基于函数挖掘的能源信息物理系统数据安全风险识别算法

邓松[1]，蔡清媛[1]，高昆仑[2,3]，张建堂[1]，饶玮[2,3]，朱力鹏[2,3]

(1. 南京邮电大学 先进技术研究院，江苏 南京 210023；2. 全球能源互联网研究院有限公司，北京 102209；
3. 电力系统人工智能(联研院)国家电网公司联合实验室，北京 102209)

摘 要：数据安全风险评估对于能源信息物理系统安全稳定运行至关重要。现有的从二次设备、信息等角度来分析数据安全风险已经无法满足能源信息物理系统广泛的能源接入和各能源之间的能量、信息交互需求。首先提出基于粗糙集的数据安全风险要素特征选择算法，对影响能源信息物理系统中数据的安全风险特征集进行特征选择，降低能源信息物理系统数据安全风险要素集的维度；在此基础上，利用基因表达式编程（gene expression programming, GEP）的函数挖掘特性，提出基于混合 GEP 的能源信息物理系统数据安全风险识别算法，通过设计小生境种群生成策略以及动态自适应变异概率动态调整策略来提高数据安全风险识别的准确率和效率。仿真实验结果表明，所提算法对于复杂高维的能源信息物理系统数据安全风险集的识别和预测具有较高的准确率和较强的实用性，可为下一步制定能源信息物理系统数据安全防护策略提供理论方法支撑。

关键词：基因表达式编程；粗糙集；特征选择；风险识别；能源信息物理系统

DOI：10.11930/j.issn.1004-9649.202007135

0 引言

数据是能源互联网的核心资产，未来能源互联网的数据来源将覆盖能源生产、传输、交易、消费等各个环节，总体呈现来源广泛、规模庞大以及类型复杂的特征[1-3]。然而，能源互联网的开放、互联及共享机制将导致恶意的网络攻击不断，这些恶意网络攻击利用能源互联网中广泛的信息物理系统间的耦合而产生交互传播的跨空间、跨系统、跨平台的连锁反应，从而不可避免地对能源互联网生产、传输、交易及消费各环节业务系统数据在采集、传输、存储、处理、交换和销毁等全生命周期过程中的安全性产生极大的威胁。

和智能电网相比，能源互联网具有更加开放的信息网络，更多层面的数据来源；同时由于开放的信息网络通道使得发生网络攻击的频率越来越频繁，范围越来越广，这都给能源互联网下的数据安全防御提出了更高的要求。及时有效的数据安全风险识别和评估是能源信息物理系统下数据总体安全防护的基础。国内外众多学者对此开展了深入研究。文献 [4] 提出一种基于攻击预测的电力 CPS 风险评估方法，通过 IEEE 33 节点仿真验证了系统的可行性和有效性。文献 [5] 构建了智慧城市信息安全风险评估指标体系，利用贝叶斯网络对中国 20 个智慧城市试点地区的信息安全风险进行量化评估。文献 [6] 将层次分析法（AHP）引入到风险评估机制中，设计出一种基于模糊数学的新型信息安全风险评估模型。文献 [7] 提出了基于云计算的船舶通信网络安全风险评估模型。文献 [8] 利用可拓识别方法对高校信息系统的安全性进行风险综合评估。文献 [9] 利用小波神经网络算法有效解决集合信息存在的虚假相关，能有效提高信息安全风险评估精度。文献 [10] 基于等级测评和风险评估相结合的理论，将信息风险和对应的风险等级建立连接。文献 [11] 利用层次分析法建立风险评估层次分析模

收稿日期：2020-07-27；**修回日期：**2021-01-18。

基金项目：国家自然科学基金资助项目(网络攻击下能源互联网数据容侵评估及可靠存储机制研究，51977113；面向有源配电网的数据传输优化及智能过滤机制，51507084)。

型。文献 [12] 基于 D-S 证据理论确定各指标体系的权重，实验表明组合规则显著提高了网络风险水平的可靠性。文献 [13] 提出一种基于 D 数层次分析法（D-AHP）与灰色理论的信息安全风险评估方法。文献 [14] 提出一种基于模糊层次分析法的电力边缘计算信息系统安全风险评估方法。文献 [15] 提出了一种基于全概率公式和条件风险价值的风险度量。文献 [16] 通过改进型 AHP 与证据理论来规避评估过程中的主观性和不确定性。文献 [17] 基于贝叶斯网络建立了一个风险概率传递关系模型。文献 [18] 运用网络分析法和灰色统计理论确定各威胁指标的灰数及信息系统风险等级。文献 [19] 以个性化、及时和连续的方式评估和交流用户和系统层面的风险。文献 [20] 构建了基于量子门线路神经网络的信息安全风险评估模型。但以上研究仅针对信息系统安全风险评估开展研究，鲜有针对数据安全风险评估的研究。文献 [21] 通过动态加密技术，有效保护主动配电网中各个分布式通信参与者的数据隐私性。文献 [22] 提出了面向能源互联网的数据一致性框架和协议。文献 [23-26] 梳理出针对智能电网的数据注入攻击以及数据完整性攻击形式，详细地给出相应的防护策略。

通过分析上述文献发现：现有针对能源互联网或者电力信息物理融合系统的数据安全防护技术较多，但也仅限于数据一致性、数据注入及完整性攻击方面的论述。目前鲜有文献对能源信息物理系统下数据安全风险进行量化识别和评估，从而无法对各类针对数据的网络攻击提供有效的分析和防御决策。同时现有的基于人工智能的安全风险评估方法都是定性分析，无法定量评估安全风险。基因表达式编程（gene expression programming, GEP）是一类定量挖掘样本数据函数模型的进化算法[27]。因此，本文利用 GEP 强大的函数挖掘特性，提出基于混合 GEP 的能源信息物理系统数据安全风险识别算法（data security risk recognition algorithm for energy cyber physics system based on hybrid gene expression programming, DSRR-HGEP），仿真实验结果表明本文所提的 DSRR-HGEP 算法具有较高的数据安全风险识别的准确率和效率。

1 基于粗糙集的数据安全风险要素特征选择算法

为了降低后期能源信息物理系统数据安全风险识别模型挖掘的复杂度，首先需要做的就是对整个影响能源信息物理系统中数据安全的风险要素进行特征选择，在不影响风险识别准确率的前提下，保留最少的数据安全风险要素。与传统的主成分分析和奇异值分解等特征提取方法相比，基于粗糙集的特征选择在降维的同时，还不影响降维后原始数据的决策。为此，本文提出基于粗糙集的数据安全风险要素特征选择算法（feature selection algorithm of data security risk features based on rough set, FSDSRF-RS）。

为了更好地描述问题，首先给出有关 FSDSRF-RS 算法中的相关定义。

（1）定义 1。设$S=<U,C\cup D,V,f>$，$C\cup D=R$，$V=v_r\cup y_r, r\in R$，$f:U\times R\to V$，则称满足上述条件的S为能源互联网数据安全风险决策表。其中U为能源信息物理系统中所有数据安全风险要素及风险等级值的集合；$C=\{c_i\}, i\in[1,n]$为影响能源互联网数据安全风险要素集的条件属性集合（包括防火墙、入侵检测、加解密、访问控制等）；$D=\{d_i\}, i\in[1,m]$为影响能源互联网数据安全风险要素集中的风险等级集合（包括低、中、高 3 个等级）；V是影响能源互联网数据安全风险要素集中各类风险值及对应的风险等级值的集合，v_r表示条件属性集合C中任意一个数据安全风险要素的取值，y_r表示风险等级集合D的取值；f表示U中每一对象x的属性值，即对于$\forall r\in R, x\in U$，有$f(x,r)\in v_r\cup y_r$。

图 1 给出了能源互联网数据安全风险决策表的示意。

（2）定义 2。设能源互联网数据安全风险决策表$S=<U,C\cup D,V,f>$，其中$C\cup D=R$，对于$\forall P\subseteq R$，且$x,y\in U$，当且仅当对于$\forall r\in P$，$f(x,r)=f(y,r)$时，称能源互联网数据安全风险要素及风险等级集合U中的对象x和y是不可分辨的，记为$\mathrm{IND}(P)=\{(x,y)\in U|\forall r\in P, f(x,r)=f(y,r)\}$或$U/R$。

（3）定义 3。设能源互联网数据安全风险决策表$S=<U,C\cup D,V,f>$，若$U/C=U/(C-c_i)$，则称影响能源互联网数据安全风险要素集的条件属

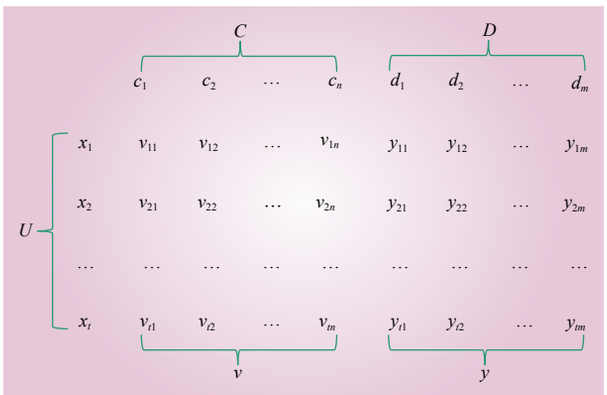

图 1　能源互联网数据安全风险决策表示意

Fig. 1　Data security risk decision-making diagram for energy Internet

性集合中某一要素c_i可约简。

整个 FSDSRF-RS 算法描述如下。

输入：$S=<U,C\cup D,V,f>$；

输出：C'；

1. $A\leftarrow U/C$；
2. for $1\leftarrow i$ to $|C|$ do
3. $\quad B\leftarrow U/(C-\{c_i\})$；
4. $\quad$if $A=B$ then
5. $\quad\quad C\leftarrow C-\{c_i\}$；
6. $\quad$end if
7. end for
8. $C'\leftarrow C$；
9. return C'；

2　基于混合 GEP 的能源信息物理系统数据安全风险识别算法

从定量分析的角度，数据安全风险识别可以理解为挖掘影响数据安全的风险因素和安全风险类型之间的函数模型。基因表达式编程是一种智能化、自动化的函数模型挖掘算法，因此在 FSDSRF-PCA 算法基础上，本文提出的 DSRR-HGEP 算法无须事先设置函数模型，直接通过对影响数据安全的风险因素进行基因编码，同时借助相应的生物进化操作最终挖掘出相应的数据安全风险识别函数模型，以此模型来识别数据安全风险。

2.1　基于小生境的 GEP 种群生成策略

在自然界中，小生境（Niche）是指特征相似的种群聚集在一起，并在同类中交配繁衍后代，在基因表达式编程算法中，各类遗传操作是基于一定概率随机的，这种方式在算法初始阶段的确保持了种群的多样性，但在进化到一定代数后，大量个体的适应度值都会集中在某一个局域，从而后代会造成近亲繁殖，大大降低种群的多样性。

因此，本文将小生境技术运用到基因表达式编程中，提出基于小生境的 GEP 种群生成策略（population generation for GEP based on niche, PG-NGEP）。其基本思想是：首先计算 GEP 初始种群中所有个体的适应度值，从中选择前 K 个最大适应度值的个体组成小生境；然后在小生境的所有个体中两两计算海明距离，并基于该距离动态调整适应度值较小的个体，使得该个体能被遗传到下一代的概率大大降低；最后对所有调整后的个体适应度进行排序，产生下一代种群，循环往复，直到算法结束。

2.2　基于种群密度的变异概率自适应调整策略

变异概率的选择会直接影响 GEP 算法的收敛性。变异概率过小，GEP 算法不易产生新的个体，种群多样性会受到很大影响；变异概率过大，GEP 算法就变为纯粹的随机搜索算法。因此如何选择一个适当的变异概率值对于 GEP 挖掘数据安全风险识别函数模型至关重要。

从生物进化的角度来看，种群中个体越密集，则进化出新物种的概率就越小。因此，本文提出基于种群密度的变异概率自适应调整策略（adaptive adjustment of mutation probability based on population density, AAMP-PD）。

设当前种群的最大适应度值为$f_{\max}$，平均适应度值为f_{avg}，AAMP-PD 算法中，变异概率P_{m}的调整策略可表示为

$$P_{\text{m}}=\begin{cases}\alpha P_{\text{m}}, & \chi f_{\max}<f_{\text{avg}}\\ \beta P_{\text{m}}, & \text{其他}\end{cases}\qquad(1)$$

式中：$2\leqslant\alpha\leqslant5$；$0<\beta<0.5$；$0.5<\chi<1$。

当$\chi f_{\max}<f_{\text{avg}}$时，表明当前种群中个体较为集中，容易陷入局部最优，通过增加变异概率P_{m}来使得个体更加多样化；否则，则表明当前种群中个体较为分散，通过减小变异概率P_{m}来保持个体多样化，避免陷入局部最优。

2.3　DSRR-HGEP

针对能源信息物理融合系统中的数据安全风险识别的目标是识别能源生产、传输、交易及消

费过程中信息物理系统自身及交互所面临的数据威胁。为了全局掌握能源信息物理系统中多维度数据的安全态势，本文提出混合 GEP 的能源信息物理系统数据安全风险识别算法（DSRR-HGEP），利用基因表达式编程算法来挖掘针对能源信息物理系统数据安全的风险要素与风险等级之间的复杂函数关系模型，定量识别能源信息物理系统下数据安全风险等级。

为了更好理解 GEP 挖掘能源信息物理系统下数据安全风险要素与风险等级之间的函数关系，首先给出如下定义。

定义 4：设函数集F包含基本初等数学函数，终端集$T=\{d_1,d_2,\cdots,d_m\}$，则称$D_g=\langle F,T|h,t\rangle$为能源信息物理系统数据安全风险识别基因。其中$d_i,i\in[1,m]$表示影响能源信息物理系统下数据安全风险要素，$h$、$t$分别表示为上述基因的头长和尾长，二者之间的关系为

$$t=h(n-1)+1 \tag{2}$$

式中：n表示函数集F中初等函数所包含的最大运算操作目数。例如，初等函数为$+,-,*,/$等时，$n=2$；初等函数为$\sin,\cos,\log,\exp$等时，$n=1$。

一个或多个D_g构成能源信息物理系统数据安全风险识别染色体。

整个算法描述如下所示。

输入：Pop, popSize, maxGen, P_s, P_m, P_t, P_r, δ,α,β,χ；

输出：bestFunction；

1. Pop ← InitPop(popSize, Data);

2. f ← CalFit(Pop, popSize);

3. Pop ← PG − NGEP(f, Pop, popSize, maxGen, P_s, P_m, P_t, P_r, δ)；

4. f ← CalFit(Pop, popSize);

5. $f_{\max}$ ← max(f);

6. f_{avg} ← Avg(f);

7. while gen < maxGen do

8. Pop ← Select(P_s, Pop);

9. Pop ← Mutate(AAMP − PD($f_{\max}$, f_{avg}, α,β,χ), Pop);

10. Pop ← ISTransposition(P_t, Pop);

11. Pop ← RISTransposition(P_t, Pop);

12. Pop ← GeneTransposition(P_t, Pop);

13. Pop ← OnePointRecombination(P_r, Pop);

14. Pop ← TwoPointRecombination(P_r, Pop);

15. Pop ← GeneRecombination(P_r, Pop);

16. f ← CalFit(Pop, popSize);

17. $f_{\max}$ ← max(f);

18. f_{avg} ← avg(f);

19. end while

20. return bestFunction

3 仿真实验与结果分析

为了更好地验证本文所提出算法的可行性和有效性，在实验室环境下做了相应的仿真实验。其中数据安全风险要素特征选择基于 Python 实现，实验平台为 Win10 + Python 3.7+ PyCharm 2019.2.2；数据安全风险识别模型挖掘基于 Java 实现，实验平台为 Win10 + Eclipse 3.2+Java1.8。

本实验数据以电网业务系统中数据安全风险来模拟，假设电网业务系统中数据主要考虑传输数据机密性破坏、传输数据完整性破坏以及传输数据被篡改等几个方面的安全风险，并结合边界、网络、主机及应用等 4 个方面构建如表 1 所示的用电信息采集系统数据安全风险要素集。

根据表 1 给出的数据安全风险要素集，结合网络安全日志文件，并通过量化后生成相应的仿

表 1 电网业务系统数据安全风险要素集

Table 1 Data security risk element set of power grid business system

一级指标	二级指标
边界（O_1）	防火墙系统（O_{11}），入侵检测系统（O_{12}），网闸（O_{13}），VLAN间访问控制技术（O_{14}），用户访问控制策略（O_{15}）
网络（O_2）	建立VPN网络（O_{21}），路由器、交换机加固措施（O_{22}），安全认证芯片加密（O_{23}），入侵检测系统（O_{24}），恶意代码防范（O_{25}），安全弱点扫描（O_{26}）
主机（O_3）	操作系统安全加固（O_{31}），防窃、防破坏安全措施（O_{32}），安全认证机制（O_{33}），数据加解密（O_{34}），主机病毒防护（O_{35}），弱点扫描（O_{36}）
应用（O_4）	主机加解密（O_{41}），安全芯片（O_{42}），密钥管理（O_{43}），用户数据安全防护（O_{44}）

真实验数据集。该数据集共包括 30 条实验数据，其中 21 个条件特征，1 个风险等级特征，数据安全风险分为低、中、高 3 个等级。整个实验数据集分为训练数据集（前 20 条数据）和测试数据集（后 10 条数据）。表 2 给出实验数据集描述。

表 2　实验数据集描述

Table 2　Description of experimental dataset

数据集名称		条件属性个数（安全风险要素个数）	决策属性个数（安全风险等级个数）	数据规模
RData	TrainRData	21	3	30
	TestRData			

（1）实验 1：针对表 2 中给出的实验数据集，表 3 给出 FSDSRF-RS、主成分分析法（principal component analysis, PCA）、互信息法（mutual information，MI）、随机森林（random forest, RF）以及方差过滤（variance threshold, VR）进行特征选择前后条件属性个数变化。表 4 显示上述 3 种特征选择算法最后的结果。

从表 3 可以看出，与特征选择前相比，基于 FSDSRF-RS 算法的特征选择后的条件属性个数减少了 76.19%。与 PCA（按信息量保存 65%，95% 和 98%），MI，RF 以及 VR 算法相比，基于 FSDSRF-RS 算法的特征选择后的条件属性个数分别减少了 28.57%，66.67%，70.59%，54.55%，50%，37.5%。由此可见，针对表 2 中所示的实验数据，FSDSRF-RS 算法是有效的。同时，从表 4 可以看出，特征选择后，FSDSRF-RS 和 MI 算法

表 3　基于5 种算法的特征选择前后条件属性个数变化

Table 3　The number of conditional attributes before and after feature selection based on FSDSRF-RS, PCA, MI, RF and VR

算法名称	特征选择前的条件属性个数	特征选择后的条件属性个数
FSDSRF-RS	21	5
PCA（按信息量保存 65%计算）	21	7
PCA（按信息量保存 95%计算）	21	15
PCA（按信息量保存 98%计算）	21	17
MI	21	11
RF	21	10
VR	21	8

表 4　实验数据集描述

Table 4　Description of experimental dataset

算法名称	特征选择结果
FSDSRF-RS	O_{11}，O_{14}，O_{36}，O_{41}, O_{42}
PCA	—
MI	O_{11}，O_{13}，O_{14}，O_{15}，O_{23}，O_{24}，O_{26}，O_{33}，O_{42}，O_{43}，O_{44}
RF	O_{13}，O_{14}，O_{15}，O_{21}，O_{23}，O_{24}，O_{26}，O_{31}，O_{36}，O_{41}，O_{42}
VR	O_{11}，O_{12}，O_{13}，O_{15}，O_{21}，O_{24}，O_{34}，O_{36}

所保留的条件属性中 3 个相同的条件属性，FSDSRF-RS 算法所保留的条件属性中全部都在基于 RF 特征选择的结果中，这也说明不同的特征选择算法可以选择出对数据安全风险等级识别最关联的条件属性；而 PCA 由于属于特征的线性组合内容，基于 PCA 的特征选择结果不是原有的条件属性，而是原有条件属性之间的组合。

（2）实验 2：在实验 1 的基础上，针对约简后形成的能源信息物理系统数据安全风险决策表，本实验阐述了 DSRR-HGEP 的性能。整个实验中 DSRR-HGEP 的参数如表 5 所示。图 2 给出了重复 5 次实验中，特征选择前后的能源信息物理系统数据安全风险决策表进行函数挖掘时的最优适应度与最大适应度值差值的比较。图 3 显示了在重复 5 次实验，每次实验算法运行 10 次的条件下，特征选择前后数据安全风险识别函数挖掘得到最优解的耗时比较。图 4 比较了传统 GEP 算法和 DSRR-HGEP 算法的收敛速度。图 5 显示了基

表 5　DSRR-HGEP 参数

Table 5　Parameters of DSRR-HGEP

变量名称	变量值	变量名称	变量值
函数集	+-*/SCT	一点交叉率	0.3
变量集	x_0, x_1, x_2, x_3, x_4	两点交叉率	0.3
连接函数	+	Gene交叉率	0.1
基因个数	3	适应度函数	$f_i = \sum_{j=1}^{n}(100 - \lvert P_{ij} - T_j \rvert)$
基因头长	9	运行代数	30 000
种群大小	5 000	δ	0.5
变异率	0.044	α	3.0
IS插串率	0.3	β	0.27
RIS插串率	0.3	χ	0.65
Gene插串率	0.1		

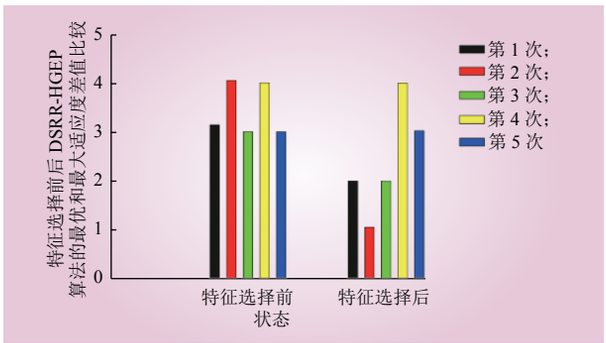

图 2 特征选择前后最优适应度值与最大适应度值差值比较

Fig. 2 Comparison of the difference between the optimal fitness value and the maximum fitness value before and after feature selection

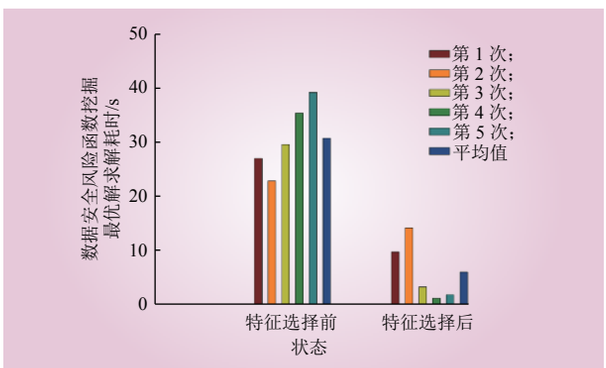

图 3 特征选择前后数据安全风险识别函数挖掘得到最优解的耗时比较

Fig. 3 Time-consuming comparison of data security risk identification function mining to obtain the optimal solution before and after feature selection

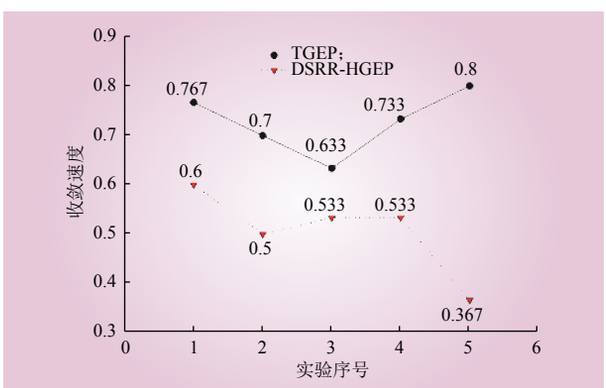

图 4 GEP 算法和 DSRR-HGEP 算法的收敛速度比较

Fig. 4 Comparison of convergence speed between traditional GEP and DSRR-HGEP

于 DSRR-HGEP 挖掘得到的数据安全风险识别函数模型对特征选择前后测试数据的模型值与真实值之间的比较。

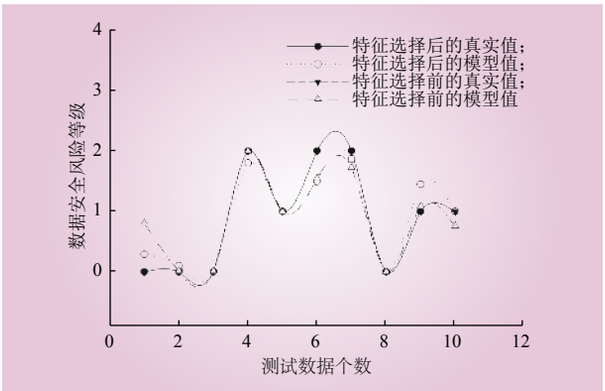

图 5 特征选择前后测试数据真实值与模型值比较

Fig. 5 Comparison between real value and model value for testing data before and after feature selection

从图 2 可以看出，针对表 2 所示的训练数据集，与特征选择前相比，特征选择后基于 DSRR-HGEP 算法进行数据完全风险识别函数挖掘所得到的最优适应度值与最大适应度值差值最大为 64.92%。这表明针对高维数据安全风险数据集，在不改变现有该数据集风险决策能力的前提下，特征选择大大提高数据完全风险识别函数挖掘的成功率。同时与传统的 GEP 算法相比，DSRR-HGEP 算法中所采用的小生境种群生成以及动态自适应变异概率动态调整策略也大大加速了算法收敛。与此同时，图 3 显示，针对表 2 所示的训练数据集，特征选择大大降低了数据完全风险识别函数挖掘的平均耗时，5 次相同参数的实验中平均耗时最大下降 80.33%。

同时为了验证 DSRR-HGEP 算法比传统 GEP 算法（traditional GEP, TGEP）的性能要优越，本文还比较 2 种算法运行 5 次的收敛速度。设 f_{max} 为对应当前样本数据下 TGEP 和 DSRR-HGEP 算法的最大适应度值，N 为 TGEP 和 DSRR-HGEP 算法的最大运行代数，P 为 TGEP 和 DSRR-HGEP 算法运行到最优解 f_{opt} 时所对应的运行代数，则 $C_s = \frac{P}{N}$ 表示 TGEP 和 DSRR-HGEP 算法的收敛速度。在保证求解到问题最优解的条件下，C_s 越小，表明算法的收敛速度越快。从图 4 可以看出，5 次实验运行过程中，DSRR-HGEP 算法的收敛速度都要优于 TGEP 算法。这也说明 DSRR-HGEP 算法中采用的小生境种群生成以及动态自适应变异概率动态调整策略大大加速了算法收敛，从而加快了求解出最优解的速度，减少了算法的耗时。

图 5 反映了特征选择前后测试数据真实值与模型值之间的拟合程度。从图 5 可以看出，特征选择前测试数据真实值与模型值之间最大的误差为 0.81，最小为 0。而特征选择后真实值与模型值之间最大的误差为 0.49，最小为 0.000 8。由此可以看出该模型具有较高的预测精度。

4　结语

为了更好地处理能源互联网下数据安全风险识别，本文提出了基于混合 GEP 的能源信息物理系统数据安全风险识别算法。首先构建能源信息物理系统数据安全风险决策表，并基于粗糙集对该数据安全风险决策表进行特征选择；在此基础上通过构建小生境种群生成策略以及动态自适应变异概率动态调整策略来构建基于混合基因表达式编程的数据安全风险识别模型。仿真实验表明本文所提出算法具有较强的高维数据处理能力以及数据安全风险识别准确率和预测精度。

数据安全风险识别是制定数据安全防护策略的前提和基础，本文研究工作可为能源互联网数据安全防护提供方法支撑。实际中能源互联网数据来源广泛，数据类型复杂，数据量较大，影响数据安全的风险要素众多，为了能对能源互联网数据实现全生命周期的安全防护，下一步将从数据采集、传输、存储、应用等角度梳理数据安全风险要素，构建一个数据全生命周期的数据安全风险要素集，并通过 UML 建模和关联分析的方法分析每一个数据安全风险要素之间的关系。

参考文献：

[1] 王继业, 郭经红, 曹军威, 等. 能源互联网信息通信关键技术综述 [J]. 智能电网, 2015, 3(6): 473–485.
WANG Jiye, GUO Jinghong, CAO Junwei, *et al*. Review on information and communication key technologies of energy Internet[J]. Smart Grid, 2015, 3(6): 473–485.

[2] 袁智勇, 赵懿祺, 郭祚刚, 等. 面向能源互联网的综合能源系统规划研究综述 [J]. 南方电网技术, 2019, 13(7): 1–9.
YUAN Zhiyong, ZHAO Yiqi, GUO Zuogang, *et al*. Research summary of integrated energy systems planning for energy Internet[J]. Southern Power System Technology, 2019, 13(7): 1–9.

[3] 王继业, 孟坤, 曹军威, 等. 能源互联网信息技术研究综述 [J]. 计算机研究与发展, 2015, 52(5): 1109–1126.
WANG Jiye, MENG Kun, CAO Junwei, *et al*. Information technology for energy Internet: a survey[J]. Journal of Computer Research and Development, 2015, 52(5): 1109–1126.

[4] 韩丽芳, 胡博文, 杨军, 等. 基于攻击预测的电力 CPS 安全风险评估 [J]. 中国电力, 2019, 52(1): 48–56.
HAN Lifang, HU Bowen, YANG Jun, *et al*. A new security risk assessment method for cyber physical power system based on attack prediction[J]. Electric Power, 2019, 52(1): 48–56.

[5] 毛子骏, 梅宏, 肖一鸣, 等. 基于贝叶斯网络的智慧城市信息安全风险评估研究 [J]. 现代情报, 2020, 40(5): 19–26, 40.
MAO Zijun, MEI Hong, XIAO Yiming, *et al*. Risk assessment of smart city information security based on Bayesian network[J]. Journal of Modern Information, 2020, 40(5): 19–26, 40.

[6] 梁智强, 林丹生. 基于电力系统的信息安全风险评估机制研究 [J]. 信息网络安全, 2017(4): 86–90.
LIANG Zhiqiang, LIN Dansheng. Information security risk assessment mechanism research based on power system[J]. Netinfo Security, 2017(4): 86–90.

[7] ZHOU H Z, YU G, LI L G. Cloud communication based ship communication network security risk assessment model[J]. Journal of Coastal Research, 2020, 95(S1): 991.

[8] 王丰, 张春平, 林瑜, 等. 军事院校信息系统安全风险的可拓识别评估 [J]. 武汉理工大学学报(信息与管理工程版), 2018, 40(6): 606–609.
WANG Feng, ZHANG Chunping, LIN Yu, *et al*. Extension identification assessment of information system security risk in military academies[J]. Journal of Wuhan University of Technology (Information & Management Engineering), 2018, 40(6): 606–609.

[9] 王皓然, 严彬元. 依赖小波神经网络算法的信息安全风险评估方法 [J]. 信息技术, 2018, 42(12): 93–96.
WANG Haoran, YAN Binyuan. Information security risk assessment method based on wavelet neural network algorithm[J]. Information Technology, 2018, 42(12): 93–96.

[10] 任贝贝. 一种风险评估和等级防护相结合的信息风险预测系统 [J]. 计算机应用与软件, 2019, 36(2): 151–154.
REN Beibei. An information risk prediction system combining risk assessment and level protection[J]. Computer Applications and Software, 2019, 36(2): 151–154.

[11] 柴继文, 王胜, 梁晖辉, 等. 基于层次分析法的信息安全风险评估要素量化方法 [J]. 重庆大学学报, 2017, 40(4): 44–53.
CHAI Jiwen, WANG Sheng, LIANG Huihui, *et al*. An AHP-based

quantified method of information security risk assessment elements[J]. Journal of Chongqing University, 2017, 40(4): 44–53.

[12] YU J J, HU M, WANG P. Evaluation and reliability analysis of network security risk factors based on D-S evidence theory[J]. Journal of Intelligent & Fuzzy Systems, 2018, 34(2): 861–869.

[13] 许硕, 唐作其, 王鑫. 基于 D-AHP 与灰色理论的信息安全风险评估 [J]. 计算机工程, 2019, 45(7): 194–202.
XU Shuo, TANG Zuoqi, WANG Xin. Information security risk assessment based on D-AHP and grey theory[J]. Computer Engineering, 2019, 45(7): 194–202.

[14] 詹雄, 郭昊, 何小芸, 等. 国家电网边缘计算信息系统安全风险评估方法研究 [J]. 计算机科学, 2019, 46(增刊2): 428–432.
ZHAN Xiong, GUO Hao, HE Xiaoyun, *et al*. Research on security risk assessment method of state grid edge computing information system[J]. Computer Science, 2019, 46(S2): 428–432.

[15] 殷加玦, 赵冬梅. 基于全概率风险度量的电力系统备用风险评估方法 [J]. 电力自动化设备, 2020, 40(1): 156–162.
YIN Jiafu, ZHAO Dongmei. Reserve risk assessment method of power system based on total probability risk measure[J]. Electric Power Automation Equipment, 2020, 40(1): 156–162.

[16] 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(1): 49–55.
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(1): 49–55.

[17] 时召伟, 魏松杰. 基于贝叶斯网络的 Android 应用风险评估的研究 [J]. 合肥工业大学学报(自然科学版), 2020, 43(6): 753–757.
SHI Zhaowei, WEI Songjie. Research on Android application risk assessment based on Bayesian network[J]. Journal of Hefei University of Technology (Natural Science), 2020, 43(6): 753–757.

[18] 赵刚, 吴天水. 结合灰色网络威胁分析的信息安全风险评估 [J]. 清华大学学报(自然科学版), 2013, 53(12): 1761–1767.
ZHAO Gang, WU Tianshui. Information security risk assessment based on G-ANP[J]. Journal of Tsinghua University (Science and Technology), 2013, 53(12): 1761–1767.

[19] ALOHALI M, CLARKE N, FURNELL S. The design and evaluation of a user-centric information security risk assessment and response framework[J]. International Journal of Advanced Computer Science and Applications, 2018, 9(10): 148–163.

[20] 周超, 潘平, 黄亮. 基于量子门线路神经网络的信息安全风险评估 [J]. 计算机工程, 2018, 44(12): 39–45.
ZHOU Chao, PAN Ping, HUANG Liang. Risk assessment of information security based on quantum gate circuit neural networks[J]. Computer Engineering, 2018, 44(12): 39–45.

[21] 董朝阳, 陈莹莹, 罗逢吉. 未来主动配电网中的新型数据驱动应用: 技术, 展望与挑战 [J]. 电力建设, 2017, 38(5): 2–10.
DONG Zhaoyang, CHEN Yingying, LUO Fengji. Innovative data-driven applications in future active distribution network: technologies, prospect and challenges[J]. Electric Power Construction, 2017, 38(5): 2–10.

[22] 杨英仪. 面向能源互联网的数据一致性框架 [J]. 广东电力, 2017, 30(12): 22–28.
YANG Yingyi. A data consensus framework for energy Internet[J]. Guangdong Electric Power, 2017, 30(12): 22–28.

[23] KHANNA K, PANIGRAHI B K, JOSHI A. Bi-level modelling of false data injection attacks on security constrained optimal power flow[J]. IET Generation, Transmission & Distribution, 2017, 11(14): 3586–3593.

[24] 王电钢, 黄林, 刘捷, 等. 考虑负荷虚假数据注入攻击的电力信息物理系统防御策略 [J]. 电力系统保护与控制, 2019, 47(1): 28–34.
WANG Diangang, HUANG Lin, LIU Jie, *et al*. Cyber-physical system defense strategy considering loaded false data injection attacks[J]. Power System Protection and Control, 2019, 47(1): 28–34.

[25] YANG Q Y, LI D H, YU W, *et al*. Toward data integrity attacks against optimal power flow in smart grid[J]. IEEE Internet of Things Journal, 2017, 4(5): 1726–1738.

[26] LIU X, BAO Z, LU D, *et al*. Modeling of local false data injection attacks with reduced network information[J]. IEEE Transactions on Smart Grid, 2015, 6(4): 1686–1696.

[27] FERREIRA C. Gene expression programming: a new adaptive algorithm for solving problems[J]. Complex Systems, 2001, 13(2): 87–129.

作者简介：

邓松 (1980—)，男，博士，副研究员，从事电网信息安全与防护，电力大数据及数据挖掘研究，E-mail: ds16090311@163.com；

蔡清媛 (1997—)，女，硕士研究生，从事电网信息安全与防护、电力大数据及数据挖掘研究，E-mail: dmccxysc@163.com；

高昆仑 (1972—)，男，博士，高级工程师 (教授级)，从事电力系统自动化与信息化技术研究，E-mail：gkl@geiri.sgcc.com.cn。

（责任编辑　李博）

(下转第 37 页)

Application of Formal Methods in Power Grid Cyber Physical Systems

HUANG Li[1,2], LIANG Yun[1,2], HUANG Hui[1,2], ZHAO Ruohan[3]

(1. Global Energy Interconnection Research Institute Co., Ltd., Beijing 102209, China; 2. Laboratory of Electric Power Intelligent Sensing Technology and Application, Beijing 102209, China; 3. State Grid Electric Power Research Institute, Nanjing 210032, China)

Abstract: The embedded terminals in the power grid cyber physical systems need to not only have the ability of information interaction, but also meet the real-time requirements of measurement and control under resource constraints. It is necessary to introduce formal methods to verify the reliability of complex systems with large scale. This paper analyzes the application of formal methods in power grid cyber physical systems, designs and realizes a formal method and model checking software tool for analyzing the information interaction process of embedded system in the power grid cyber physical systems. The application process of the model checking tool is analyzed in detail through a practical case, and the application results show that the formal method can shorten the distance from high level design to code implementation and improve the reliability of the products. The model checking software tool can provide a reference solution to the reliability assurance problem caused by the rapid increase of embedded devices in terms of scale and complexity.

This work is supported by the National Key Research and Development Program of China(Basic Theories and Methods of Analysis and Control of the Cyber Physical Systems for Power Grid, No.2017YFB0903000).

Keywords: power grid cyber physical systems; formal method; resource constraints; information interaction; model checking

(上接第 30 页)

Data Security Risk Recognition Algorithm for Energy Cyber Physics System Based on Function Mining

DENG Song[1], CAI Qingyuan[1], GAO Kunlun[2,3], ZHANG Jiantang[1], RAO Wei[2,3], ZHU Lipeng[2,3]

(1. Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; 2. Global Energy Interconnection Research Institute Co., Ltd., Beijing 102209, China; 3. Artificial Intelligence on Electric Power System State Grid Corporation Joint Laboratory (GEIRI), Beijing 102209, China)

Abstract: Data security risk assessment is essential for the safe and stable operation of energy cyber physics system (CPS). The existing data security risk analysis from the perspective of secondary equipment and information cannot meet the requirements for extensive energy access as well as energy and information interaction between various energy sources in the energy CPS. Firstly, a feature selection algorithm for data security risk elements based on rough set (FSDSRF-RS) is proposed to select the data security risk feature sets in the energy CPS, consequently reducing the dimensions of the data security risk element sets of the energy CPS. And then, a data security risk recognition algorithm for energy cyber physics system based on hybrid gene expression programming (DSRR-HGEP) is proposed. In the DSRR-HGEP, a niche-based population generation strategy and a dynamic adaptive mutation probability adjustment strategy are designed to improve the accuracy and efficiency of data security risk identification. Simulation and experimental results show that the proposed algorithm in this paper has a high recognition and prediction accuracy for the complex and high-dimensional data security risk sets in the energy CPS, and can provide a theoretical support for formulating data security protection strategies of the energy cyber physical system in the future.

This work is supported by the National Natural Science Foundation of China (Data Tolerance Intrusion Assessment and Reliable Storage for Energy Internet under Cyber Attacks, No.51977113, Data Transmission Optimization and Intelligent Filtering Mechanism for Active Distribution Network, No.51507084).

Keywords: gene expression programming; rough set; feature selection; risk recognition; energy cyber physics system