·安全技术· 文章编号：1000-3428(2019)04-0119-05 文献标志码：A 中图分类号：TM769

基于 word2vec 的配电网恶意控制指令检测算法

郑佩祥[1]，陈 彬[1]，卢 昕[2]，徐文渊[2]

（1. 国网福建省电力有限公司，福州 350003；2. 浙江大学 电气工程学院，杭州 310027）

摘 要： 现有的配电网恶意控制指令检测方法基于电力系统运行规则，但规则维护困难、规则匹配耗时较长。根据配电网上行测量信息和下行控制指令之间存在的上下文一致性关系，提出基于 word2vec 的恶意控制指令检测算法。在配电网仿真平台模拟各类工况并获取标注数据集，结果表明，该算法能够取得 100% 的精确度和 87.2% 的召回率，具有较高的检测精度。

关键词： 配电网；恶意控制指令；上下文；word2vec 模型；异常检测

中文引用格式： 郑佩祥，陈彬，卢昕，等. 基于 word2vec 的配电网恶意控制指令检测算法[J]. 计算机工程，2019，45(4)：119-123，129.

英文引用格式： ZHENG Peixiang，CHEN Bin，LU Xin，et al. Malicious control command detection algorithm in power distribution network based on word2vec[J]. Computer Engineering，2019，45(4)：119-123，129.

Malicious Control Command Detection Algorithm in Power Distribution Network Based on word2vec

ZHENG Peixiang[1]，CHEN Bin[1]，LU Xin[2]，XU Wenyuan[2]

（1. State Grid Fujian Electric Power Company，Fuzhou 350003，China；
2. College of Electrical Engineering，Zhejiang University，Hangzhou 310027，China）

【Abstract】 The existing detection method of malicious control command of distribution network is based on the operating rules of power system，but the rules are difficult to maintain and the rule matching takes a long time. According to the context consistency relationship between the uplink measurement information of the distribution network and the downlink control instructions，a malicious control detection algorithm based on word2vec is proposed. Simulating various working conditions on the power distribution network simulation platform and obtaining the labeled data set，the results show that the algorithm can achieve 100% accuracy and 87.2% recall rate，and has high detection accuracy.

【Key words】 power distribution network；malicious control command；context；word2vec model；abnormal detection

DOI： 10.19678/j.issn.1000-3428.0049685

0 概述

近年来，信息通信已成为智能配电网的关键支撑技术，广泛应用于电力领域。智能配电网将发展为电力空间和信息空间（包括各类信息系统及信息设备）高度融合的配电网信息物理系统（Cyber-Physical System，CPS）[1-2]，其中，电力空间覆盖电力一次设备是配电网 CPS 的承载基础，信息空间涵盖电力二次系统与电力信息系统作为配电网 CPS 的控制与运算平台。

当处于离散时空维度的信息空间和处于连续时空维度的电力空间深度融合后，信息空间中原有的各类安全风险也被引入到智能配电网中并严重影响其安全稳定运行。由于配电网规模庞大、结构复杂，一旦遭受来自具有电力系统背景的黑客发起网络攻击，即使是局部微小扰动，都极有可能导致电力系统的大面积瘫痪[3]。此类针对配电网控制系统的网络攻击称为恶意控制指令攻击。在 2015 年 12 月 23 日发生的乌克兰大规模停电事件中，黑客通过获取控制权限并下达恶意停电指令等方式攻击乌克兰配电网络，直接导致数十万用户停电[4]。恶意控制指令攻击对配电网的安全稳定运行危害巨大，亟需研究针对恶意控制指令攻击的检测技术。

本文采用异常检测方法检测配电网恶意控制指令。根据配电网的上行测量信息和下行控制指令之间存在上下文一致性关系，通过 word2vec 方法提取

基金项目： 国家高技术研究发展计划（2015AA050202）；国家电网公司科技项目（52130415000P）。

作者简介： 郑佩祥（1970—），男，高级工程师，主研方向为智能配电网安全；陈 彬，高级工程师；卢 昕，硕士研究生；徐文渊，教授、博士。

收稿日期： 2017-12-13 **修回日期：** 2018-03-12 **E-mail：** 602666317@qq.com

该上下文关系生成电力通信报文对应离散事件的特征向量和序列向量,利用孤立森林算法建立白名单模型并检测恶意控制指令。

1 相关工作

目前,工业界和学术界尚未提出完善的恶意控制指令攻击的检测技术。电力控制系统采取控制指令预演的方式检测恶意控制指令,即在系统中仿真控制指令下发后的电力系统潮流,观察该控制指令是否会导致电力系统失去稳定[5]。文献[6]提出一种基于电力通信报文语义分析的电力网络入侵检测算法。该算法通过解析电力通信报文的内容并分析报文是否符合对应的电力通信协议标准,从而判断是否存在电力网络入侵行为。文献[7]在电力通信报文语义分析基础上,提出基于预先潮流计算的恶意控制指令攻击检测算法,为减少潮流计算导致的延时,根据系统节点规模对潮流计算的迭代轮数进行优化。上述检测方法均基于提取电力一次系统或二次系统运行规则,存在规则提取依赖专家经验、规则提取和维护困难、规则匹配耗时会随系统规模增大而急剧增加的缺点。

基于机器学习的检测算法是一种数据驱动的检测算法,在训练过程中可从大量数据中自动地学习和提取相应的“规则”,有利于解决基于规则的检测算法存在的缺陷。机器学习应用于恶意控制指令攻击检测的最大难点在于缺乏足够的负样本数据。虽然配电信息网络中存在大量的正常通信报文,但攻击样本十分稀少,即正负样本分布极度不均衡,因此会对模型的学习和训练造成巨大的困难。此外,由于知识的局限性,防守方人工构造的负样本与黑客精心构造的、变化多样的攻击往往存在差异,因此人工构造负样本训练的模型在实际中效果并不理想[8]。目前关于网络攻击检测的研究热点在于异常检测,其原理是使用大量正常的数据训练白名单模型,检测时将与白名单模型偏差较大的数据判定为为异常数据。异常检测方法相对于传统的“硬规则”更难被黑客绕过,且有望发现零日漏洞[9-10]。

2 配电网恶意控制指令检测算法框架

针对攻击样本稀少且变化多样、正负样本分布极度不均衡导致模型难以学习和训练的问题,本文在异常检测基础上提出基于 word2vec 的配电网恶意控制指令检测算法,整体框架如图 1 所示,包括以下 3 个部分:

1)数据获取、标注和划分。在某配电网仿真平台上模拟各种工况(包括正常场景、故障场景和攻击场景),以仿真平台自带的故障处理系统和控制指令预演系统的输出结果标注数据,记录电力通信报文对应的离散事件和标签。本文采用异常检测框架,将数据集中 80% 的正样本作为训练数据,其余 20% 正样本和所有的负样本作为测试数据。

2)训练白名单模型。根据配电网的上行测量信息和下行控制指令之间存在上下文一致性关系,通过 word2vec 以最大化后验概率 P(当前信息|上下文信息)为目标训练电力通信报文对应离散事件的特征向量,利用孤立森林算法建立白名单。

3)模型测试。根据 word2vec 结果生成测试数据离散事件的特征向量,在白名单模型中进行搜索和匹配是否存在满足条件的故障点定位向量,若存在满足条件的故障点定位向量,则输出正常;否则输出异常。

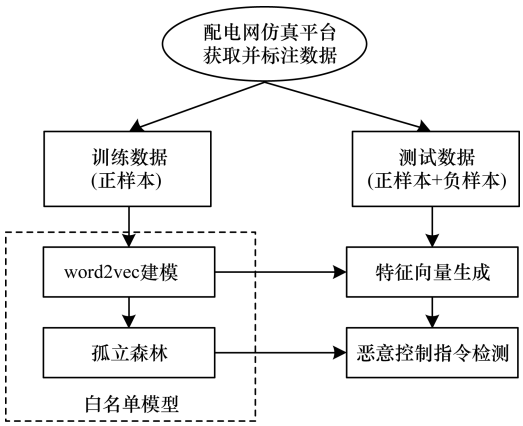

图 1 配电网恶意控制指令检测算法框架

3 配电网故障处理

配电网故障处理是依据配电网的网架结构和设备运行的实时信息,结合故障信号,进行故障的定位、隔离和非故障失电区域的恢复供电。图 2 所示为配电网结构,以其作为算例说明故障处理的逻辑过程。其中,S_1、S_2、S_3 为变电站出线开关,其余为配电网开关,开关黑色实心表示开关闭合,白色空心表示开关断开,开关之间的黑色实线表示电缆线。

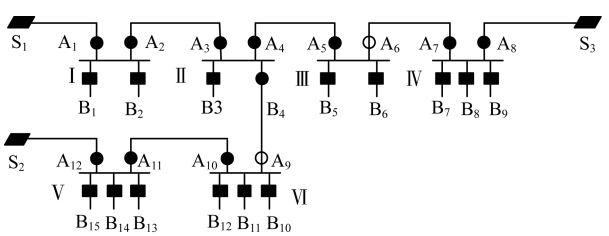

图 2 配电网结构

配电网故障处理分为以下 3 个步骤:1)故障点定位,即根据设备的测量信息定位故障点;2)故障隔离,即断开与故障点直接相连的开关;3)非故障失电区域的恢复供电。下文以断路器出口故障和电缆线故障为例,具体说明配电网故障处理的原理。

图 3 所示为断路器出口故障,当箭头处发生故障时,设备会上传以下测量信息:

1)断路器 S_1 开关分闸;

2)断路器 S_1 的保护动作。

根据测量信息可判定 S_1 ~ A_1 之间区域发生故障,即出口断路器 S_1 故障。故障处理包括以下 2 个步骤:

1)断开 A_1 完成故障区域隔离;

2)合上 A_9 或者 A_6 恢复故障下游(在供电充足情况下)。

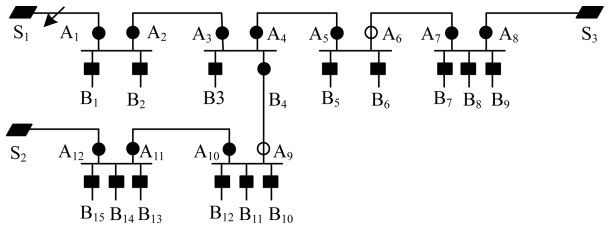

图 3　断路器出口故障

图 4 所示为电缆线出口故障,当箭头处发生故障时,设备会上传以下测量信息:

1)断路器 S_1 开关分闸;

2)断路器 S_1 的保护动作;

3)A_1 保护动作;

4)A_2 保护动作。根据测量信息,可判定 A_2 ~ A_3 之间区域发生故障,即电缆线故障。故障处理包括以下3 个步骤:

1)断开开关 A_2、A_3 隔离故障区域;

2)合上 S_1 恢复上游供电;

3)合上 A_9 或者 A_6 恢复故障下游供电(在供电充足情况下)。

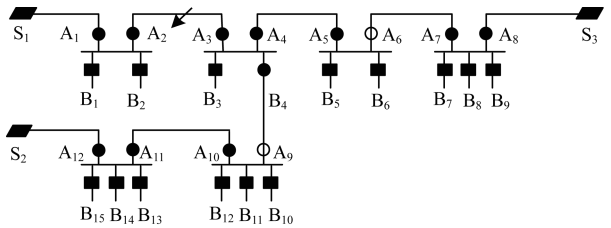

图 4　电缆线故障

4　word2vec 模型

4.1　词向量

词的向量化就是将自然语言中的基本单位词表示成一个向量,使之能用数学方法描述并被计算机处理,是自然语言处理的基础任务之一。词的向量化主要有独热表示方式和词嵌入方式 2 种。

1)独热表示方式

独热表示方式[11]的基本思想是用一个高度稀疏的向量表示一个词,向量的长度为词典大小,向量中某一维的值取 1 代表词表中对应位置的词出现,值取 0 代表词表中对应位置的词未出现。例如,“番茄”对应的词向量为[1 0 0 … 0 0 0],而“西红柿”对应的词向量则为[0 1 0 … 0 0 0]。用独热方式表示一个词语简单易行且便于理解,但是存在无法有效表示词语的语义信息的问题。例如,“番茄”和“西红柿”、“土豆”和“马铃薯”尽管是同义词,但是如果利用常规的向量距离度量方式(如欧式距离和余弦相似度)计算“番茄”和“西红柿”间的相似度,“番茄”和“土豆”间的相似度,通常不能得到“番茄”和“西红柿”相似度较高而“番茄”和“土豆”相似度较低的结果,独热表示方式提取的语义特征不能较好地表达词之间的相似性。对于本文的问题,同时断开开关 A_1 和开关 A_2,断开开关 A_1 和开关 A_4,断开开关 A_1 和开关 A_8 的独热表示分别为[1 1 0 0 0 0 0 0 0 …],[1 0 0 1 0 0 0 0 0 0 …],[1 0 0 0 0 0 0 0 1 0 …],两两计算这 3 个向量的欧式距离和余弦相似度的结果相同,但是断开开关 A_1 和 A_8 导致的停电范围远大于断开开关 A_1 和 A_2,因此独热表示方式难以提取配电网故障信息的语义特征。

2)词嵌入方式

为克服独热表示方式提取的语义特征难以表达词之间相似性的缺点,文献[12]提出词嵌入方式,也叫词的分布式表示方式。它的基本思想是将独热表示的词向量从原本高维、稀疏、0/1 表示的向量空间根据词的上下文信息映射到一个低维(一般为 50 维或者 100 维)、稠密的实数向量空间,使得词义越相近的词向量之间的相似度越高或者在空间的距离越近。文献[13]提出 word2vec 模型。该模型利用词的上下文信息将一个词转化成一个低维实数向量,越相似的词在向量空间中越相近。word2vec 模型有 2 种:连续词袋模型(Continuous Bag-of-Word,CBOW)和跳字模型(Skip Gram,SG)。CBOW 模型是给定上下文来预测中心词出现的后验概率,而 SG 模型是给定中心词来预测上下文出现的后验概率。由于本文白名单模型是通过 CBOW 模型训练得到,且 CBOW 和 SG 的原理十分类似,因此下文将重点介绍 CBOW 模型原理。

4.2　CBOW 模型

CBOW 模型的一个基本假设是每个词出现的概率和它的上下文有关,如果上下文的长度为 2,则对于给定词序列 $\boldsymbol{w}_{t-2}\boldsymbol{w}_{t-1}\boldsymbol{w}_t\boldsymbol{w}_{t+1}\boldsymbol{w}_{t+2}$,中心词 $\boldsymbol{w}_t$ 出现的概率依赖于其上下文词,即:

$$P(\boldsymbol{w}_t)=P(\boldsymbol{w}_t|context(\boldsymbol{w}_t))=P(\boldsymbol{w}_t|\boldsymbol{w}_{t-2}\boldsymbol{w}_{t-1}\boldsymbol{w}_{t+1}\boldsymbol{w}_{t+2})$$

CBOW 模型结构由一个上下文词向量的累加器和一个单层神经网络组成,如图 5 所示。其中,输入层为中心词 $\boldsymbol{w}_t$ 对应各个上下文词的词向量,映射层负责将各上下文词的词向量累加得到上下文向量,并作为单层神经网络的输入,输出层代表中心词 $\boldsymbol{w}_t$ 出现的后验概率,同时也是单层神经网络的输出。该神经网络的目标函数可表示为:

$$L(\boldsymbol{\theta}) = \prod_{w \in \boldsymbol{C}} p(\boldsymbol{y} = \boldsymbol{w} \mid \boldsymbol{x}) = \prod_{w \in \boldsymbol{C}} \frac{\exp(\boldsymbol{\theta}_w^{\mathrm{T}} \boldsymbol{x})}{\sum_{i=1}^{N} \exp(\boldsymbol{\theta}_i^{\mathrm{T}} \boldsymbol{x})}$$

其中，$\boldsymbol{C}$ 表示语料库，w 表示语料库中出现的词，$\boldsymbol{x}$ 表示上下文，$\boldsymbol{y}$ 是神经网络输出，即词表中各个词的概率，$\boldsymbol{\theta}$ 是神经网络权重矩阵，$\boldsymbol{\theta}_i^{\mathrm{T}}$ 是矩阵的第 i 行向量。

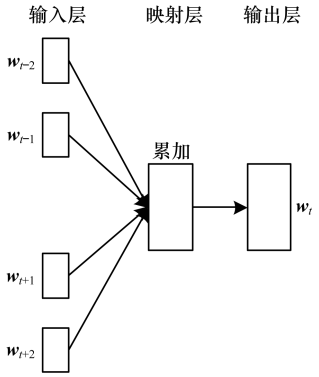

图 5 CBOW 模型原理

5 恶意控制指令检测算法

5.1 word2vec 建模

测量信息、故障点定位信息、控制信息之间存在上下文关系，即在确定三元组（测量信息，故障点定位信息，控制信息）中 2 个的前提下，可推断出另一个出现的后验概率。因此，可以用 word2vec 中的 CBOW 模型对配电网离散事件进行建模。

配电网离散事件的词表建立包括以下 5 部分：1）XX 开关异常分闸；2）XX 开关处保护动作；3）故障点在 XX 处；4）控制 XX 开关分闸；5）控制 XX 开关合闸。其中，第 1 部分和第 2 部分属于测量信息，第 3 部分属于故障点定位信息，第 4 部分和第 5 部分属于控制信息。然后本文按照词表提取配电网故障处理过程的离散事件作为词序列，依次将每个词序列的每个词作为中心词，剩余词作为上下文词，用 CBOW 模型进行建模。在模型训练完成后，将词序列中的所有词向量累加，得到序列向量。

5.2 孤立森林

本文采用的单分类模型为孤立森林算法[14]。孤立森林算法的基本思想是对各个维度的数据构建一系列的随机二叉树，这些随机二叉树每个节点要么有 2 个孩子，要么就是叶子节点。通过在取值范围内随机取值，将该范围内的数据划分成 2 个分支，然后在 2 个分支中继续随机取值进行划分。不断重复上述步骤，直至二叉树不可分割或者二叉树的高度超过限定高度。由于异常点通常是样本空间中的离群点且数量十分稀少，在随机树中容易很快被划分到叶子节点中，因此根据样本在叶子节点到根节点之间的平均路径长度是否超过阈值，即可判定该样本否为异常样本。以 5 个测试样本为例来说明孤立树的生成和算法的思想，如图 6 所示。从图 6 可以看出，节点 E 最早被孤立，代表的数据最有可能是异常数据。

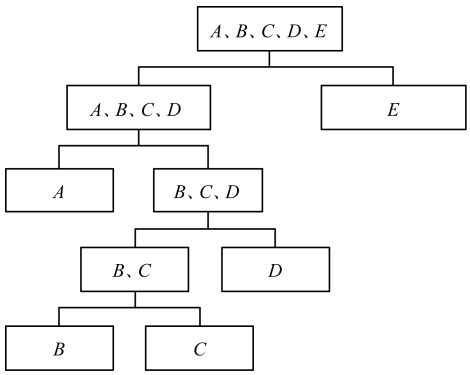

图 6 5 个测试样本的孤立树遍历示意图

孤立树的实现步骤描述如下：

步骤 1 从训练集中随机抽取 n 个点样本点作为二叉树的根节点。

步骤 2 随机选择一个维度，在样本点的特征向量在该维度数据的最大值和最小值之间随机产生一个分割点 p。

步骤 3 以此分割点生成一个超平面，然后将当前节点数据空间划分为 2 个子空间，把指定维度里小于 p 的数据放在当前节点的左孩子，大于等于 p 的数据放在当前节点的右孩子。

步骤 4 在孩子节点中递归步骤 2 和步骤 3，不断构造新的孩子节点，直到孩子节点中只有一个数据（无法再继续切割）或孩子节点已到达限定高度 。

获得所有的孤立树之后，孤立森林训练完成。对于一个测试数据 $\boldsymbol{x}$，令其遍历每一棵孤立树，然后计算 $\boldsymbol{x}$ 最终落在每个树第几层（$\boldsymbol{x}$ 在树上的高度），计算 $\boldsymbol{x}$ 在孤立森林的高度平均值。设置一个高度平均值的阈值（边界值），低于此阈值的测试数据即为异常数据。

6 实验结果与分析

6.1 实验设置

本文在配电网半实物仿真平台上模拟各类工况，如图 7 所示。该配电网半实物仿真平台分成上位机和下位机 2 个部分，其中，上位机是某厂家的配电自动化系统，下位机是数字仿真，双方的通信按照国家标准使用电力专有协议 IEC60870-5-104[15] 通信。如果拓扑处在故障状态，下位机会通过潮流计算方式仿真得到故障状态下拓扑运行状态（开关的开合，节点的电压、电流、功率等），并将相关的测量信息以 IEC60870-5-104 协议发送给上位机，上位机负责解析报文并根据 IEC61850[5] 分析故障，在弹出的故障分析界面给出相关的故障分析信息和故障处理建议，故障分析界面如图 8 所示，相关人员确认后控制指令会下发给下位机执行，下位机执行后将重新进行潮流计算，直至隔离故障。

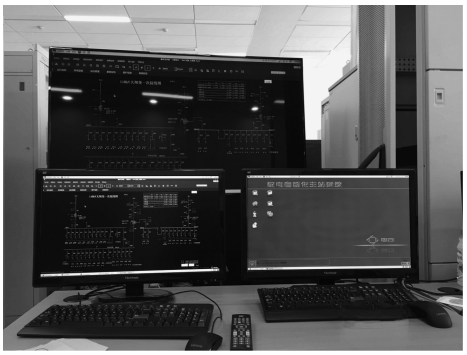

图 7　配电网半实物仿真平台

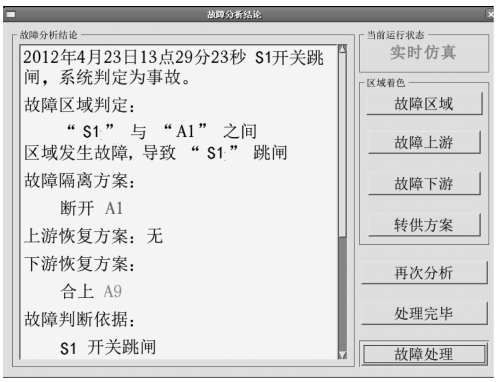

图 8　配电自动化系统故障分析界面

在该配电网半实物仿真平台上模拟正常状态和故障场景、无攻击和有攻击场景,获取数据并标注,数据集包括以下 2 个部分:

1)训练集。在仿真平台上仿真正常场景、单点故障处理和恢复场景,用仿真平台自带的模块对测量信息、故障点定位信息、控制信息进行标注,仿真 5 d,共获得 1 375 个数据。

2)测试集。测试集包括正样本和负样本 2 个部分。正样本分成 2 d 仿真数据,第 1 天仿真和训练集相同的场景,第 2 天仿真多点故障和故障信息缺失下故障处理过程中的测量信息、故障点定位信息、控制信息,共 250 个数据。负样本通过仿真以下3 类场景共获得 116 个数据:(1)在正常场景下任意发送合闸指令导致断电,对联络开关发送合闸指令导致短路;(2)发送无效的控制指令,如对已闭合的开关下发闭合指令或对已断开的开关下发断开指令;(3)在故障场景下发送不正确的故障处理指令。

6.2　结果与分析

编写 python 程序,用训练集的数据训练电力离散事件的 word2vec 模型,其中,词向量长度为 50,激活函数为 ReLU,训练方式为 adam,学习率为 0.000 2。孤立森林模型的参数设置:孤立树个数为 100,假设离群点比例为 0.05,最大采样数为 min(256,样本数),训练数据 bootstrap 采样,在测试集上验证算法的性能指标,如表 1 所示。

表 1　本文算法在测试集上的实验结果

性能指标	正样本		负样本
	第 1 天	第 2 天	
TP	196	22	0
TN	0	0	116
FP	0	0	0
FN	24	8	0

本文算法在测试集上的准确度(Accuracy)、精确度(Precision)、召回率(Recall)计算结果为:

$$A_{\text{Accuracy}} = \frac{TP + TN}{P + N} \times 100\% = 91.3\%$$

$$P_{\text{Precision}} = \frac{TP}{TP + FP} \times 100\% = 100\%$$

$$R_{\text{Recall}} = \frac{TP}{TP + FN} \times 100\% = 87.2\%$$

ROC 曲线如图 9 所示(AUC 为 0.95)。本文算法对恶意控制指令有较高的检测精度,但是由于机器学习的过拟合,导致模型过多地学习训练集中的正样本的特征和测试集上较高的误警率。考虑到恶意控制指令的危害,通常需要设置较高的警戒阈值以尽可能多地筛选出可疑的数据供工作人员判断,因此一定的误警率在某种程度上可以容忍。

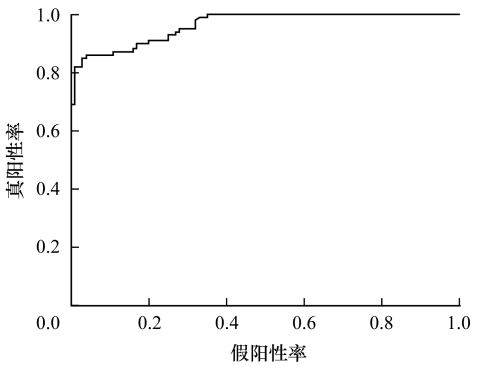

图 9　本文算法 ROC 曲线

7　结束语

针对基于规则的配电网恶意控制指令检测方法存在的规则构造维护门槛高、成本大等问题,本文提出基于 word2vec 的恶意控制指令检测算法。该算法基于数据驱动,通过挖掘配电网测量信息和控制信息的上下文,能够快速有效检测恶意控制指令。通过配电网仿真平台模拟各类工况并获取标注数据,结果验证了该算法的有效性。下一步将研究在实际数据集中如何降低算法的误警率。

参考文献

[1]　赵俊华,文福拴,薛禹胜,等. 电力 CPS 的架构及其实现技术与挑战[J]. 电力系统自动化,2010,34(16):1-7.

[2]　郭庆来,辛蜀骏,孙宏斌,等. 电力系统信息物理融合建模与综合安全评估:驱动力与研究构想[J]. 中国电机工程学报,2016,36(6):1481-1489.

(下转第 129 页)

2006:24.
[7] DWORK C. Differential privacy[C]//Proceedings of the 33rd International Conference on Automata, Languages and Programming. Berlin, Germany: Springer, 2006: 1-12.
[8] DWORK C. Differential privacy: a survey of results[C]//Proceedings of the 5th International Conference on Theory and Applications of Models of Computation. Berlin, Germany: Springer, 2008: 1-19.
[9] DWORK C. A firm foundation for private data analysis[M]. New York, USA: ACM Press, 2011.
[10] 张啸剑,孟小峰. 面向数据发布和分析的差分隐私保护[J]. 计算机学报,2014,37(4):927-949.
[11] YAROSLAVTSEV G, PROCOPIUC C M, CORMODE G, et al. Accurate and efficient private release of datacubes and contingency tables[C]//Proceedings of 2013 IEEE International Conference on Data Engineering. Washington D. C., USA: IEEE Computer Society, 2013: 745-756.
[12] DWORK C, MCSHERRY F, NISSIM K. Calibrating noise to sensitivity in private data analysis[M]. Berlin, Germany: Springer, 2006: 637-648.
[13] ZHANG J, CORMODE G, PROCOPIUC C M, et al. PrivBayes: private data release via bayesian networks[M]. New York, USA: ACM Press, 2014.
[14] XU J, ZHANG Z, XIAO X, et al. Differentially private histogram publication[J]. VLDB Journal, 2013, 22(6): 797-822.
[15] LI N, LI T, VENKATASUBRAMANIAN S. t-closeness: privacy beyond k-anonymity and l-diversity [C]// Proceedings of 2007 IEEE International Conference on Data Engineering. Washington D. C., USA: IEEE Press, 2007: 106-115.
[16] WONG R C, LI J, FU A W, et al. (α, k)-anonymity: an enhanced k-anonymity model for privacy preserving data publishing[C]//Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press, 2006: 754-759.
[17] FRIEDMAN A, SCHUSTER A. Data mining with differential privacy[C]//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press, 2010: 493-502.
[18] 熊平,朱天清,王晓峰. 差分隐私保护及其应用[J]. 计算机学报,2014,37(1):101-122.
[19] MCSHERRY F, TALWAR K. Mechanism design via differential privacy [C]//Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science. Washington D. C., USA: IEEE Press, 2007: 94-103.
[20] SPIRTES P, GLYMOUR C, SCHEINES R. Causality from probability [EB/OL]. [2017-12-30]. https://philpapers.org/rec/SPICFP-2.
[21] COOPER G F, HERSKOVITS E. A Bayesian method for the induction of probabilistic networks from data [J]. Machine Learning, 1992, 9(4): 309-347.
[22] 曾千千,曾安,潘丹,等. 基于最大信息系数的贝叶斯网络结构学习算法[J]. 计算机工程,2017,43(8):225-230.
[23] 王良,王伟平,孟丹. 基于加权贝叶斯网络的隐私数据发布方法[J]. 计算机研究与发展,2016,53(10):2343-2353.

编辑　司森森

(上接第 123 页)

[3] 蔡铭,谢晓玲,王雪畅,等. 智能电网脆弱性分析及对策研究[J]. 信息工程大学学报,2013,14(3):376-379.
[4] LIANG G, WELLER S R, ZHAO J, et al. The 2015 Ukraine blackout: implications for false data injection attacks [J]. IEEE Transactions on Power Systems, 2017, 32(4): 3317-3318.
[5] Communication networks and systems for power utility automation: IEC/TR 61850 [EB/OL]. [2017-11-20]. https://webstore.iec.ch/publication/6028.
[6] LIN H, SLAGELL A, SAUER P W, et al. Semantic security analysis of SCADA networks to detect malicious control commands in power grids [C]//Proceedings of the 1st ACM Workshop on Smart Energy Grid Security. New York, USA: ACM Press, 2013: 29-34.
[7] LIN H, SLAGELL A, KALBARCZYK Z, et al. Runtime semantic security analysis to detect and mitigate control-related attacks in power grids [J]. IEEE Transactions on Smart Grid, 2018, 9(1): 163-178.
[8] PERDISCI R, ARIU D, FOGLA P, et al. McPAD: a multiple classifier system for accurate payload-based anomaly detection [J]. Computerand Telecommunications Networking, 2009, 53(6): 864-881.
[9] 费金龙,王禹,王天鹏,等. 基于云模型的网络异常流量检测[J]. 计算机工程,2017,43(1):178-182.
[10] 凌骏,尹博学,李晟,等. 基于监控数据的 MySQL 异常检测算法[J]. 计算机工程,2015,41(11):41-46.
[11] MANNING C D, SCHÜTZE H. Foundations of statistical natural language processing [M]. Cambridge, USA: MIT Press, 1999.
[12] PACCANARO A, HINTON G E. Learning distributed representations of concepts using linear relational embedding [J]. IEEE Transactions on Knowledge and Data Engineering, 2001, 13(2): 232-244.
[13] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. [S. l.]: Curran Associates Inc., 2013: 3111-3119.
[14] LIU F T, KAI M T, ZHOU Z H. Isolation Forest [C]// Proceedings of the 8th IEEE International Conference on Data Mining. Washington D. C., USA: IEEE Press, 2008: 413-422.
[15] Supervisory control and data acquisition: IEC 60870 [EB/OL]. [2017-11-20]. https://webstore.iec.ch/publication/3755.

编辑　赵　辉