

基于 CART 回归树的高校网络安全等级保护结论判定方法

庄恩泽^{1,2} 朱 铨^{1,2} 赵力立^{1,2} 林 玲^{1,2} 赖宏图³

¹(福建理工大学信息化建设与管理中心 福建 福州 350118)

²(福建理工大学未来智慧校园研究中心 福建 福州 350118)

³(福建理工大学计算机科学与数学学院 福建 福州 350118)

摘 要 为解决当前高校等级保护测评工作中存在主观因素、容易影响评判结论的问题,提出基于 CART 回归树的安全等级保护结论判定方法。依据网络安全等级保护的相关法规指南,结合现有信息系统的测评结论构建训练数据集;使用主成分分析法进行降维以消除各维度间的相关性;基于降维后的数据集构建 CART 回归树并进行预剪枝以防止过拟合;将生成的决策树用于高校网络安全等级保护的结论判定。结合某高校正在运行的等级保护对象进行实例分析,验证了该模型的可行性。

关键词 网络安全 等级保护 数字化校园 回归树

中图分类号 TP393 文献标志码 A DOI:10.3969/j.issn.1000-386x.2026.07.049

A REGRESSION TREE-BASED METHOD OF CONCLUSION ASSESSMENT IN CLASSIFIED PROTECTION OF CYBERSECURITY IN UNIVERSITY

Zhuang Enze^{1,2} Zhu Quan^{1,2} Zhao Lili^{1,2} Lin Ling^{1,2} Lai Hongtu³

¹(Center for Information Development and Management, Fujian University of Technology, Fuzhou 350118, Fujian, China)

²(Future Wisdom Campus Research Center, Fujian University of Technology, Fuzhou 350118, Fujian, China)

³(School of Computer Science and Mathematics, Fujian University of Technology, Fuzhou 350118, Fujian, China)

Abstract In order to reduce the influence of subjective factors which are easy to affect the conclusion assessment in current classified protection of cybersecurity in colleges and universities, a regression tree-based method of conclusion assessment in classified protection of cybersecurity is proposed. According to the relevant laws and regulations of classified protection of cybersecurity, combined with the evaluation results of the existing information system, the training data set was constructed. We reduced dimensions of training data set using PCA method to eliminate the correlation between dimensions. A regression tree was constructed based on the dimensionally-reduced data set and pre-pruned to prevent overfitting. The generated decision tree was used to assess the conclusion of classified protection of cybersecurity in universities. An example of a university was given to illustrate the practicability of the model.

Keywords Cybersecurity Classified protection Digital campus Regression tree

0 引 言

近年来,随着高校信息化建设的不断推进,如何保障校园基础网络和重要信息系统安全,维护师生利益和社会稳定,成为迫切需要解决的重点问题。网络安

全等级保护为我国保护关键信息基础设施、保障网络和信息安全的通行做法,故现阶段各高校均结合自身信息化建设实际情况,开展等级保护定级及测评工作。2019 年 12 月,《信息安全技术网络安全等级保护基本要求》(GB/T 22239—2019)^[1]、《信息安全技术网络安全等级保护测评要求》(GB/T 28448—

收稿日期:2023-04-27。2021 年度福建省中青年骨干教师(高校信息化专项)科研课题(JAT211007);福建工程学院博士科研启动基金项目(GY-Z15009)。庄恩泽,助理实验师,主研领域:数字化校园,教育信息化。朱铨,高工。赵力立,工程师。林玲,实验师。赖宏图,教授级高工。

2019)^[2]、《信息安全技术网络安全等级保护安全技术要求》(GB/T 25070—2019)^[3]正式实施,标志着网络安全等级保护制度正式进入 2.0(以下简称等保 2.0)时代。等保 2.0 中,保护对象由信息系统扩展至基础信息网络、工业控制系统、云计算平台、物联网、移动互联网、其他网络、大数据等多个系统平台;安全体系由被动防御为主转变为事前预防、事中响应、事后审计的动态保障体系。

现阶段的等级保护测评工作主要由测评人员使用调查表、技术检测、访谈等多种方式得到数据,再根据相应的等保要求评判符合程度。由于信息系统的多样性及其安全测评的复杂性,评判过程中不可避免地存在主观因素,容易影响其评判的真实结论。因此研究者通常结合信息安全风险评估方法中的定量分析方法,建立恰当的信息安全等级保护测评量化模型,以在测评过程中获得明确的、可信的判定结果。目前常见的测评模型有:基于云模型的测评模型^[4-5],即基于已有经验构建各测评模块的三个隶属云并得到信息系统结论的判定矩阵,再利用正向云发生器得到信息系统测评后最终的安全结论的判定云图,根据判定区间得到测评系统的安全结论;基于模糊层次分析法的信息安全测评^[6]和主机安全测评模型^[7],即根据各安全评估控制点对应的综合评价向量得到某安全评估要素的模糊评价矩阵和综合评价向量,将综合评价向量与评价等级向量相乘得到最终评估结果;基于 PCA-BP 的信息安全风险评估模型^[8],即首先对原始的风险性评估数据进行主成分分析以简化神经网络结构,选取其中的一部分作为训练数据对 BP 神经网络进行训练,将其余的数据作为测试数据输入 BP 神经网络中,将测试值与真实值作对比分析以验证训练模型的有效性;基于模糊集的高校网络安全等级保护测评模型^[9],即将模糊综合评判决策方法应用于信息系统的测评符合性程度判定,并给出实际的测评结果;融合网络安全等级保护系列标准的测评量化方法^[10],即在国家等级保护的多个标准基础上,基于均衡性准则、整体性准则和有效性准则,给出相应的等级测评结论量化判定方法。上述方法均能实现对测评结果的量化表示,然而仍然存在诸如评判区间过大、未对测评指标进行权重考量等问题。决策树(Decision Tree)算法^[11]是一种基于属性分类的机器学习算法,常应用于分类^[12]和回归^[13]问题,故本文基于等保 2.0 体系下的安全要求构建了对应的数据集,采用 CART(Classification and Regression Tree)算法生成用于等级保护结论判定的回归树。此方法可用于在测评工作中辅助测评人员,给出对测评对象的客观、定量的评价结果。

1 等保 2.0 下的安全措施控制体系

在等保 2.0 体系中,《GB/T 28448—2019 信息安全技术网络安全等级保护测评要求》明确了各级别等级保护对象的测评方法和评分方式。对某个对象而言,其等级测评由单项测评和整体测评两部分组成。其中:单项测评为针对每一个具体安全要求项的测评,整体测评是在单项测评的基础上,对等级保护对象整体安全保护能力的判断。等保 2.0 的测评要求包括安全测评通用要求和安全测评扩展要求,其下分别设置安全物理环境、安全通信网络、安全区域边界等安全类,每个安全类由若干个安全控制点组成。以某第三级(S3A3G3)等级保护对象通用要求为例,其测评指标如表 1 所示。

表 1 等保 2.0 下的某第三级等级保护对象安全通用要求指标

技术/管理	安全类	安全控制点数量	测评项数
安全通用要求			
安全技术要求	安全物理环境	10	22
	安全通信网络	3	8
	安全区域边界	6	20
	安全计算环境	11	34
	安全管理中心	4	12
安全管理要求	安全管理制度	4	7
	安全管理机构	5	14
	安全管理人员	4	12
	安全建设管理	10	34
	安全运维管理	14	48
安全通用要求指标数量统计			211

设某安全控制点下有 n 个测评项,由多个对象在各测评项的符合情况(符合、部分符合、不符合、不适用)得到对应的测评项算术平均分 $\bar{x}_1 \sim \bar{x}_n$,测评项权重依次为 $w_1 \sim w_n$,则该安全控制点的得分由式(1)得出:

$$s = \frac{\sum_{k=1}^n \bar{x}_k \times w_k}{\sum_{k=1}^n w_k} \times 10 \quad (1)$$

当得到所有安全控制点的得分后,即可根据综合得分计算式给出等级测评结论,实现对等级保护对象的整体测评:

$$S = 100 - \frac{\sum_{j=1}^q \frac{\sum_{k=1}^{p(j)} (\sum_{i=1}^{m(k)} w'' + \sum_{i=1}^{m(k)} w' \times 0.5)}{\sum_{k=1}^{p(j)} \sum_{i=1}^{m(k)} w_k}}{q} \times 100 \quad (2)$$

式中: q 为被测对象涉及的安全类, $p(j)$ 为某安全类对应的总测评项数, $m(k)$ 为测评项 k 对应的测评对象数, w'' 为判定结果为“不符合”的测评项权重, w' 为判定结果为“部分符合”的测评项权重。将综合得分划分为 $[90, 100]$ 、 $[80, 90)$ 、 $[70, 80)$ 、 $[0, 70)$ 四个区间, 依次对应等级测评结论的“优”“良”“中”“差”。

2 基于安全控制点得分的回归树构建

2.1 构建训练数据集

由第 1 节可知, 等级保护对象的测评结论受多个安全控制点得分的影响, 设作为评估案例样本的应用系统个数为 m , 每个样本的安全控制点个数为 n , 则可构建如式(3)所示的原始数据矩阵 X 。

$$X = (x_{ij})_{m \times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (3)$$

式中: x_{ij} 为第 i 个应用系统的第 j 个安全控制点得分。由于等保 2.0 体系下的安全控制点数量众多, 为在保留样本集主要信息的同时消除各特征间的相关性, 简化生成的 CART 回归树规模, 本文采取文献[6-7]中的方式, 使用主成分分析法 (Principal Components Analysis, PCA) 对原始数据矩阵进行降维, 具体步骤如下:

(1) 原始数据矩阵标准化。对 X 中符合情况为“不适用”(N/A)的安全控制点, 其得分以该列中其余安全控制点得分的均值进行填充。后进行去中心化操作, 将各安全控制点得分 x 减去其所在列的均值 $\bar{x}_j$ 后再除以该列得分的标准差 σ_j , 以便寻找原始数据矩阵中数据分布的最大方差方向。设对 X 进行标准化操作后的矩阵为 X' :

$$X' = (x'_{ij})_{m \times n}, x'_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \quad (4)$$

(2) 求矩阵 X' 的协方差矩阵 C :

$$C = \frac{1}{m} X'^T X' \quad (5)$$

(3) 计算 C 的特征值和对应的特征向量, 将特征值从大到小进行排序, 记作 $\lambda_1, \lambda_2, \cdots, \lambda_m$ 。

(4) 根据主成分的累计方差贡献率确定主成分个数

k , 将对应的特征向量排列成矩阵 $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_k)$ 。

(5) 计算主成分 $R = X' \alpha$, 作为 CART 回归树的新评估案例样本并输入。

2.2 构建 CART 回归树

设经 2.1 节中步骤后, 得到的经降维的特征空间为 R :

$$R = (r_{ij})_{m \times k} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ r_{21} & r_{22} & \cdots & r_{2k} \\ \vdots & \vdots & & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mk} \end{bmatrix} \quad (6)$$

对应的各应用系统综合得分为 $Y = \{y_1, y_2, \cdots, y_m\}$, 则基于该特征空间构建其对应的 CART 回归树, 具体步骤如下:

(1) 遍历 R 中的各特征 (即每一列), 将当前特征 j 上的对应值 r 依次作为切分点 s , 即可将原始特征空间划分为两个子空间 R_1, R_2 :

$$R_1(j, s) = \{r | r^{(j)} \leq s\} \quad (7)$$

$$R_2(j, s) = \{r | r^{(j)} > s\} \quad (8)$$

(2) 计算 R_1 和 R_2 相应的输出值 (即各应用系统综合得分) 的均值:

$$\hat{e}_m = \frac{1}{N_{m r_i \in R_m(j, s)}} y_i \quad r \in R_m, m = 1, 2 \quad (9)$$

(3) 通过步骤(1)中的遍历寻找最优切分特征 j' 和最优切分点 s' , 使 R_1 和 R_2 输出值的均方误差之和最小, 即:

$$\min_{j', s'} \left[\min_{c_1} \sum_{r_i \in R_1(j', s')} (y_i - c_1)^2 + \min_{c_2} \sum_{r_i \in R_2(j', s')} (y_i - c_2)^2 \right] \quad (10)$$

(4) 将由 j' 和 s' 对 R 划分得到的 R_1, R_2 作为根节点 R 的左右子树。继续在左右子树重复步骤(1) - 步骤(4), 递归生成回归树。

为避免形成的回归树对训练数据集的过拟合, 增强其对验证数据集的泛化能力, 使用预剪枝 (Pre-Prune) 算法^[14-15]减少形成的叶节点数目。本例中使用均方误差的下降程度作为是否进行剪枝的依据, 假设对某节点进行划分得到的左右子树的均方误差之和与未划分前均方误差的差值小于某阈值, 则不对该节点进行划分, 此时该节点即为回归树的某一叶节点, 其值即为其中记录的综合得分的平均值。最终生成的 CART 回归树可表示为:

$$f(x) = \sum_{m=1}^M \hat{e}_m I \quad x \in R_m \quad (11)$$

(5) 经以上步骤(1) - 步骤(4)完成回归树的构建后, 即可将待测系统的各安全控制点得分代入, 求得

与之匹配的叶节点值,即为该待测系统的等级保护结论的预测值。

3 实验与结果分析

3.1 实验拓扑

为验证本文模型的有效性,选取某两所高校(以下称高校 A、高校 B)的校园网及相关业务系统作为实验对象。该两所高校校园网的拓扑结构如图 1、图 2 所示,可见高校 A 校园网受等级保护的区域包括出口区、校园网核心区、核心监控与优化区、专网服务区、数据中心区;高校 B 校园网受等级保护的区域包括出口区、办公区域、宿舍区域、数据中心区、安全运维区。

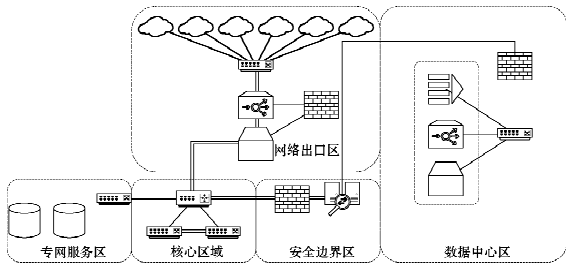


图 1 高校 A 校园网络实验拓扑

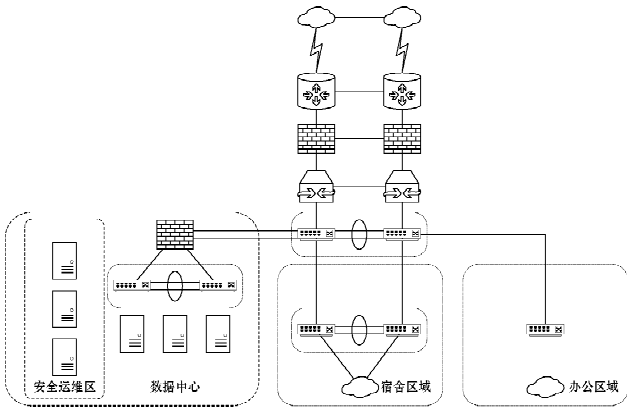


图 2 高校 B 校园网络实验拓扑

3.2 构建训练数据集

选取部署于高校 A 数据中心区的 20 个业务系统、高校 B 数据中心区的 8 个业务系统分别作为测试集,利用漏洞扫描工具、渗透测试工具等多种测试工具对各系统进行已有安全控制措施和主要安全问题的汇总分析,并由式(1)得出各系统的安全控制点(记作 $A_1 - A_{71}$ 、 $B_1 - B_{81}$)得分列表,再由式(2)计算出综合得分,如表 2、表 3 所示。

表 2 高校 A 业务系统的安全控制点得分及综合得分(部分)

系统编号	A_1	A_2	A_3	A_4	A_5	...	A_{66}	A_{67}	A_{68}	A_{69}	A_{70}	A_{71}	综合得分
1	10.0	10.0	—	10.0	10.0	...	—	3.15	10.0	10.0	6.25	—	70.23
2	10.0	10.0	—	10.0	10.0	...	—	10.0	10.0	10.0	10.0	—	93.74
3	10.0	10.0	—	10.0	10.0	...	10.0	6.30	10.0	7.3	5.0	10.0	70.07
4	10.0	10.0	—	10.0	10.0	...	0.0	0.0	10.0	10.0	8.75	10.0	75.38
5	10.0	10.0	—	10.0	10.0	...	—	10.0	10.0	10.0	10.0	—	93.87
6	10.0	10.0	7.92	10.0	10.0	...	0.0	10.0	10.0	10.0	7.5	—	88.33
7	10.0	10.0	—	10.0	10.0	...	10.0	—	10.0	10.0	5.0	5.0	81.87
8	10.0	10.0	—	10.0	10.0	...	5.0	10.0	10.0	10.0	7.5	10.0	80.89
9	10.0	10.0	—	10.0	10.0	...	0.0	10.0	10.0	10.0	10.0	10.0	81.45
10	10.0	10.0	10.0	10.0	10.0	...	10.0	10.0	10.0	10.0	0.0	10.0	87.74
11	10.0	10.0	—	10.0	10.0	...	10.0	10.0	10.0	10.0	5.0	10.0	80.97
12	10.0	10.0	—	10.0	10.0	...	10.0	—	5.83	10.0	5.0	5.0	81.26
13	10.0	10.0	—	10.0	10.0	...	10.0	0.0	10.0	6.3	10.0	—	76.37
14	10.0	10.0	—	10.0	10.0	...	0.0	10.0	10.0	10.0	10.0	—	81.89
15	10.0	10.0	—	10.0	10.0	...	10.0	10.0	10.0	10.0	5.0	10.0	80.73
16	10.0	10.0	—	10.0	10.0	...	10.0	—	10.0	10.0	5.0	5.0	82.37
17	10.0	10.0	—	10.0	10.0	...	10.0	0.0	10.0	6.3	10.0	—	74.86
18	10.0	10.0	—	10.0	10.0	...	0.0	10.0	10.0	10.0	10.0	—	80.81
19	10.0	10.0	—	10.0	10.0	...	0.0	10.0	10.0	10.0	10.0	—	73.65
20	10.0	10.0	—	10.0	10.0	...	10.0	10.0	10.0	10.0	5.0	10.0	80.91

表 3 高校 B 业务系统的安全控制点得分及综合得分(部分)

系统 编号	B_1	B_2	B_3	B_4	B_5	...	B_{76}	B_{77}	B_{78}	B_{79}	B_{80}	B_{81}	综合 得分
1	10.0	10.0	10.0	10.0	10.0	...	—	10	—	—	10.0	—	82.36
2	10.0	10.0	10.0	10.0	10.0	...	10	—	10	—	10.0	—	82.43
3	10.0	10.0	10.0	10.0	10.0	...	10	—	10	—	10.0	—	83.91
4	10.0	10.0	10.0	10.0	10.0	...	—	10	—	—	10.0	—	83.54
5	10.0	10.0	10.0	10.0	10.0	...	10	—	10	—	10.0	10.0	83.52
6	10.0	10.0	10.0	10.0	10.0	...	10	—	10	—	10.0	—	82.53
7	10.0	10.0	10.0	10.0	10.0	...	10	—	10	—	10.0	—	82.47
8	10.0	10.0	10.0	10.0	10.0	...	10	—	10	—	10.0	10.0	81.57

经标准化和 PCA 降维后,得到用于生成 CART 回归树的评估案例样本。对于高校 A 和高校 B 业务系统的安全控制点得分,本文分别取

方差贡献率最大的前 10 个主成分(记作 $C_1 - C_{10}$)和前 8 个主成分(记作 $D_1 - D_8$),如表 4、表 5 所示。

表 4 经标准化和 PCA 降维后得到的评估案例样本(高校 A)

系统 编号	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	综合 得分
1	4.95	14.09	3.64	-6.31	-6.49	-0.47	-3.32	-4.59	-2.77	1.86	70.23
2	-13.10	-3.25	-3.72	5.33	-5.67	3.32	-1.52	0.74	1.50	-0.64	93.74
3	6.60	4.51	6.61	-4.13	0.21	-3.65	-1.88	-1.62	11.72	8.02	70.07
4	4.13	1.00	2.97	-10.00	6.55	13.46	-6.40	-0.80	4.51	-3.19	75.38
5	-13.10	-2.97	-3.66	4.85	-5.64	3.54	-1.65	0.67	1.34	-0.43	93.87
6	-7.77	-4.71	-6.36	-7.57	6.35	3.39	11.18	-0.54	-2.46	2.10	88.33
7	15.94	-8.81	-11.80	1.17	-3.08	-0.23	-1.15	0.94	-1.63	-0.08	81.87
8	2.06	-1.92	-4.13	-3.69	-6.63	-10.90	7.05	-7.28	3.21	-3.69	80.89
9	-5.93	-8.30	5.09	-2.03	8.35	-6.38	-6.97	-3.46	-7.04	-2.37	81.45
10	2.02	-3.04	10.04	8.28	-4.04	2.12	4.02	7.19	-4.26	7.20	87.74
11	5.79	-10.90	12.54	3.22	1.32	0.09	1.79	-0.66	-0.32	-2.41	80.97
12	16.84	-8.33	-12.40	2.11	-1.23	0.00	-1.88	1.13	0.03	-0.49	81.26
13	1.54	17.16	-1.31	11.33	8.78	0.23	2.23	1.91	1.93	-2.62	76.37
14	-12.70	-4.30	-1.63	-2.63	1.86	-4.47	-3.57	1.94	3.93	0.54	81.89
15	5.85	-10.40	13.05	2.56	1.90	-0.47	0.77	-0.66	-0.06	-1.07	80.73
16	16.43	-6.82	-12.40	0.04	0.13	0.89	-3.55	1.85	-1.03	1.44	82.37
17	5.54	20.10	-3.66	9.26	8.20	-2.77	2.26	0.20	2.05	-1.61	74.86
18	-9.20	6.34	-4.27	4.62	3.05	-1.71	-6.96	-6.70	-6.68	5.19	80.81
19	-13.60	-3.33	-1.52	-6.35	1.91	-6.24	-3.32	7.61	2.63	-0.60	73.65
20	4.95	-10.80	13.05	3.38	1.23	-0.49	1.60	-0.87	0.29	-1.94	80.91

表 5 经标准化和 PCA 降维后得到的评估案例样本(高校 B)

系统 编号	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	综合 得分
1	1.874 96	-0.562 18	-0.077 85	-0.381 76	-0.103 58	-0.081 95	-0.129 56	0.047 55	82.36
2	-0.584 44	-0.005 02	0.121 63	0.580 00	-0.051 58	-0.000 37	-0.077 54	0.019 69	82.43
3	1.832 94	1.378 74	-0.001 44	0.628 55	0.176 09	-0.016 02	-0.002 12	-0.026 54	83.91
4	-0.879 44	0.899 22	-0.917 05	-0.105 58	0.034 65	0.179 62	0.053 28	0.061 14	83.54
5	-0.583 41	-0.758 82	0.079 11	0.283 15	-0.108 49	-0.126 42	0.065 13	0.025 28	83.52
6	-0.671 17	-0.664 84	0.404 83	-0.211 35	0.696 27	0.025 95	-0.020 81	-0.005 53	82.53
7	-0.586 16	-0.753 58	0.081 17	0.288 24	-0.102 85	-0.130 20	0.085 49	-0.002 93	82.47
8	-0.555 79	-0.777 08	0.112 71	0.123 02	-0.254 55	0.251 03	-0.076 93	-0.045 23	81.57

3.3 构建 CART 回归树

以上述两个样本作为输入,取 0.5 作为均方误差下降程度的阈值,分别构建基于业务系统安全控制点得分的 CART 回归树模型。以对高校 A 评估案例样本的第一次划分为例,经遍历后可得最优切分特征为 C_2 ,最优切分点为 -0.46 ,将样本划分为左右子树,如图 3 所示。

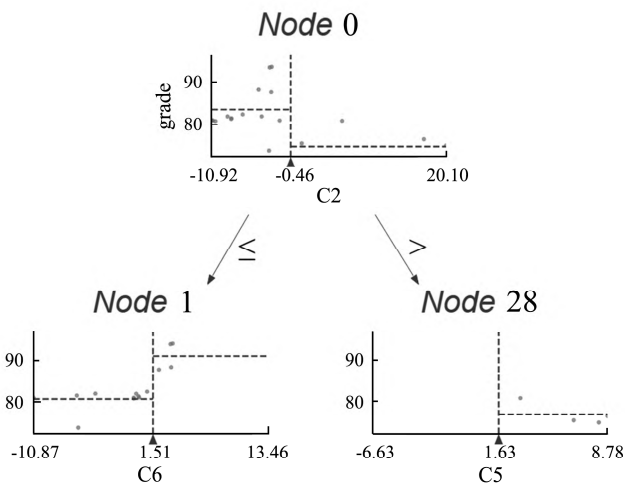


图 3 对高校 A 评估案例样本进行第一次划分的示意图

3.4 实验结果与分析

选择部署于高校 A 校园网的另外 5 个业务系统和

部署于高校 B 校园网的另外 3 个业务系统,经漏洞分析、渗透测试后得到对应的安全控制点得分,基于得分构建测试集(以下称测试集 A、测试集 B),对训练的决策树模型进行评估。本文使用平均绝对百分误差(Mean Absolute Percentage Error, MAPE)、均方误差(Mean Square Error, MSE)、决定系数(R^2)三个指标评价所构建模型的拟合程度,计算式如下:

$$E_{\text{MAP}} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \tag{12}$$

$$E_{\text{MS}} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \tag{13}$$

$$R^2 = 1 - \frac{S_{\text{SE}}}{S_{\text{ST}}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{14}$$

式中: n 为测试集的样本个数, y_i 为第 i 个样本综合得分的真实值, $\hat{y}_i$ 为第 i 个样本综合得分的预测值, $\bar{y}$ 为 n 个样本综合得分真实值的均值, S_{SE} 为误差平方和, S_{ST} 为总平方和。对测试集 A、测试集 B 的评估结果如表 6、表 7 所示。

表 6 对测试集 A 的评估结果

系统 编号	y_i	$\hat{y}_i$	$(\hat{y}_i - y_i)^2$	$(y_i - \bar{y})^2$	$\left \frac{\hat{y}_i - y_i}{y_i} \right $	MAPE	MSE	R^2
1	88.50	88.035	0.216 225	810.000 0	0.005 254 2	0.029 360 9	7.441 51	0.919 4
2	71.13	75.120	15.920 100	70.056 9	0.056 094 5			
3	72.02	70.150	3.496 900	55.950 4	0.025 965 0			
4	72.18	76.370	17.556 100	53.582 4	0.058 049 3			
5	93.67	93.805	0.018 225	200.788 9	0.001 441 2			

表 7 对测试集 B 的评估结果

系统 编号	y_i	$\hat{y}_i$	$(\hat{y}_i - y_i)^2$	$(y_i - \bar{y})^2$	$\left \frac{\hat{y}_i - y_i}{y_i} \right $	MAPE	MSE	R ²
1	83.46	83.91	0.202 5	0.132 0	0.005 4	0.002 366 7	0.074 033 3	0.657 9
2	82.44	82.44	0.000 0	0.431 2	0.000 0			
3	83.39	83.53	0.019 6	0.086 0	0.001 7			

由表 6、表 7 可知,对测试集 A、测试集 B 测试结果的平均绝对百分误差 MAPE 分别为 0.029 和 0.002 366 7,接近于 0,说明通过该模型得到的综合得分预测值与真实值误差较小;决定系数 R² 分别为 0.9 和 0.657 9,反映了综合得分的 90% 和 65% 的变化可由本模型中存在的自变量(主成分 C₁ - C₁₀、D₁ - D₈)解释。综上所述,本文模型具备较好的对等级保护安全结论的预测能力。

在相同的实验拓扑环境下,使用文献[8]提出的基于 PCA-BP 的信息安全风险评估模型分别对两组训练样本数据进行训练,并通过训练好的 BP 神经网络对测试集进行综合得分的评估。将 PCA-BP 算法的评估结果与本文算法的评估结果进行对比,如图 4、图 5 所示(创建的 BP 神经网络输入层分别包含 10 个节点和 8 个节点,隐藏层包含 10 个节点,输出层包含 1 个节点,采用带动量的批处理梯度下降法作为神经网络的训练方法,动量因子为 0.8,学习率为 0.01,最大训练次数为 20 000 次)。

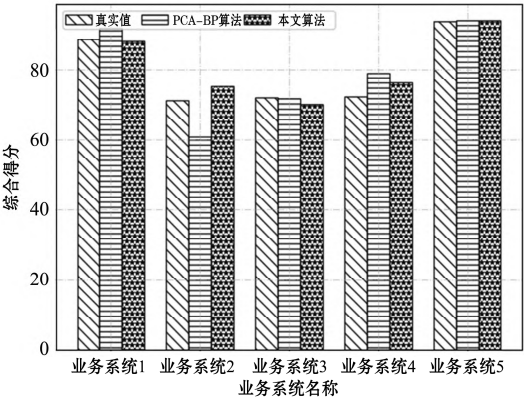


图 4 本文算法与 PCA-BP 算法评估结果的对比(测试集 A)

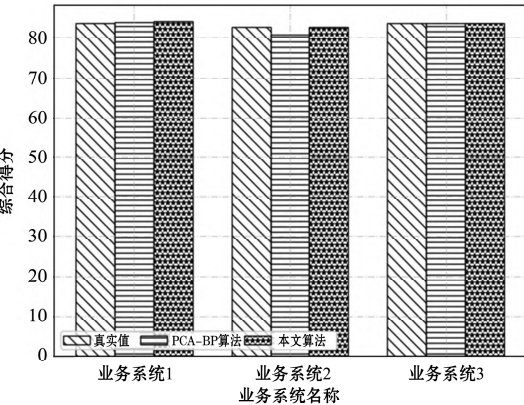


图 5 本文算法与 PCA-BP 算法评估结果的对比(测试集 B)

其中,每组条形图代表测试集中的 1 个业务系统的得分情况,从左至右依次为该业务系统的真实测评得分、PCA-BP 算法的评估结果和本文算法的评估结果。相应的 MAPE、MSE、R² 如表 8、表 9 所示。

表 8 本文算法与 PCA-BP 的比较(测试集 A)

算法 名称	MAPE	MSE	R ²	运行 时间/ms
本文算法	0.029 360 9	7.441 51	0.919 4	268 016
PCA-BP	0.055 970 4	32.003 56	0.653 2	724 645

表 9 本文算法与 PCA-BP 的比较(测试集 B)

算法 名称	MAPE	MSE	R ²	运行 时间/ms
本文算法	0.002 366 7	0.074 033 3	0.657 9	72 004
PCA-BP	0.008 337 6	16.205 6100	-0.346 3	189 011

由表 8、表 9 可知,使用本文算法进行综合得分预测的结果的 MAPE 相较于 PCA-BP 分别下降了 3% 和 0.6% 左右,同时由于本文算法省略了传统 BP 神经网络的构建、训练等一系列繁琐的步骤,算法的运行时间相较于 PCA-BP 有较为明显的提升。

4 结 语

根据网络安全等级保护 2.0 体系的基本要求,构建基于 CART 回归树的安全等级保护结论判定模型。本文模型充分考虑各安全控制点对测评结论的影响程度,并通过降维和剪枝等方式控制回归树的规模,保证预测过程的运行效率。实例表明,本文模型可有效地实现对信息系统测评安全结论得分的预测,其结果与实际测评中经专家审核测评结果得到的得分可以较好地拟合,具有一定的应用前景和研究价值。

参 考 文 献

[1] 马力,祝国邦,陆磊.《网络安全等级保护基本要求》(GB/T 22239—2019)标准解读[J]. 信息安全, 2019(2):77-84.

[2] 陈广勇,祝国邦,范春玲.《信息安全技术 网络安全等级保护测评要求》(GB/T 28448—2019)标准解读[J]. 信

- 息网络安全,2019(7):1-7.
- [3] 国家市场监督管理总局,中国国家标准化管理委员会. 信息安全技术网络安全等级保护安全设计技术要求:GB/T 25070-2019[S]. 北京:中国标准出版社,2019.
- [4] 张鹏,谢晓尧. 基于云模型的信息系统测评安全结论判定[J]. 武汉大学学报(理学版),2014,60(5):429-433.
- [5] 朱丹,谢晓尧,徐洋,等. 基于云模型与贝叶斯反馈的网络安全等级评估方法[J]. 山东大学学报(理学版),2018,53(1):53-62.
- [6] 牛晓博,方群,朱飞. 基于模糊层次分析法的信息安全测评[J]. 软件,2021,42(9):154-157.
- [7] 齐富民,谢晓尧,徐洋. 基于模糊理论的 Web 安全评估模型[J]. 计算机应用研究,2014,31(6):1883-1888.
- [8] 任远芳,刘志杰,景凤宣,等. 基于 PCA-BP 的信息安全风险评估模型[J]. 计算机仿真,2014,31(6):212-216,281.
- [9] 林志兴,张武威,余建,等. 基于模糊集的高校网络安全等级保护体系研究[J]. 计算机应用与软件,2021,38(4):305-310.
- [10] 邹鑫灏,李新建,潘伟,等. 融合网络安全等级保护系列标准的测评量化方法探讨[J]. 信息安全,2021(S1):190-194.
- [11] Sathyadevan S, Nair R. Comparative analysis of decision tree algorithms: ID3, C4.5 and random forest [C]//Computational Intelligence in Data Mining,2015:549-562.
- [12] 李康乐,任志磊,周志德,等. 基于决策树算法的 API 误用检测[J]. 计算机科学,2022,49(11):30-38.
- [13] 王辉,张文杰,刘杰,等. 基于分类回归决策树算法的航班延误预测模型[J]. 中国民航大学学报,2022,40(3):35-40.
- [14] Esposito F, Malerba D, Semeraro G, et al. A comparative analysis of methods for pruning decision trees [J]. IEEE Computer Society,1997,19(5):476-491.
- [15] Rastogi R, Shim K. PUBLIC: A decision tree classifier that integrates building and pruning[J]. Data Mining and Knowledge Discovery,2000,4(4):315-344.
- a revolutionary and flawed grand experiment in blockchain: The DAO attack[J]. Journal of Cases on Information Technology,2019,21(1):19-32.
- [8] 钱鹏,刘振广,何钦铭,等. 智能合约安全漏洞检测技术研究综述[J]. 软件学报,2022,33(8):3059-3085.
- [9] Lai E, Luo W. Static analysis of integer overflow of smart contracts in ethereum[C]//4th International Conference on Cryptography, Security and Privacy,2020:1-8.
- [10] 董伟良,刘哲,刘逵,等. 智能合约漏洞检测技术综述[J]. 软件学报,2024,35(1):38-62.
- [11] Garfatta I, Klai K, Gaaloul W, et al. A survey on formal verification for solidity smart contracts [C]//Proceedings of the 2021 Australasian Computer Science Week Multiconference,2021:1-10.
- [12] Nam W, Kil H. Formal verification of blockchain smart contracts via atl model checking[J]. IEEE access, 2022, 10: 8151-8162.
- [13] Jensen K. Coloured petri nets[M]//Petri nets: central models and their properties. Springer, 1987: 248-299.
- [14] 徐倩,林俊亭. RSSP-II 协议消息鉴定层的形式化描述及验证[J]. 计算机应用与软件,2023,40(6):77-82,180.
- [15] Bhargavan K, Delignat-Lavaud A, Fournet C, et al. Formal verification of smart contracts: Short paper [C]//ACM Workshop on Programming Languages and Analysis for Security,2016:91-96.
- [16] Amani S, BéGEL M, Bortin M, et al. Towards verifying ethereum smart contract bytecode in Isabelle/HOL [C]//7th ACM SIGPLAN International Conference on Certified Programs and Proofs,2018:66-77.
- [17] Hildenbrandt E, Saxena M, Rodrigues N, et al. KEVM: A complete formal semantics of the ethereum virtual machine [C]//31st Computer Security Foundations Symposium, 2018:204-217.
- [18] Abdellatif T, Brousmiche K L. Formal verification of smart contracts based on users and blockchain behaviors models [C]//9th IFIP International Conference on New Technologies, Mobility and Security,2018:1-5.
- [19] 赵颖琪,朱雪阳,李广元,等. 带时间约束的智能合约验证[J]. 应用科学学报,2021,39(1):1-16.
- [20] Petri C A. Kommunikation mit automaten[D]. Bonn: University of Bonn,1962.
- [21] Liu Z T, Liu J. Formal verification of blockchain smart contract based on colored petri net models [C]//43rd Annual Computer Software and Applications Conference, 2019, 2: 555-560.
- [22] 董春燕,谭良. 基于 CPN 模型 Auction 智能合约的形式化验证[J]. 小型微型计算机系统,2020,41(11):2292-2297.
- [23] 郑红,刘泽润,黄建华,等. 基于 CPN 的智能合约交易顺序依赖漏洞的验证[J]. 系统仿真学报,2022,34(7):10.

(上接第 345 页)

- [4] Szabo N. Smart contracts: Building blocks for digital markets [J]. EXTROPY: The Journal of Transhumanist Thought, 1996,18(2):28.
- [5] 李继海,邵祥东,刘佳宁. 基于区块链技术的福利乐透事业优化路径探究[J]. 社会福利(理论版),2021(5):33-38.
- [6] Fire Lotto. Fire lotto: The first truly transparent decentralized blockchain lottery [EB/OL]. [2023-10-12]. <https://www.prnewswire.com/news-releases/fire-lotto-the-first-truly-transparent-decentralized-blockchain-lottery-300600719.html>.
- [7] Mehar M I, Shier C L, Giambattista A, et al. Understanding