓扑关系识别的二维工程CAD图纸矢量化方法[J]. 计算机与现代化, 2018(11): 40-45.  
ZHANG Qi, YE Ying. 2D engineering CAD drawings vectorization method based on recognition of object legends and their topology[J]. Computer and Modernization, 2018(11): 40-45.
- [7] 范帆,关侗红. 工程图纸字符串及标注信息提取[J]. 计算机工程与应用, 2012, 48(7): 161-164.  
FAN Fan, GUAN Jihong. Recognition of characters and annotations on engineering drawings[J]. Computer Engineering and Applications, 2012, 48(7): 161-164.
- [8] 宋小鸿. 工程图纸的关键信息智能提取技术[D]. 北京:北京交通大学, 2019.  
SONG Xiaohong. Intelligent extraction technology of key information in engineering drawings[D]. Beijing: Beijing Jiaotong University, 2019.
- [9] 刘清泉,陈天英,郭少飞,等. 基于图谱理论的变电站二次系统图纸活化技术研究与应用实践[J]. 电测与仪表, 2021, 58(7): 97-103.  
LIU Qingquan, CHEN Tianying, GUO Shaofei, et al. Research and application of drawing activation technology of substation secondary system based on graph theory[J]. Electrical Measurement & Instrumentation, 2021, 58(7): 97-103.
- [10] 朱江,孙家广,邹北骥,等. 电气原理图的自动识别[J]. 计算机工程与科学, 2007, 29(1): 56-58, 69.  
ZHU Jiang, SUN Jiaguang, ZOU Beiji, et al. Automated recognition of electrical circuit drawings[J]. Computer Engineering & Science, 2007, 29(1): 56-58, 69.
- [11] 江再玉,石文娟,马晶,等. 基于深度神经网络的电气元件符号识别算法[J]. 电力系统及其自动化学报, 2022, 34(2): 48-55.  
JIANG Zaiyu, SHI Wenjuan, MA Jing, et al. Recognition algorithm for electrical component symbols based on deep neural network[J]. Proceedings of the CSU-EPSA, 2022, 34(2): 48-55.
- [12] 郝晓光,赵宇皓,尹项根,等. 基于中间模型文件的智能变电站虚回路校验研究[J]. 电力科学与技术学报, 2020, 35(5): 132-137.  
HAO Xiaoguang, ZHAO Yuhao, YIN Xianggen, et al. Intelligent substation virtual circuit check based on the intermediate model file[J]. Journal of Electric Power Science and Technology, 2020, 35(5): 132-137.
- [13] 曾凡智,周燕,余家豪,等. 基于无监督学习的二维工程CAD模型端到端检索算法[J]. 计算机科学, 2019, 46(12): 298-305.  
ZENG Fanzhi, ZHOU Yan, YU Jiahao, et al. End-to-end retrieval algorithm of two-dimensional engineering CAD model based on unsupervised learning[J]. Computer Science, 2019, 46(12): 298-305.
- [14] 于丰畅,陆伟. 基于机器视觉的PDF学术文献结构识别[J]. 情报学报, 2019, 38(4): 384-390.  
YU Fengchang, LU Wei. Structural recognition of PDF academic literature based on computer vision[J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(4): 384-390.
- [15] 李铁成,任江波,曾四鸣,等. 基于CAD图自动拾取的变电站电气二次回路连接关系的方法,系统及介质:CN110795809A[P]. 2020-02-14.

- LI Tiecheng, REN Jiangbo, ZENG Siming, et al. Transformer substation electrical secondary circuit connection relation method and system based on CAD drawing automatic pickup and medium; CN110795809A [P]. 2020-02-14.
- [16] 高旭,于庆广,马迎新,等. 基于图论深度遍历算法的智能变电站光纤虚实回路对应方法研究[J]. 电测与仪表, 2020,57(2):1-6.  
GAOXu, YU Qingguang, MA Yingxin, et al. Research on mapping method between logical circuit and physical circuit of optical fiber loop in smart substation based on deep traversal algorithm of graph theory[J]. Electrical Measurement & Instrumentation, 2020,57(2):1-6.
- [17] 董怀普,王位杰,刘文彪,等. 基于深度搜索的二次虚实回路融合故障诊断技术[J]. 电力自动化设备, 2019,39(5):219-223.  
DONG Huaipu, WANG Weijie, LIU Wenbiao, et al. Fault diagnosis technology of secondary virtual loop and physical fiber loop fusion based on depth first search[J]. Electric Power Automation Equipment, 2019,39(5):219-223.
- [18] 李耕,邱智勇,饶丹,等. 基于SPCD的智能变电站虚实链路映射方案研究[J]. 电力工程技术, 2020,39(1):124-129.  
LI Geng, QIU Zhiyong, RAO Dan, et al. Scheme for mapping between virtual and real link in smart substation based on SPCD [J]. Electric Power Engineering Technology, 2020, 39(1):124-129.
- [19] 戴志辉,鲁浩,刘媛,等. 基于改进 D-S 证据理论的智能站保护二次回路故障诊断方法[J]. 电力系统保护与控制, 2020,48(9):59-67.  
DAI Zhihui, LU Hao, LIU Yuan, et al. A fault diagnosis method for the secondary circuits of protection systems in smart substations based on improved D-S evidence theory [J]. Power System Protection and Control, 2020,48(9):59-67.
- [20] 黄子雄,黄树贤. 通信线路施工图自动分幅方法与实现[J]. 计算机应用与软件, 2018,35(9):323-326,333.  
HUANG Zixiong, HUANG Shuxian. Method and implementation of automatic division of communication line construction drawing [J]. Computer Applications and Software, 2018, 35(9):323-326,333.

作者简介:



侯进(1969),女,博士,副教授,研究方向为电力系统自动化、人工智能在电网中的应用 (E-mail:jhou@swjtu.edu.cn);

王祥宇(1995),男,硕士在读,研究方向为计算机视觉、图像处理技术在电力系统中的应用;

郝彦超(1996),男,硕士在读,研究方向为人工智能在电力系统中的应用。

**Realization of physical circuit modeling method for substation drawings**

HOU Jin<sup>1,2</sup>, WANG Xiangyu<sup>1,2,3</sup>, HAO Yanchao<sup>1,2,3</sup>, WEN Zhilong<sup>1,2,3</sup>, LI Yunzhe<sup>1,2,3</sup>, KANG Pingping<sup>1,2</sup>

(1. Laboratory of Intelligent Perception and Smart Operation & Maintenance, School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China; 2. National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Southwest Jiaotong University, Chengdu 611756, China; 3. School of Tangshan, Southwest Jiaotong University, Tangshan 063000, China)

**Abstract:** At present, the original drawing information of the substation is stored in a single way, and the degree of digital analysis is not high. In order to reduce the huge workload of manually analyzing the original drawing of the substation, a method is proposed to intelligently extract the information of the equipment and connection relationship of the substation based on the portable document format (PDF) drawing for building a physical circuit model, thus implementing the method according to the smart substation physical configuration description (SPCD) file. Firstly, the primitive information is extracted and processed from drawings. Then electrical symbol recognition is realized through string similarity matching running Karp-Rabin greedy string tiling (RKR-GST) algorithm, and substation images are classified by gradient boosting decision tree-logistic regression (GBDT-LR) hybrid algorithm based on features. Finally, the digital description from the original drawing to the physical circuit model is completed according to the SPCD file. Experiments show that the correct rate of electrical symbol matching is 93%, and the correct rate of physical circuit identification is more than 90% when there are errors in primitives.

**Keywords:** smart substation; two-dimensional drawing; configuration description file; physical circuit; image processing; drawing activation; auxiliary design

(编辑 吴楠)