

文章编号:1671-6833(2019)05-0038-06

基于 GRU 网络的配电网故障数量等级预测方法

高金峰¹, 庞 昊¹, 杜耀恒²

(1.郑州大学 产业技术研究院,河南 郑州 450001; 2.国家电网烟台供电公司,山东 烟台 264000)

摘 要: 配电网故障数量的多少直接影响配电网的运行维护与用户的用电体验,目前业界关于配电网故障数量等级预测的研究较少.给出了一种基于 GRU 网络的配电网故障数据分析与故障数量等级预测方法.通过条件熵来衡量配电网故障数量等级的历史依赖性,采用距离相关系数对诸多气象特征因素进行相关性强弱考察,筛选出最优特征子集,最后通过训练 GRU 网络实现了配电网故障数量等级的预测.算例结果证明了预测方法的有效性.

关键词: 配电网故障数量等级; 循环神经网络; GRU 网络; 历史依赖性; 相关性

中图分类号: TM755

文献标志码: A

doi: 10.13705/j.issn.1671-6833.2019.05.007

0 引言

据统计,80%以上的用户停电事故是配电网故障导致的^[1].我国配电网运维检修普遍存在检修工作量大、检修时间紧等问题^[2].作为一种辅助最优决策与调度的手段,配电网故障数量等级评估与预测对提升检修效率、科学调配人力资源具有现实意义.

导致配电网故障的因素分为内部因素和外部因素.内部因素包括设备及线路老化程度、性能参数^[3],部分指标难以量化,并且由于各地配电网智能化、信息化程度不一,造成了数据积累匮乏,使得数据难以被全部、准确地采集并使用;外部因素包括社会、季节及各种气象因素,在恶劣气象条件下,更易发生配电网故障^[3-4].我国早已建立了完整的气象观测体系,气象及季节因素相对容易量化,获取数据的途径更具普适性和便捷性.

国内外关于此类问题的研究,大致可以分为基于统计和基于数据两大类,研究外部因素的影响居多.文献[5]针对电压、电流等内部因素,提出了基于长短期记忆网络(long short-term memory, LSTM)与支持向量机(support vector machine, SVM)的预测方法;文献[6]则针对气象等外部因素,提出了一种基于 SVM 的预测方法;文献[7]主要研究了风暴气象下配电网故障的预测方法;文献[8-9]基于

统计的方法提出了一种在恶劣气象条件下配电网故障的预测方法;文献[10-13]分别基于多元线性回归、时间序列、随机森林等算法建立了计及气象因素的预测模型.上述研究中,基于统计的方法建模环节繁琐,在实际的社会、自然及电力系统环境中,模型的鲁棒性较差;其他诸如机器学习等基于数据的方法,虽然在处理非线性问题的鲁棒性方面优于传统方法,但其在处理大数据、挖掘深层关系上捉襟见肘.

在电力系统其他领域的研究中,文献[14]采用 LSTM 神经网络算法,实现了对电力变压器的运行状态预测.文献[15]提出了一种基于门控循环单元网络与卷积神经网络(convolution neural network, CNN)的风电功率预测方法.以上研究采用了基于循环神经网络(recurrent neural network, RNN)的 LSTM、门控循环单元网络(gated recurrent Unit, GRU)等算法.循环神经网络以其在处理时间序列方面的出色表现^[16],已经被用于研究电力系统的部分问题,但尚未被用于配电网故障数量等级预测等相关问题.

针对预测精确度有待提高、内部因素原始数据获取困难等问题,结合深度学习、数据挖掘技术现状,笔者提出了一种基于 GRU 网络的配电网故障数据分析与故障数量等级预测方法.该方法以气象、季节因素为研究对象,通过数据分析、特征

收稿日期:2018-12-17;修订日期:2019-03-11

作者简介:高金峰(1963—),男,河南项城人,郑州大学教授,博士,主要研究方向为电工理论与新技术, E-mail: jfgao@zzu.edu.cn.

筛选、训练GRU网络等步骤实现预测,算例结果证明了方法的有效性。

1 数据分析与特征筛选

1.1 数据预处理

(1) 数据聚类.为了便于和文献[6]作对比,根据配电网故障数据实际分布情况,笔者将离散的原始配电网故障数量聚类为0~5、6~10及10次以上3个等级,分别对应等级1、等级2和等级3。

(2) 归一化.气温数据存在正值和负值,降水量数据也较为离散,采用公式(1)对上述两类数据进行归一化。

$$x_n = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (1)$$

式中: x 为归一化前的数据; x_{\max} 和 x_{\min} 为原始数据中同类数据的最大值和最小值。

(3) 特征映射.季节因素常用文字描述,需要量化为数字形式,转换规则如表1。

表1 季节因素的转换规则

Tab.1 Transformation rules of seasonal factors

季节	转换规则	转换形式
春季	连续7日日平均气温在10℃以上	1
夏季	连续7日日平均气温在23℃以上	2
秋季	连续7日日平均气温在22℃以下	3
冬季	连续7日日平均气温在9℃以下	4

1.2 历史依赖性分析

电力系统在大多数情况下都保持运行状态,为了判断当前配电网故障数量等级与历史故障数量等级间的联系,笔者采用条件熵来衡量历史依赖性的强弱.定义配电网故障数量等级的条件熵为:

$$H(S|N) = - \sum_{q=1}^n \sum_{j=1}^k (p(n_q, s_j) \lg p(s_j | n_q)), \quad (2)$$

式中: S 为配电网故障数量等级,分为等级1、等级2和等级3,即 $S = \{s_1, s_2, s_3\}$; N 为聚类后的故障数量等级时间序列, $N = \{n_1, n_2, \dots, n_n\}$ 代表 n 天前的故障数量等级数列; $p(n_q, s_j)$ 为 n_q, s_j 的联合分布概率; $p(s_j | n_q)$ 为 s_j, n_q 的条件概率. $H(S|N)$ 越小,说明在此条件下,配电网故障数量等级的不确定性越低,越容易预测,反之说明不确定性越高。

1.3 相关性分析

采用距离相关系数来衡量特征因素与配电网故障数量等级间的相关性强弱,以此来剔除冗余的特征因素,确定最优特征子集.因笔者无法收集

到足够充足有效的内部因素数据,此处特征因素仅指气象因素,此法对于内部因素数据同样适用。

距离相关系数是皮尔斯相关系数的改进,克服了原方法只对线性相关性敏感的短板,且不需要任何假设的前提条件,使得其无论在线性或者非线性相关性分析中均有广泛的应用。

定义两个变量 M 和 V ,其间的距离相关系数定义为 $dCor(M, V)$. $dCor(M, V)$ 为0,说明 M 和 V 相互独立; $dCor(M, V)$ 越大,说明 M 和 V 的相关性越强. $dCor(M, V)$ 的计算公式如下:

$$dCor(M, V) = \left(\frac{dCov(M, V)}{\sqrt{dCov(M, M) dCov(V, V)}} \right)^{1/2}. \quad (3)$$

当上式分母中一项或者全部为0时,经验距离协方差 $dCov(M, V)$ 计算公式如下:

$$dCov(M, V) = \frac{1}{n^2} \sum_{k,l=1}^n M_{kl} \cdot V_{kl}; \quad (4)$$

$$M_{kl} = m_{kl} - \hat{m}_k - \bar{m}_l + \bar{m}; \quad (5)$$

$$\bar{m}_k = \frac{1}{n} \sum_{l=1}^n m_{kl}; \quad (6)$$

$$\bar{m}_l = \frac{1}{n} \sum_{k=1}^n m_{kl}; \quad (7)$$

$$\bar{m} = \frac{1}{n^2} \sum_{k,l=1}^n m_{kl}; \quad (8)$$

$$m_{kl} = |M_k - M_l|_p, \quad (9)$$

其中, $M \in R^p$, 对于 $V \in R^q$ 计算公式类似。

2 GRU网络及其训练

2.1 网络结构

GRU网络是RNN的一种变体,实际中单纯的RNN网络结构很容易出现梯度消失或梯度爆炸问题。

GRU网络可以解决RNN的前述问题. GRU网络的隐藏层单元结构如图1.包含重置门 r_t 和更新门 z_t 两个状态门,通过控制状态门的权重在0~1变化,实现对“历史记忆”的加强与遗忘.具体的公式表示如下:

$$r_t = \sigma(U_r \cdot x_t + W_r \cdot h_{t-1}); \quad (10)$$

$$z_t = \sigma(U_z \cdot x_t + W_z \cdot h_{t-1}); \quad (11)$$

$$H_t = \tanh[U_h \cdot (r_t \cdot h_{t-1}) + W_h \cdot x_t]; \quad (12)$$

$$h_t = (1 - z_t) \cdot H_t + z_t \cdot h_{t-1}; \quad (13)$$

式中: h_t 为候选隐藏层状态; \tanh 为隐藏层激活函数,其他符号含义与前述相同。

笔者构建的3层神经网络架构如图2,将外部因素序列作为数据输入,通过3层GRU神经网络

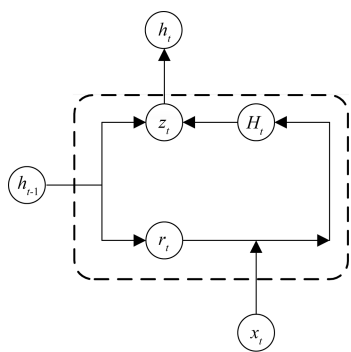


图1 GRU 的隐藏层单元结构

Fig.1 Hidden layer unit structure of GRU

络进行前向计算,挖掘数据纵向时序联系以及横向非线性关系,最终通过 Softmax 分类器根据最大概率准则得出配电网故障数量等级预测结果。

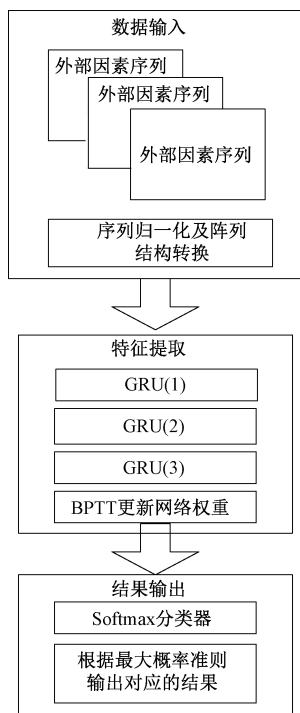


图2 网络构架

Fig.2 Network architecture

2.2 训练网络

将筛选后的特征变量与每日配电网故障数据整合至同一数据库中,便于统一输入网络,数据格式统一采用科学计数法.再将集成后的数据集随机拆分为训练集和测试集,为了避免训练集数目不同造成不同故障等级学习及预测效果的差异,根据前述第2部分中配电网故障数量的3个等级,每个故障数量等级各对应相同组数据作为训练集,对网络采用有监督学习的方式进行训练,并采用沿时间反向传播算法(back propagation trough time, BPTT)更新网络权重以及(root mean square prop, RMSprop)优化算法进行梯度优化。

网络的超参数主要包括网络层数、隐藏层神经元个数、输出层神经元个数和学习率.鉴于笔者采集的原始数据量有限,为避免梯度消失过早,采用了较小的学习率进行训练.经过对不同的结构搭配及超参数调试,最终的网络由3层隐藏层和1层 Softmax 全连接输出层堆叠而成.其中,隐藏层的激活函数采用 tanh, units 值分别为 64、32 和 16, dropout 为 0.4, Softmax 输出层 units 为 3, learning rate 为 0.005, 学习轮数为 1 000。

2.3 评估网络

笔者采用精确度 A (accuracy)、查准率 P (precision) 和查全率 R (recall) 来评估网络的分类预测能力,定义 N_{TP} 、 N_{FP} 、 N_{TN} 和 N_{FN} 分别为每个故障数量等级预测结果中真正例、假正例、真反例和假反例的数量.计算公式如下:

$$A = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{FN} + N_{FR} + N_{TN}}; \quad (14)$$

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}; \quad (15)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}}. \quad (16)$$

3 算例及结果分析

3.1 原始数据分析及历史依赖性分析

数据来源于国家电网华东某地级市供电公司 10 kV 配电网故障抢修日志,2005 年 1 月至 2010 年 12 月期间,该地级市 10 kV 配电网故障次数共 4 115 次,每月故障数量统计如图 3.这些故障可以分为状态故障、设备故障和客户内部故障 3 大类,具体比例如图 4.客户内部故障虽然主要由客户自身原因导致,但考虑到会对局部配电网造成一定的影响,故在统计时保留此部分故障。

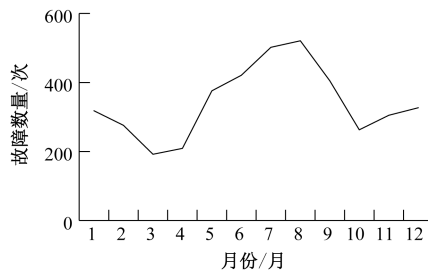


图3 每月故障数量统计趋势

Fig.3 Monthly failure statistics trend

由图3、图4可得,该地级市配电网的故障数量根据月份、季节不同呈现一定的规律,而时间、季节与气象因素有较为直接的因果关系.其中六、

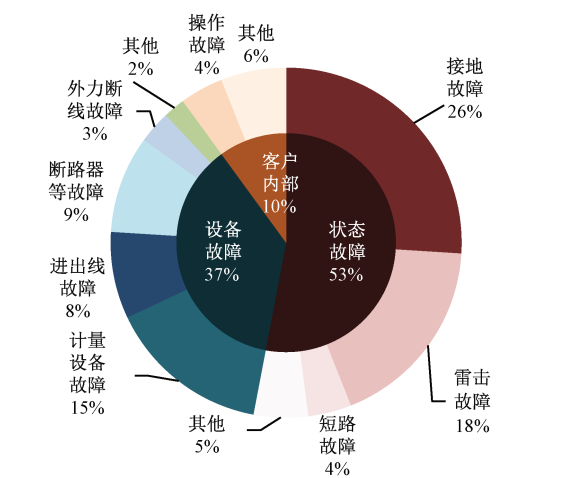


图 4 故障类型
Fig.4 Fault type

七、八三月份及一月、十二月份的故障数量达到半年内局部最大值,三、四及十月为半年内局部最小值.这与该地级市根据季节、温度变化的生产生活规律相符.

配电网故障数量等级的条件熵如图 5 所示.由图 5 可知,条件熵随着 n 的增大呈下降趋势,即配电网故障数量等级对历史数据有较大的依赖,随着历史数据的增加,配电网故障数量等级可预测性逐渐增强,挖掘时间序列的前后关系正是 RNN 及 GRU 网络的优势.由此进一步印证

了笔者所研究的问题与算法之间的匹配性与适用性.

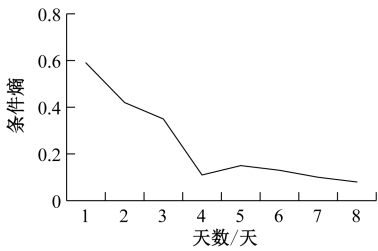


图 5 不同天数情况下的条件熵
Fig.5 Conditional entropy under different days

3.2 气象因素的距离相关系数

气象因素主要包括气温、气压、风速、风向、降水量、相对湿度和光照小时数 7 种.笔者根据国家气象信息中心网站获取的历史气象数据,7 种气象因素的距离相关系数计算数值从大到小排列如表 2,根据统计学经验及惯例,可以认为距离相关系数计算值小于同批次最大值 25%的因素与配电网故障数量具有弱相关性,故将气压、风向、日照小时数剔除,由此确定出气象因素的最优特征子集为气温、风速、降水量、相对湿度.值得说明的是,由于配电网所处的地区不同、数据特征不同,各气象因素的相关性强弱也可能与本研究不尽相同.

表 2 各气象因素的距离相关系数

Tab.2 The distance correlation of meteorological factors

气象因素	气温	降水	风速	相对湿度	气压	风向	日照小时	参考阈值
dCor	0.786	0.613	0.549	0.396	0.136	0.106	0.090	0.197

根据前述分析,故障数量的分布也与季节有关,在气象因素最优特征子集的基础上,引入季节组成下一步输入网络的特征集合.同时考虑到有相当一部分配电网故障是在夜间发生的,将气温、风速、降水量按照日间和夜间分别统计,即季节、日间最高气温、夜间最低气温、日间(夜间)最大风速、日间(夜间)最小风速、日间(夜间)降水量、日间(夜间)相对湿度,共 11 种.

3.3 预测结果及评估

经过数据集成,利用 24 090 个经前述方法处理分析的数据训练、测试网络,经过 1 000 轮迭代后,网络在测试集上的精确度达到 80.41%.本节将针对网络的各项指标进行分析评估.

将前述网络与反向传播神经网络(back propagation neural network, BPNN)及文献[6]所述支持向量机的对比.为保证与前述 GRU 网络构架参数相同,BPNN 也采用 3 层隐藏层和 1 层 Softmax

全连接输出层,隐藏层的激活函数采用 tanh,units 值分别为 64、32 和 16,dropout 为 0.4,Softmax 输出层 units 为 3,learning rate 为 0.005,同样训练 1 000 轮,GRU 网络与 BPNN 的查准率与查全率对比如表 3,精确度如图 6.从表 3 可以看出,GRU 网络在 3 个故障数量级的评价指标上均优于 BPNN,说明 GRU 网络对不同类别的样本均有较好的适应性.由图 6 可以看出在相同数据集的情况下,BPNN 的预测精确度停留在较低的水平.这是由于 BPNN 是将动态的时间序列转换至静态空间建模,与 RNN 及 GRU 有本质区别,难以充分挖掘时间序列内部历史关系.也从侧面印证了前述对于配电网故障数量等级历史相关性分析的正确性.

图 7 为 GRU 与文献[6]SVM 算法的精确度对比.SVM 算法基于 Scikit-learn 机器学习框架实现,根据文献[6]所述,采用高斯径向基核函数,

表 3 查准率与查全率的对比

Tab.3 Comparison between precision rate and recall rate / %

类别	GRU			BPNN		
	1	2	3	1	2	3
查准率	79.28	81.92	77.66	57.13	48.57	60.03
查全率	64.89	72.12	76.20	52.59	56.82	50.26

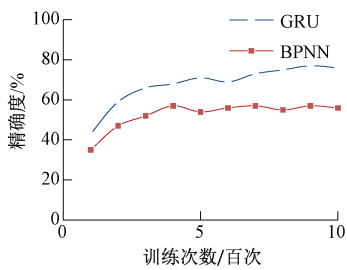


图 6 GRU 与 BPNN 的对比

Fig.6 Comparison of GRU and BPNN

gamma 为 0.34,C 为 30.可以看出在相同的分类条件下,GRU 网络相较于 SVM 算法,在精确度上有一定的提高,尤其是在 3 分类的精确度上,提升了约 10 个百分点,但由于笔者仅以外部因素为研究对象及输入特征,以及特征维度及数据集大小的原因,预测精确度的提升也是有限的.

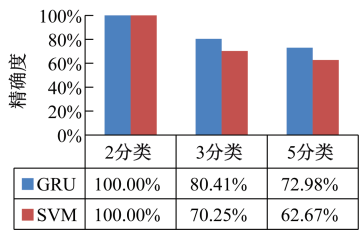


图 7 GRU 与 SVM 的对比图

Fig.7 Comparison of GRU and SVM

(1) 预测结果与网络深度的关系.笔者尝试增加网络的深度以期提升网络的预测精确度,经测试发现,在网络深度为 4 层时,精确度与 3 层网络几乎相同,但当网络深度为 5 层时,测试集精确度反而大幅下降,产生严重的过拟合问题.这是由于随着网络深度的增加,需要学习的网络参数陡增,而训练集大小及特征维度不能与之相匹配.

(2) 预测结果与特征选择的关系.根据表 1,现选取三组气象因素:第一组选取表 1 中全部气象因素,共 7 种;第二组选取相对湿度、气压、风向、日照小时数,共 4 种;第三组为前述根据距离相关系数法选取的 4 种气象因素,即气温、降水、风速、相对湿度.3 组特征除气象因素外,季节因素均相同,分别训练前述 GRU 网络,网络结构、参数及训练次数均与前述相同,测试集精确度如表 4.对比第一组

与第三组可得,在不剔除冗余特征的情况下,将增加网络训练工作量并降低预测精确度;对比第二组与第三组可得,选择具有强相关性的特征变量将显著提升网络的预测精确度,由此也印证了前述距离相关系数法的实际意义.

表 4 精确度对比

Tab.4 Comparison of accuracy rates %

组别	第一组	第二组	第三组
精确度	74.87	53.19	80.41

4 结论

(1) 所提方法在有效筛选故障特征的基础上,使用 GRU 网络能更加充分地利用多变量信息、挖掘时间序列历史关系,较 BPNN、SVM 等算法,提高预测精度.

(2) 如果采用数值天气预报等高精度、分时段、分区域的外部特征数据,在笔者所述方法的基础上,可以提高预测时效及精确度.

(3) 随着电网智能化、信息化程度的不断提升,将产生和积累海量的数据,笔者所提方法在处理大数据方面的优势将不断凸显;在外部因素的基础上,结合更多的配电网内部因素,可以增加 GRU 网络的深度,在预测配电网故障数量等级的基础上,为预测配电网故障风险等级提供有效途径.

参考文献:

[1] BILLINTON R, WANG P. Reliability-network-equivalent approach to distribution-system-reliability evaluation[J]. Generation, Transmission and Distribution, 145(2):149-153.

[2] 张稳,盛万兴,刘科研,等.计及气象因素相关性的配电网故障风险等级预测方法[J].电网技术, 2018,42(8):2391-2398.

[3] 刘科研,吴心忠,石琛,等.基于数据挖掘的配电网故障风险预警[J].电力自动化设备,2018,38(5): 148-153.

[4] 马瑞,周谢,彭舟,等.考虑气温因素的负荷特性统计指标关联特征数据挖掘[J].中国电机工程学报, 2015,35(1):43-51.

[5] ZHANG S L, WANG Y X, LIU Z J, et al. Data-based line trip fault prediction in power systems using LSTM networks and SVM [J]. IEEE Access, 2018 (3):7675-7686.

[6] BAI Y L, LI Y H, LIU Y G, et al. Short-term prediction of distribution network faults based on support vector

- machine [C]//Proceeding of the 2017 12th IEEE Conference on Industrial Electronics and Application, 2018;1421-1426.
- [7] ZHU D, CHENG D L. Storm modeling for prediction of power distribution system outages [J]. Electric Power Systems Research, 2007, 77(8): 973-979.
- [8] LIU H B, DAVIDSON R A, ROSOWSKY D V, et al. Negative binomial regression of electric power outages in hurricanes [J]. Journal of Infrastructure Systems, 2005, 11(4): 258-267.
- [9] LIU H B, DAVIDSON R A, APANASOVICH T V. Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms [J]. Reliability Engineering and System Safety, 2008, 93(6): 875-890.
- [10] 张鹏飞, 瞿海妮, 肖其师, 等. 基于气象因素和时间序列分析的配电网故障数量预测[J]. 陕西电力, 2016, 44(1): 68-72.
- [11] 蔡冬阳, 彭涛. 基于多元线性回归的故障报修受理数量预测研究[J]. 江苏科技信息, 2016(20): 61-65.
- [12] 程森海, 楼俏, 王琼, 等. 基于随机森林算法的配网抢修故障量预测方法[J]. 计算机系统应用, 2016, 25(9): 137-143.
- [13] 王继业, 魏晓菁, 郝悍勇, 等. 基于灰色投影随机森林算法的配网故障量预测模型[J]. 自动化技术与应用, 2018, 37(4): 75-78.
- [14] 代杰杰, 宋辉, 盛戈崑, 等. 采用 LSTM 网络的电力变压器运行状态预测方法研究[J]. 高电压技术, 2018, 44(4): 1099-1106.
- [15] 牛哲文, 余泽远, 李波, 等. 基于深度门控循环单元神经网络的短期风功率预测模型[J]. 电力自动化设备, 2018, 38(5): 36-42.
- [16] 赵淑芳, 董小雨. 基于改进的 LSTM 深度神经网络语音识别研究[J]. 郑州大学学报(工学版), 2018, 39(5): 63-67.

A Method for Predicting the Number of Faults in Distribution Network Based on GRU Neural Network

GAO Jinfeng¹, PANG Hao¹, DU Yaoheng²

(1. Industrial Technology Research Institute, Zhengzhou University, Zhengzhou 450001, China; 2. State Grid Yantai Power Supply Company, Yantai 264000, China)

Abstract: The number of faults in distribution network was a direct impact on the operation and maintenance of distribution network and the user's power consumption experience. At present, there were few studies on the prediction of the number of faults in distribution network. To measure the historical dependence of distribution network fault magnitude, the optimal feature subset was selected by using the distance correlation coefficient to investigate the correlation of many meteorological features. Finally, the GRU neural network was trained to predict the fault magnitude of distribution network accurately. The results proved the feasibility of this method.

Key words: number of faults in distribution network; RNN; GRU neural network; historical dependence; relevance