

# 基于大数据的电力通信网络风险辨识与评估方法研究

罗云,高艳宏,王志强

(中通服创立信息科技有限责任公司,四川成都 610093)

**摘要:**为了高效、准确地识别网络庞大、结构复杂电力通信网中所潜在的风险点,从网络结构、承载业务等方面分析其对电力通信的影响,并结合电力事故事件调查规程的评定标准,客观且科学地评估其对电力通信引起的电力安全事件,为电力通信风险管控提供决策支撑。本文结合电力通信风险管控具体内容及特点,分析电力通信网事故发生原因,推理事故演变过程,提出解决思路和方案,并通过实际测试案例,验证电力事故评估模型具有较高的计算效率和准确性,为准确定位事故发生原因和评判风险等级提供科学依据。该模型结合复杂网络理论可靠性因素的抗毁性、生存性、有效性等特征,计算其相关特性的风险因子值;利用通径系数分析技术,验证风险因子的完备性;根据已标注的风险等级样例数据,构建深层卷积神经网络 CNN 模型,实时评测事故的风险等

**关键词:**电力通信网络;重要度评价;风险因子;通径系数分析;深层神经网络;大数据

**文章编号:**2096-4633(2019)11-0064-06 **中图分类号:**TM76 **文献标志码:**B

当今时代的快速发展和科学技术的日新月异,电力技术也在随之迅猛地进步,伴随而来的是在电力领域中越来越多样化<sup>[1]</sup>。电力通信网作为电网二次系统不可或缺的组成部分,其为电力系统提供调度、生产、经营和管理等服务,同时保障了电力系统安全稳定运行。随着电网智能化步伐不断加快,导致电力通信网络结构日趋复杂多变,而网络所承载的业务数量与日俱增,网络链路(设备)的中断会对用户产生各种风险,因此对电力通信网的节点重要性评估和风险评估至关重要<sup>[2]</sup>。

国内外也有很多学者提出了一些风险因子判别和风险评估模型。因子判别,例如主成分分析、基于粗糙集约简等理论<sup>[3]</sup>。风险评估模型,例如基于粗糙集与 DS 证据理论模型解决设备运营分析评估;基于直觉模糊熵权法模型解决主型设备安装风险评估;基于改进型 AHP 与证据理论解决应答器系统风险评估<sup>[4]</sup>;基于 BP 神经网络模型,解决交通网络风险评估。

电力通信网络相对于传统的网络来说,有其特有的行业特征,而且数据量也非常大,传统解决方案已经无法满足电力通信行业需求。大数据具有 5V 特征,即大量(Volume)、多样(Variety)、难辨识(Veracity)、快速(Velocity)和价值大密度低(Value)。针对这一问题,本文利用分布式的深度神经网络模型实现电力通信资源网络风险评估。

## 1 风险等级制定

### 1.1 风险等级标准

某公司电力事故事件调查规程的生产实时控制业务通信通道中断评定标准附如表 1。

### 1.2 风险等级因子

网络是节点和链路的集合,电力通信网亦是如此。根据电力通信网静态拓扑结构确定通信网对应的无向网络,确定网络中每个节点和边所承载业务信息(执行业务的类型、数量、重要性)。复杂网络可以用  $G(V, E)$  来表示,其中  $G$  表示一个无向连通图,  $V$  为复杂网络点的集合,  $E$  为边的集合。根据复杂网络评价标准所涉及抗毁性、生存性、有效性、同步性等特性,设计风险因子,分析因子如表 2。

电力通信网关键节点的辨识,对其保护,能够有效增强电力通信网抵抗攻击的能力,降低电网的运行风险,保障电力系统的安全稳定运行。计算节点静态重要度(网络凝聚度)和节点业务重要度(节点所承担的业务量种类以及数量)。采用三标度层次分析法确定两个重要度评价指标,辨识电力通信网关键节点。网络凝聚度定义如下<sup>[5]</sup>:

$$\partial[G] = \frac{1}{n \cdot l} = \frac{1}{n \cdot \sum_{i! = j \in N} d_{ij}} = \frac{n-1}{\sum_{i! = j \in N} d_{ij}} \quad (1)$$

表 1 电力生产安全事件等级划分标准

Tab. 1 Classification standards of power production safety incidents

电力生产安全事件 类别	电力生产安全事件等级			
	二级事件	三级事件	四级事件	五级事件
电力安全事件(影响调度业务和实时通信功能)	生产实时控制业务通信通道中断	省区级以上调度控制中心与直接调度范围内超过 30% 的厂站调度电话、调度自动化等通信业务全部中断	地区调度机构与调度对象的调度电话、调度自动化业务全部中断	县级调度机构通信业务全部中断
		500 kV 以上线路或厂站间生产实时控制业务通信通道全部中断 30 分钟以上	500 kV 以上线路或厂站间生产实时控制业务通信通道全部中断 30 分钟以下	通信电路非计划停运造成 220 kV 以上线路远方跳闸保护、远方切机、切负荷等装置由双通道改为单通道运行 24 小时以上
		220 kV 线路或厂站间生产实时控制业务通信通道全部中断 30 分钟以上	220 千伏线路或厂站间生产实时控制业务通信通道全部中断 30 分钟以下	110 kV 线路或厂站间生产实时控制业务通信通道全部中断 1 小时以上

表 2 风险因子

Tab. 2 Risk factors

因子类别名称	因子类别细节	因子名称
网络结构	固有特性	总边数
		总节点数
	计算特性	重要节点数目
		重要边数目
承载业务	业务数量	自动化业务数量
		继电保护业务数量
		安稳控制业务数量
	业务的特性	其他业务数量
		业务中断率
		业务连通度
		业务分离度
		业务跳数
	业务固有特性	资源利用率
		电压等级
影响业务	影响业务	中断业务数量
		影响业务数量
	影响时间	未影响业务数量
		中断业务时间
影响范围	网、省、地	

式(1)中,  $d_{ij}$  表示电力通信网节点  $i$  到节点  $j$  链路间加权路径的最短距离, 可以利用改进后的 Dijkstra 算法计算最短距离<sup>[6]</sup>。

资源利用率是指每条链路已经被使用的时隙长短占链路总时隙的百分比。网络的资源利用率越低, 资源的分配越均衡, 阻塞率越低, 网络传输的质量越高。

业务路由分离率很大程度地影响网络的抗毁能力。

业务中断率是利用 N - X 分析确定受中断影响的业务程度。

### 1.3 风险等级因子预处理

我们获得的数据中往往存在很多的无用信息, 甚至存在错误信息, 数据的健康程度对于算法结果的影响极大。数据预处理就是让那些冗余混乱的源数据变得能满足其应用要求。在进行特征选择之前, 一般会先进行数据无量纲化处理, 这样, 表征不同属性的各特征之间才有可比性。本文采用第三方模块 sklearn 的 preprocessing 分支 StandardScaler 函数, 作为数据标准化的方式, 消除不同特征之间差异, 利用 inverse\_transform 函数恢复特征数据的原貌。

### 1.4 风险等级因子评价

路径系数分析对自变量和因变量之间相关性进

行分解,来研究自变量对因变量的直接重要性和间接重要性,据此挑选出重要指标<sup>[7-8]</sup>。

通路系数以  $p_{ij}$  表示,定义式为:

$$p_{ij} = b_i \frac{\delta_{x_i}}{\delta_y} \quad i = 1, 2, \dots, p \quad (2)$$

式(2)中: $b_i$  偏回归系数,  $\delta_{x_i}$   $x$  的标准差,  $\delta_y$   $y$  的标准差。

通过计算,  $b_{y_1} = 0.041 < 0.05$ , 可知电力通信资源网风险等级相关的因数均已考虑在内。

## 2 风险等级评价

本文采用 tensorflow 实现卷积神经网络。TensorFlow 是 Google 深度学习框架,用于数值计算的开源软件库。

### 2.1 卷积神经网络架构

卷积神经网络(CNN)在深度学习领域起着至关重要的作用。CNN 的基本结构由输入层、卷积层、取样层、全连接层及输出层构成<sup>[9]</sup>。由于各层间的神经元的连接是非全连接的,而且同一层中某些神经元之间的连接的权重共享(即相同的),降低了网络模型的复杂度,减少了权值的数量,解决了传统神经网络权值过多计算量过大产生的性能问题。

#### 2.1.1 基本结构

借鉴借鉴了经典的 VGGNet<sup>[10]</sup>、GoogleNet 和 AtriaNet<sup>[11]</sup>的结构,构建卷积神经网络模型包含三个子卷积网络,每个子卷积网络都有单独的输入,三个子卷积网络最后在全连接层(fully connected layer, FC)进行融合,全连接层之后是 Softmax 输出层, CNN 结构如下图所示。

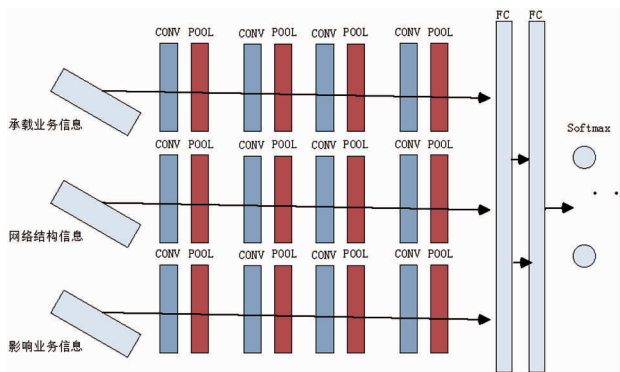


图 卷积神经网络结构

Fig. Convolutional neural network structure

各子卷积网络均包含四个卷积层(convolutional

layer, CONV)和四个池化层(pooling layer, POOL)来自动提取特征。每个特征面由多个神经元组成,每一个神经元通过卷积核与上一层特征面的局部区域连接。卷积层计算公式如下:

$$X(l) = f(Wl \otimes X(l-1) + b(l)) \quad (3)$$

其式(3)中  $X(l)$  和  $X(l-1)$  代表层  $l$  层和  $l-1$  层的神经元,  $Wl$  代表卷积核,  $b$  代表偏置。

卷积神经网络利用 Softmax 进行分类<sup>[12-14]</sup>, Softmax 函数估计输入  $x$  属于特定类别  $j \in k$  的概率:

$$P(y = j | x) = \frac{\exp(xTW_j)}{\sum_{k=1}^K \exp(xTW_k)} \quad (4)$$

Dropout 是 CNN 中防止过拟合的一个有效手段。在神经网络训练过程中,最后一个池化层按照一定的概率 0.25 丢弃网络神经元。

#### 2.1.2 卷积层

卷积层是卷积神经网络的核心部分,也被称为过滤器(filter)或者卷积核,即动物视觉中的感受野。卷积核用三维矩阵表示,矩阵的长度与宽度一般由人为指定,而深度需与前一层网络节点矩阵深度保持一致。卷积层的前向传播过程,过滤器滑动指定大小的窗口,与相应大小数据进行运算(加权求和),实现卷积层的矩阵计算。

一个卷积层可以有多个不同的卷积核,而每一个卷积核都对应一个滤波后的一种特征,通过增加卷积核的个数,可以提升卷积神经网络的表达能力。

卷积的好处是不管输入数据维度有多大,需要训练的权值数量仅与卷积核长宽、卷积核的数量有关,从而减少了网络中的参数数量。

在 TensorFlow 中函数 `tf.nn.conv2d()` 可以实现卷积网络前向传播算法,核的大小为  $3 \times 3$ , 填充方式为 SAME, 滑动窗口  $[1, 1, 1, 1]$ 。

#### 2.1.3 池化层

池化层能有效的缩小数据维度,减小最后全连接层中的参数数量,防止过拟合发生。池化层的前向传播过程,过滤器滑动指定大小的窗口,与相应大小数据进行运算(最大值或者平均值),实现池化层的矩阵计算。常用的池化层包括最大池化层和平均池化层。

池化层起到二次提取特征的作用,它的每个神

经元对局部接受域进行池化操作

在 TensorFlow 中函数 `tf.nn.max_pool` 实现了最大池化层, `tf.nn.avg_pool` 实现了平均池化层。本实验采用最大池化层方式,核的大小为  $2 \times 2$ ,滑动窗口 `[1,2,2,1]`,填充方式为 SAME

## 2.2 卷积神经网络优化

### 2.2.1 非线性的变换

激活函数是神经网络的重要组成部分,对输入数据进行非线性的变换,可以提升网络非线性的表达能力,使网络具备非线性分割的能力。

Sigmoid 函数导数值从 0 开始,其经历若干层传递后,其很快又趋于 0,容易出现梯度消失现象,使学习停滞。Relu 函数减少了参数相互依存关系,缓解过拟合问题发生的概率,容易学习优化。

本文将选用 Relu 激活函数和交叉熵代价函数进行试验去提高实验的准确率。Tensorflow API 提供 `tf.nn.relu` 函数,方便调用。

### 2.2.2 目标函数的优化

代价函数(cost function)是用来计算预测值与实际值的误差,用于找到最优解的目标函数<sup>[15-17]</sup>。

二次代价函数:

$$C = \frac{1}{2n} \sum_x ||y(x) - aL(x)||^2 \quad (5)$$

式(5)中  $w$  和  $b$  的梯度跟激活函数的梯度成正比,激活函数的梯度越大,  $w$  和  $b$  的大小调整得越快,训练收敛得就越快。初始的代价(误差)越大,训练越慢;而与我们的期望不符,错误越大,修正的幅度越大,学习越快,适合输出神经元是线性的函数。

交叉熵代价函数(cross-entropy):

$$C = -\frac{1}{n} \sum_x [y \ln a + (1-y) \ln(1-a)] \quad (6)$$

权重的学习的速度受  $\sigma(z) - y$ ,也就是输出中的误差的控制,更大的误差,更快地学习速度。这也是我们期望达到的效果,适合输出神经元是 S 型函数。

Tensorflow 框架 API 提供了 `softmax_cross_entropy_with_logits` 函数,其将交叉熵代价函数与 `softmax/sigmoid` 函数配合使用,实现分类预测功能。

### 2.2.3 计算梯度的方法

优化类提供了为损失函数计算梯度的方法,其中包含比较经典的优化算法,比如 GradientDescent

和 Adagrad。

计算梯度算法有两个重要的控制因子:一个是步长,由学习率控制;一个是方向,由梯度指定。常用的学习率包括固定值、可调节的自适应值。Tensorflow API 提供 GradientDescentOptimizer 旨在对所有步骤中的所有变量使用恒定的学习率。还提供现成的自适应优化器,包括 AdagradOptimizer 和 AdamOptimizer。Tensorflow 提供了一种自动将指数衰减应用于学习速率张量的操作:`tf.train.exponential_decay`。本文采用 `tf.train.AdamOptimizer` 方法计算损失函数梯度。

## 2.3 评价指标

预测正确方式包括:TP(true positive)代表被模型预测为正的样本、TN(true negative)代表被模型预测为负的样本;预测错误方式包括:FP(false positive)代表被模型预测为正的负样本, FN(false negative)代表被模型预测为负的正样本。

常用的评价指标 ROC(receiver operating characteristic) 和 AUC(area under curve)。AUC 是 ROC 曲线和轴围成的面积。ROC 曲线越左上倾斜, AUC 越大,说明模型的准确性越高。

## 3 实验验证

### 3.1 环境搭建

TensorFlow 包括单机模式和分布式模式。由于该实验计算量比较庞大,故利用 Tensorflow 所支持分布式计算能力,实现分布式的神经网络。Tensorflow 集群包括一个 ParamServer 节点(保存和更新参数)和两个 WorkerServer 节点(负责计算)。

在三个机器节点均安装如下软件:GCC v 6.1.0 libstdc++ +6.8.3.0-7\_i386、glibc-2.25.tar、Anaconda3-5.2.0、tensorflow-1.10.0-cp36。

### 3.2 实验设置

本实验数据包含 10 万条风险等级评价数据。首先对数据进行预处理,例如补齐丢失的数据、连续变量的归一化处理、将数据统一变换为无单位的数据集(消除量纲)、剔除噪声点数据。将预处理后的数据转化为矩阵和向量,作为输入数据,喂给 CNN。CNN 利用自身的网络结构,调整权重和偏置参数,建立自变量与因变量映射关系,预测风险等级。为了更好地对模型进行评估,将数据划分为训练集、验证集和测试集。



### 3.3 实验过程

由于每次训练后模型都会更新权重和偏值。在前期训练过程中,准确率有所提升,但经历若干次训练后,模型的准确率提高幅度几乎为零。

在模型训练的过程中,出现了过拟合的现象,通过利用 tensorflow 提供的函数 tf.nn.dropout、tf.contrib.layers.l2\_regularizer,防止过拟合的发生。

在若干次实验过程中,不断调整网络层次结构,通过增加和减少层次数量和神经元数目,替换不同激活函数和损失函数,提高训练和测试数据的准确率。

### 3.4 结果分析

本文从电力通信网络风险评估应用出发,提出利用分布式的深度神经网络模型实现电力通信资源网络风险辨识与评估方法。通过仿真表明,实验准确率为 83.5%,但仍有优化的空间。通过调整网络结构,进行训练和测试,能进一步提高正确识别率。

## 4 总结

目前深度学习已经被广泛应用于学术界和工业界。卷积神经网络在多种神经网络中表现的尤为突出,无论是在速率,还是在准确率方面。本实验利用神经网络优化理论,结合 tensorflow 框架提供的 api,验证该模型具有较高准确率,对构建更复杂的卷积神经网络具有一定的参考意义。

#### 参考文献:

- [1] 石际. 关于应用 OTN 技术对电力通信网的网络优化[J]. 电力大数据,2018,21(06):82-86.  
SHI Ji. The application of OTN technology to the network optimization of power communication network[J]. Power Systems and Big Data,2018,21(06):82-86.
- [2] 狄立,郑征,夏旻,等. 基于快速密度聚类的电力通信网节点重要性评估[J]. 电力系统保护与控制,2016,44(13):90-95.  
DI Li,ZHENG Zheng,XIA Min, et al. Node importance evaluation of electric power communication network based on fast density clustering[J]. Power System Protection and Control,2016,44(13):90-95.
- [3] 孙林,潘俊方,张霄雨,等. 一种基于邻域粗糙集的多标记专属特征选择方法[J]. 计算机科学,2018,45(01):173-178.  
SUN Lin,PAN Junfang,ZHANG Xiaoyu, et al. Multi-label-specific feature selection method based on neighborhood rough set[J]. Computer Science,2018,45(01):173-178.

- [4] 戴胜华,谢旭旭. 基于改进型 AHP 与证据理论的应答器系统风险评估[J]. 安全与环境学报,2019,19(01):55-61.  
DAI Shenghua,XIE Xuxu. Risk assessment of balise system based on the improved AHP and evidence theory[J]. Journal of Safety and Environment,2019,19(01):55-61.
- [5] 王汪兵,王先培,尤泽樟,等. 电力通信网关键节点辨识方法研究[J]. 电力系统保护与控制,2018,46(01):44-49.  
WANG Wangbing,WANG Xianpei,YOU Zezhang, et al. Research on key node identification method in electric power communication network[J]. Power System Protection and Control,2018,46(01):44-49.
- [6] 任伟建,左方晨,黄丽杰. 基于 GIS 的 Dijkstra 算法改进研究[J]. 控制工程,2018,25(02):188-191.  
REN Weijian,ZUO Fangchen,Huang Lijie. The improvement research of dijkstra algorithm based on GIS[J]. Control Engineering of China,2018,25(02):188-191.
- [7] 李进涛,刘彦陟,杨园园,等. 1985-2015 年京津冀地区城市建设用地时空演变特征及驱动因素研究[J]. 地理研究,2018,37(01):37-52.  
LI Jintao,LIU Yansui,YANG Yuanyuan, et al. Spatial-temporal characteristics and driving factors of urban construction land in Beijing-Tianjin-Hebei region during 1985-2015[J]. Geographical Research,2018,37(01):37-52.
- [8] 佚名. 新疆艾比湖流域潜在蒸散变化特征与成因分析[J]. 灌溉排水学报,2018,37(02):89-94.  
YI Ming. Variety characteristics and cause analysis of potential evapotranspiration in the Ebinur lake basin in Xingjiang[J]. Journal of Irrigation and Drainage,2018,37(02):89-94.
- [9] 周飞燕,金林鹏,董军. 卷积神经网络研究综述[J]. 计算机学报,2017,40(06):1229-1251.  
ZHOU Feiyan,JIN Linpeng,DONG Jun. Review of convolutional neural network[J]. Chinese Journal of Computers,2017,40(06):1229-1251.
- [10] XIANGYU ZHANG,JIANHUA ZOU,KAIMING HE, et al. Accelerating very deep convolutional networks for classification and detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2016,38(10):1943-1955.
- [11] ZHAOHAN XIONG,VADIM V. FEDOROV,JICHAO ZHAO, et al. Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imaging using a dual fully convolutional neural network[J]. IEEE Transactions on Medical Imaging,2019,38(02):515-524.
- [12] WEIWEI SHI,YIHONG GONG,XIAOYU TAO, et al. Fine-grained image classification using modified DCNNs trained by cascaded softmax and generalized large-margin losses[J]. IEEE Transactions on Neural Networks and Learning Systems,2019,30(03):683-694.
- [13] 傅隆生,冯亚利,Elkamil Tola,等. 基于卷积神经网络的田间多簇猕猴桃图像识别方法[J]. 农业工程学报,2018,34(02):205-211  
FU Longsheng,FENG Yali, ELKAMIL Tola, et al. Image

- recognition method of multi-cluster kiwifruit in field based on convolutional neural networks [J]. Transactions of the Chinese Society of Agricultural Engineering, 2018, 34(02): 205 - 211.
- [14] 董振江, 高燕, 吴文熙. 车辆属性识别及跨场景标注方案 [J]. 中兴通讯技术, 2017, 23(04): 24 - 28.  
DONG Zhenjiang, GAO Yan, WU Wenxi. A solution for vehicle attributes recognition and cross-scene annotation [J]. ZTE Technology Journal, 2017, 23(04): 24 - 28.
- [15] 汪济洲, 鲁昌华, 蒋薇薇. 一种基于随机场多运动目标跟踪算法 [J]. 电子测量与仪器学报, 2017, 31(06): 909 - 913.  
WANG Jizhou, LU Changhua, JIANG Weiwei. New multi-target tracking algorithm based on conditional random field [J]. Journal of Electronic Measurement and Instrumentation, 2017, 31(06): 909 - 913.
- [16] 魏勇. 关联语义结合卷积神经网络的文本分类方法 [J]. 控制工程, 2018, 25(02): 367 - 370.  
WEI Yong. A text classification method based on associative semantics and convolution neural network [J]. Control Engineering of China, 2018, 25(02): 367 - 370.
- [17] 姜有泉, 汤亚芳, 黄良, 等. 智能配电网无线通信系统可靠性研究 [J]. 电力大数据 2017, 20(12): 52 - 55.  
JIANG Youquan, TANG Yafang, HUANG Liang, et al. Study on reliability of wireless communication system in smart distribution network [J]. Power systems and big data. 2017, 20(12): 52 - 55.

收稿日期: 2019 - 09 - 02

作者简介:



罗 云(1980), 男, 本科, 高级工程师, 主要从事电力系统通信网规划、运维和管理工作,

(本文责任编辑: 范 斌)

## Research on risk identification and assessment method of electric power communication network based on big data

LUO Yun, GAO Yanhong, WANG Zhiqiang

(China Comservice Enrising Information Technology Co., Ltd., Chengdu 610093 Sichuan, China)

**Abstract:** In order to efficiently and accurately identify potential risk points in a large and complex network of power communication networks, analyze the impact of power communication from the aspects of network structure and bearer services, and combine the assessment criteria of power accident incident investigation procedures, objectively and scientifically evaluate the power safety incidents caused by power communication and provides decision support for risk management and control of power communication. This paper combines the specific content and characteristics of power communication risk management and control, analyzes the causes of accidents in power communication networks, infers the evolution process of accidents, proposes solutions and solutions, and verifies the power accident assessment model with high computational efficiency and accuracy through actual test cases. Provide a scientific basis for accurately locating the cause of the accident and judging the risk level. The model that combines the invulnerability, survivability and effectiveness of reliability factors of complex network theory calculate the risk factor values of related characteristics. The path coefficient analysis technique is used to verify the completeness of risk factors; The CNN model of deep convolution neural network is constructed based on the labeled sample data of risk level, it assess risk level of accidents in a real-time manner.

**Key words:** power communication network; importance evaluation; risk factors; path coefficient analysis; deep neural network; big data