

DOI:10.3969/j.issn.1003-5060.2024.02.003

核泄漏事故风险评估中的概率分析及预测

何博文¹, 关群²

(1. 范德堡大学 土木与环境工程系, 田纳西 纳什维尔 37212; 2. 合肥工业大学 土木与水利工程学院, 安徽 合肥 230009)

摘要: 文章利用逻辑回归模型(logistic regression model, LRM)、线性判别模型(linear discriminant model, LDM)和支持向量机(support vector machine, SVM)3种统计模型,从核反应堆的内部和外部因素2个方面评估其在核泄漏事故中所体现的相关安全性能。针对每种模型,利用数理统计理论探究核反应堆相关影响因素与其发生核泄漏事故的概率。研究发现核反应堆外部因素有主导内部因素的趋势并在整个核泄漏事故风险中占有举足轻重的地位。文章提供的模型分析与预测结果可为核反应堆工程师及其相关决策者在核反应堆的选址、设计及建设运营等方面提供参考。

关键词: 核泄漏; 风险评估; 概率分析; 逻辑回归模型(LRM); 线性判别模型(LDM); 支持向量机(SVM)

中图分类号: X946 **文献标志码:** A **文章编号:** 1003-5060(2024)02-0161-08

Probability analysis and prediction associated with nuclear meltdown accident risk evaluation

HE Bowen¹, GUAN Qun²

(1. Department of Civil and Environmental Engineering, Vanderbilt University, Nashville 37212, USA; 2. School of Civil and Hydraulic Engineering, Hefei University of Technology, Hefei 230009, China)

Abstract: This paper uses logistic regression model(LRM), linear discriminant model(LDM) and support vector machine(SVM) to evaluate the safety performance of nuclear reactors in nuclear meltdown accidents from both internal and external factors. For each model, the relevant influencing factors of nuclear reactor and the probability of nuclear meltdown accident are investigated by mathematical statistics theory. It is found that the external factors tend to dominate the internal factors and play an important role in the nuclear meltdown accident risk. The model analysis and prediction results in this paper can provide reference for nuclear reactor engineers and related decision makers in the site selection, design, construction and operation of nuclear reactors.

Key words: nuclear meltdown; risk evaluation; probability analysis; logistic regression model(LRM); linear discriminant model(LDM); support vector machine(SVM)

核泄漏是一种可对人类以及周边环境造成巨大破坏的工程事故,通常可分为核芯完全熔融和核芯部分熔融^[1],具体表现为反应堆内部由于温度超过设定温度所导致的核芯和燃料棒的融化及放射性物质的泄漏^[2]。历史上曾出现过多次举世闻名的重大核泄漏灾害事故,例如,2011年的日

本福岛事件,1986年的切尔诺贝利事件及1977年的美国三里岛事件^[3]。因此,核泄漏被认为是当代最危险的工业事故之一,其重要性值得业内工程师以及科学家投入大量精力来研究。

本文基于统计分析的角度,探究核泄漏事故的发生与反应堆自身的设计因素和外部影响因素

收稿日期:2020-08-11;修回日期:2020-08-25

基金项目:国家自然科学基金资助项目(51878234)

作者简介:何博文(1994—),男,安徽合肥人,范德堡大学博士生;

关群(1962—),女,江苏扬州人,博士,合肥工业大学教授,硕士生导师,通信作者,E-mail:grz@hfut.edu.cn.

之间存在的联系,为核反应堆工程师在未来反应堆的选址、设计及相关建设运营方面提供参考。

1 统计方法

1.1 样本选取

本文重点选取在全球范围内具有核泄漏事故记录的核反应堆和从未发生过核泄漏事故的核反应堆进行研究。对于具有核泄漏记录的核反应堆,不仅采取具有全球重要影响的核泄漏记录的核反应堆,同时也采取一些具有不著名核泄漏历史记录核反应堆来组成研究样本。为了扩充样本容量,本文不仅选取大型商用核反应堆(例如切尔诺贝利核反应堆),同时也采取小型还未商用的反应堆(例如实验与科研反应堆)来组成研究样本。对于未发生过核泄漏事故的核反应堆样本,其样本个例均来自全球正在商用运行的核反应堆机组。

1.2 数据采集

本文在全球范围内选取 50 个核反应堆组成研究样本。样本反应堆的特征参数来自世界核协会与国际原子能机构^[4-5]。通过确定核反应堆的设计类型,确定其设计公司并查找该公司该产品的技术报告和设计手册的方法来最终确定该反应堆的所有相关设计参数。同时,采取大量来自研究文献的信息来评估样本反应堆是否建造在一个易发生地震与海啸的位置。例如,在样本反应堆中,有一个是由关西电力公司建造运营的三滨核反应堆。文献^[6]研究发现建造的三滨核反应堆位置下有剧烈的地壳运动痕迹,显示其位置具有很大的发生 7.4 级及以上地震的可能性,从而影响三滨核反应堆的运营。同时,由于日本是一个狭长的岛国,对于地震之后接连发生海啸有极大的可能性,而海啸对临近海岸线的建筑物包括核电站会造成巨大的破坏和影响。基于这些信息,三滨核反应堆的建造地址是一个不够理想且易于发生核泄漏事故的地理位置。同样采取大量来自研究文献的信息判断在核泄漏事故中是否存在人为操作上的失误。例如,历史上著名的三里岛核泄漏事件中,一位核电站操作人员误判了阀门的开关状态这一重大关键性的人为操作失误,直接导致了美国核工业历史上无可挽回的结果^[7]。

1.3 解释变量

本研究把可能与核泄漏事故发生相关联的因素分为内部因素和外部因素。内部因素是指由反应堆自身特性所决定的因素,例如反应堆一系列

的设计参数。外部因素是指由非反应堆自身特性所决定的因素,例如人为以及不可抗力因素。

根据人为环境因素,本文在内部因素中选取设计功率、堆芯高度、堆芯直径、燃料类型、冷却液类型和慢化剂类型 6 个变量。设计功率是衡量核反应堆输出能力的表征量;堆芯高度、堆芯直径是表征核反应堆堆芯大小的几何量;燃料类型、冷却液类型及慢化剂类型是代表核反应堆类型的表征量。同时,在这些内部因素中,设计功率、堆芯高度和堆芯直径是数值变量;燃料类型、冷却液类型及慢化剂类型是类别变量。本文使用虚变量表示不同的燃料、冷却液及慢化剂类型。

对于外部因素,选取建造地址和人为操作因素作为模型独立自变量。建造地址是反映样本反应堆的安全运行是否容易遭受由地理位置而带来的不可抗力因素,诸如地震或海啸的影响的表征量。人为因素是反映在整个核反应堆运行过程中,样本反应堆的安全运行是否容易遭受由于重大人为操作失误因素而带来的运行风险的表征量。同时,在这些外部因素中,两者均为布尔变量,即 0 代表建造地址不易受诸如地震、海啸的影响以及没有发现任何人为操作失误的历史记录;1 代表建造地址容易受诸如地震、海啸的影响及发现存在人为操作失误的历史记录。本文研究的变量类型见表 1 所列。

表 1 变量因素类型

变量因素	变量类型	变量因素	变量类型
核反应堆熔化 P_{MD}	布尔	水 L_{C1}	布尔
设计功率 P_{DC}	数值	液体钠 L_{C2}	布尔
堆芯高度 H_C/m	数值	二氧化碳 L_{C3}	布尔
堆芯直径 D_C/m	数值	慢化剂类型 M	类别
燃料类型 T_F	类别	水 M_1	布尔
二氧化铀 T_{F1}	布尔	重水 M_2	布尔
金属铀 T_{F2}	布尔	石墨 M_3	布尔
铀氢锆 T_{F3}	布尔	铍 M_4	布尔
铀混合氧化物 T_{F4}	布尔	建造地理位置 P_{LC}	布尔
冷却液类型 L_C	类别	人为因素 P_{HU}	布尔

2 模型建立

2.1 逻辑回归模型

逻辑回归模型(logistic regression model, LRM)是基于一系列独立自变量预测并分类系统应变量的统计学分析方法^[8]。LRM 不同于传统的判别式分析模型,其自变量无需服从正态分布,或要求其每个组别中都有相同的方差^[9],存在明

显的优势。此外,LRM 对其输入自变量基本不存在任何限制^[9]。在本研究中,LRM 中的独立自变量为布尔类型、类别类型、数值类型。LRM 可允许在不同复杂程度的模型中进行预测精确性的对比。最简单的模型是只用一个与应变量具有最显著数值关系的变量作为独立自变量的模型,而最复杂的模型是采取所有可能与系统应变量有关系的变量作为独立自变量的模型。本文采用具有最精确的分类预测能力,同时具有最少独立自变量的 LRM。

本研究使用可产生 2 种输出结果的二元 LRM,即输出结果可以是判定反应堆出现核泄漏事故或反应堆无核泄漏事故。采用 50 个来自不同组别的反应堆来训练回归模型,即 19 个具有核泄漏事故历史的反应堆和 31 个无核泄漏历史的反应堆。选取 19 个具有核泄漏事故历史的反应堆的原因在于该类反应堆的数量在全球范围内较少,它们的样本已经充分利用了此类样本所能提供的有限的信息。对于另一组 31 个无核泄漏历史的反应堆,首先在数据库中排除所有具有核泄漏事故历史的样本,然后采取随机选取的方式在数据库中剩余的反应堆样本里选取 31 个反应堆来组成第 2 组无核泄漏历史记录的本。

在建立 LRM 之前,本文采用卡方检验方法(chi-squared method)测试各个独立自变量与应变量的显著性水平^[10-11]。具体步骤为从建立最简单的只含有一个独立自变量的模型开始到最复杂的含有所有独立自变量的模型,利用 ANOVA chi-squared 测试方法检测每个模型之间是否产生显著性差别,判断各个独立自变量与应变量之间是否存在显著联系^[12]。独立变量及其显著性水平对应的 P 值(显著性水平概率)见表 2 所列。

从表 2 可以看出,每个独立自变量都与发生核泄漏之间存在潜在的联系。基于各个变量和 P 值,采取 98%(P 值小于 0.02)和 95%(P 值小于 0.05)2 种显著性水平来建立模型。对于 98% 显著性回归模型,只在模型中采取 P 值小于 0.02 的独立自变量作为预测变量,即堆芯高度、冷却液类型、建造地理位置以及人为因素,而放弃 P 值大于 0.02 的独立自变量。同理,对于 95% 显著性回归模型,采取所有 P 值小于 0.05 的独立自变量作为预测变量,即模型中将包括除了设计能力和慢化剂类型的所有独立自变量。

在这以上 2 种具有不同显著性水平的模型

中,样本反应堆的输出泄漏概率大于 0.5 将会被分类为发生核泄漏组,输出泄漏概率小于 0.5 将会被分类至无核泄漏发生组。使用 Nagelkerke 的 R^2 值评估模型的精确度,即通过衡量模型样本预测分类成功的概率来衡量模型评估核泄漏发生风险的能力^[13]。

表 2 独立变量与其显著性水平

变量	P 值	变量	P 值
P_{DC}	0.052	L_C	0.01
H_C	0.01	M	0.06
D_C	0.04	P_{LC}	0.016
T_F	0.023	P_{HU}	0.000 68

2.2 线性判别模型

本文采取线性判别模型(linear discriminant model, LDM)的方法来评估预测核泄漏事故发生风险,不同于 LRM, LDM 同样有其优势与劣势。在本研究中,较小的样本容量(50 个核反应堆)会导致 LRM 所产生的结果具有不稳定性^[14]。而在 LDM 中,在每组样本数据中每一个自变量服从正态分布的假设前提下,可近似地认为该模型可以很好地解决这一问题^[15],然而这也无法确定在每组样本数据中每一个自变量确实严格地服从正态分布。因此, LDM 的存在确实带来了一个明显的权衡,即模型的稳定性与不确定性之间的取舍^[16]。

对于建立的 LDM,基于表 2 采取与 LRM 相同的 2 种显著性水平来考虑。对于自由模型,采取 95% 的显著性水平,模型中包含堆芯高度、堆芯直径、燃料类别、冷却剂类别,建造地理位置及人为操作作为独立自由变量。对于保守模型,采取 98% 的显著性水平,模型中仅包含堆芯高度、冷却剂类别、建造地理位置及人为操作作为独立自由变量。在二分问题中,与 LRM 类似采取 50% 作为既定观测值来区分样本。

2.3 支持向量机

基于 LRM 和 LDM,采用支持向量机(support vector machine, SVM)的方法来学习并分类核反应堆样本。采取与前 2 个模型相同的显著性水平来建立 SVM 模型。不同的是在 SVM 中,采取径向核的方法来达到实现非线性边界决策的效果^[17]。同时采取十折交叉验证的方法,即将样本容量均等分为 2 组,即每组 25 个样本个例,分别用来训练模型和检测模型^[18]。该方法可对每种显著性水平的 SVM 选取最佳的代价函数 cost

及径向核参数 γ 。

3 结果与预测分析

3.1 逻辑回归模型

3.1.1 逻辑回归的自由模型

在自由模型中,有堆芯高度、堆芯直径、燃料类别、冷却液类别、建造地址以及人为因素这 6 个预测变量可以达到 95% 的显著性水平,因此该模型中包含这 6 个预测变量。该自由模型公式为:

$$\frac{P}{1-P} = e^x \quad (1)$$

$$x = -18.08 - 0.22H_C - 0.91D_C + 0.71T_{F1} + 9.76T_{F2} + 14.19T_{F3} + 0.001T_{F4} + 19.37L_{C1} + 15.34L_{C2} + 22.81L_{C3} - 4.25P_{LC} + 6.52P_{HU} \quad (2)$$

该模型的预测准确率为 92%,具体预测结果见表 3 所列。该模型的 R^2 值为 0.76, P 值为 $2.86E-05$,模型具体信息见表 4 所列。

表 3 LRM 中自由模型预测结果

组别	核泄漏组	无核泄漏组
核泄漏组	16	2
无核泄漏组	2	30

表 4 LRM 中自由模型计算结果

变量因素	系数	让步比	Z 值	P 值
C	-18.080	1.410E-08	-0.005 0	0.99
H_C	-0.220	0.804	-1.043 0	0.30
D_C	-0.910	0.404	-1.604 0	0.11
T_{F1}	0.710	2.040	0.340 0	0.73
T_{F2}	9.760	1.738E+04	1.650 0	0.10
T_{F3}	14.190	1.450E+06	0.005 0	0.99
T_{F4}	0.001	1.000	0.000 3	0.99
L_{C1}	19.370	2.590E+08	0.004 9	0.99
L_{C2}	15.340	4.580E+06	0.003 9	0.99
L_{C3}	22.810	8.070E+09	0.005 8	0.99
P_{LC}	-4.250	1.420E-02	-1.630 0	0.10
P_{HU}	6.520	6.770E+02	2.080 0	0.04

采用受试者工作特征曲线 (receiver operating characteristic curve, ROC) 及其线下面积 (area under the curve, AUC) 来评估本文中出现的所有判别模型的预测能力。ROC 曲线是根据一系列不同的二分类方式,以真阳性率(敏感度)为纵坐标,假阳性率(特异性)为横坐标绘制的曲线^[19]。AUC 是衡量模型优劣的一种性能指标,其越接近于 1.0,分类方法真实性越高,等于 0.5 时,则真实性最低,模型无应用价值^[20-22]。该自由模型的 ROC 曲线如

图 1 所示, AUC 值为 0.953 1。

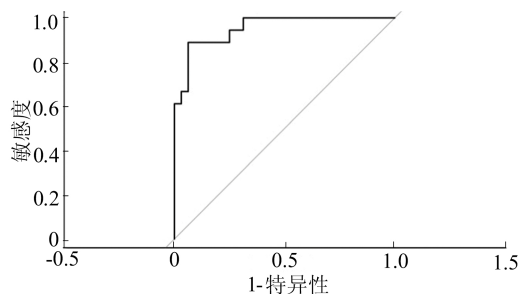


图 1 LRM 中自由模型的 ROC 曲线

3.1.2 逻辑回归的保守模型

在保守模型中,有堆芯高度、燃料类别、建造地址以及人为因素 4 个预测变量可以达到 98% 的显著性水平,因此该模型中包含这 4 个预测变量。模型公式为:

$$P/(1-P) = e^y \quad (3)$$

$$y = -19.05 - 0.39H_C + 18.69L_{C1} + 18.55L_{C2} + 21.64L_{C3} - 3.16P_{LC} + 4.79P_{HU} \quad (4)$$

该模型的预测准确率为 88%,具体预测结果见表 5 所列。该模型的 R^2 的值为 0.68, P 值为 $6.41E-06$,计算结果见表 6 所列。该保守模型的 ROC 曲线如图 2 所示, AUC 值为 0.946 2。

表 5 LRM 中保守模型预测结果

组别	核泄漏组	无核泄漏组
核泄漏组	13	1
无核泄漏组	5	31

表 6 LRM 中保守模型计算结果

变量因素	系数	让步比	Z 值	P 值
C	-19.05	5.34E-09	-0.008	0.990
H_C	-0.39	0.68	-2.166	0.030
L_{C1}	18.69	1.31E+08	0.008	0.990
L_{C2}	18.55	1.31E+08	0.008	0.990
L_{C3}	21.64	2.51E+09	0.009	0.990
P_{LC}	-3.16	4.26E-02	-1.830	0.067
P_{HU}	4.79	1.20E+02	2.605	0.009

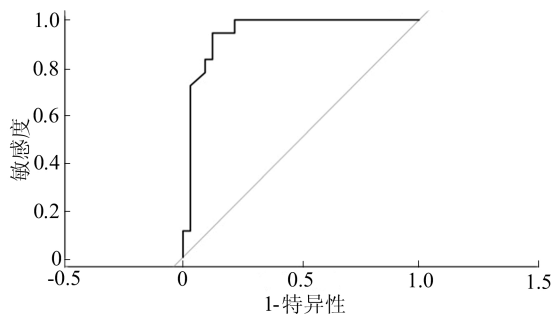


图 2 LRM 中保守模型的 ROC 曲线

3.2 线性判别模型

3.2.1 线性判别的自由模型

在线性判别的自由模型中,先验概率值 π_k 通过训练模型数据部分计算得出,核泄漏组和无核泄漏组 π_k 值依次为 0.36、0.64。在 LDM 中,假设每个预测变量都近似地服从高斯分布。高斯分布的系数平均值 μ_k 、方差 σ^2 可由样本数据来预估得到。线性判别模型中利用这些预测变量的高斯分布信息构建。在线性判别过程中主要步骤分为抽取步骤和分类步骤。

在抽取步骤中,输入的预测变量线性组合形成判别式隐藏变量。该模型中相关预测变量的系数及判别式系数分别见表 7 所列。具体的 2 组数据的判别结果如图 3 所示。从图 3 可以看出,在每组中判别式的分布差别较大,表明该模型可较好区分 2 组数据。

表 7 预测变量的 μ_k 和判别式系数

变量因素	μ_k		判别式系数
	核泄漏组	无核泄漏组	
H_C	3.17	5.62	0.13
D_C	3.23	5.24	0.16
T_{F1}	0.44	0.75	-1.14
T_{F2}	0.33	0.06	-2.02
T_{F3}	0.11	0	-1.49
T_{F4}	0.06	0.09	-0.24
L_{C1}	0.39	0.75	-2.53
L_{C2}	0.33	0.03	-2.25
L_{C3}	0.28	0.19	-4.44
P_{LC}	0.17	0.50	1.00
P_{HU}	0.67	0.19	-2.12

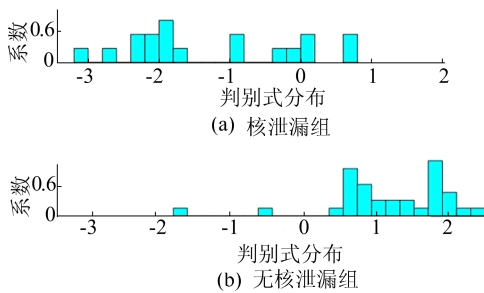


图 3 自由模型中的线性判别式

自由模型的预测准确率为 86%,具体预测结果见表 8 所列。自由模型的 ROC 曲线如图 4 所示,AUC 值为 0.960 1。

表 8 LDM 中自由模型预测结果

组别	核泄漏组	无核泄漏组
核泄漏组	13	2
无核泄漏组	5	30

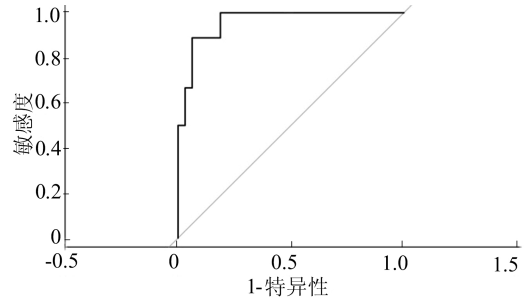


图 4 LDM 中的自由模型 ROC 曲线

3.2.2 线性判别的保守模型

保守的线性判别模型只包含堆芯高度、冷却液类别、建造地址与人为因素 4 个预测变量。该保守模型的样本先验概率值 π_k 与自由模型中的先验概率值相同,先验概率及保守模型判别式系数见表 9 所列。

表 9 预测变量的 μ_k 和判别式系数

变量因素	μ_k		判别式系数
	核泄漏组	无核泄漏组	
H_C	3.17	5.62	0.20
L_{C1}	0.39	0.75	-2.75
L_{C2}	0.33	0.031	-2.92
L_{C3}	0.28	0.19	-4.33
P_{LC}	0.17	0.5	1.00
P_{HU}	0.67	0.19	-2.25

LDM 中的保守模型判别结果如图 5 所示。保守模型的预测准确率为 88%,具体预测结果见表 10 所列。

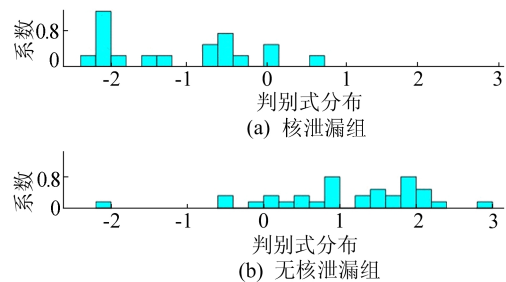


图 5 保守模型的线性判别式

表 10 LDM 中保守模型预测结果

组别	核泄漏组	无核泄漏组
核泄漏组	15	3
无核泄漏组	3	29

ROC 曲线如图 6 所示,AUC 值为 0.941 0。

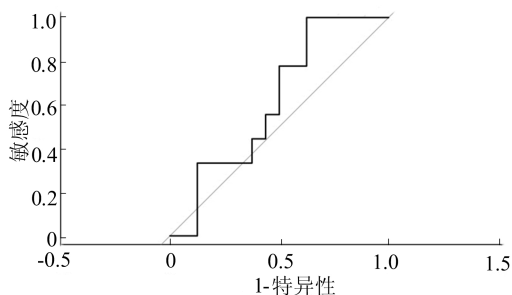


图 6 LDM 中的保守模型 ROC 曲线

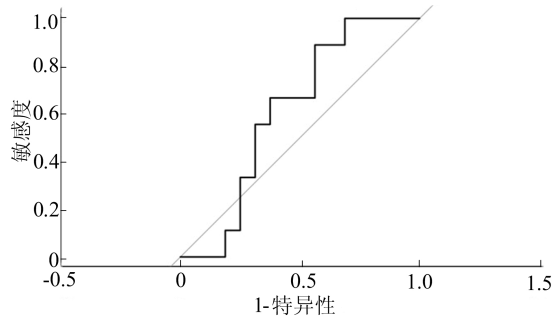


图 8 SVM 中的保守模型 ROC 曲线

3.3 SVM

3.3.1 支持向量机的自由模型

在 SVM 的自由模型中,同样包含与前面所有自由模型相同的 6 个预测变量。在 SVM 模型建立过程中,把 50 个样本核反应堆均等分为训练模型部分和检验模型部分。利用十折交叉验证的方法,SVM 自由模型最佳参数选择代价函数 cost 取 5,径向核参数 γ 取 1。SVM 中的自由模型在检验数据下的 ROC 曲线如图 7 所示,该曲线与横坐标之间的面积为 0.618 1。利用检验数据计算出模型的预测精确度为 68%。自由模型的预测结果见表 11 所列。

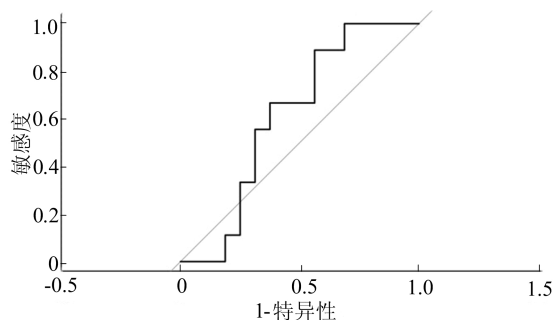


图 7 SVM 中的自由模型 ROC 曲线

表 11 SVM 的自由模型的预测结果

组别	核泄漏组	无核泄漏组
核泄漏组	5	4
无核泄漏组	4	12

3.3.2 支持向量机的保守模型

在 SVM 的保守模型中,预测变量的数量减少为 4 个。同样利用十折交叉验证的方法,选择的最佳参数代价函数 cost 取 1, γ 径向核参数取 1。SVM 中的保守模型在检验数据下的 ROC 曲线如图 8 所示,AUC 为 0.611 1。利用检验数据计算出该保守模型的预测精确度为 84%,高于自由模型。SVM 的保守模型预测结果见表 12 所列。

表 12 SVM 中保守模型预测结果

组别	核泄漏组	无核泄漏组
核泄漏组	5	4
无核泄漏组	0	16

4 讨 论

4.1 重要影响因素

通过统计模型的构建,确定影响核泄漏事故的重要因素。对比自由模型与保守模型可看出,在核泄漏事故中,堆芯高度、建造地理位置及人为因素是影响事故发生的重要因素。值得注意的是,在自由模型和保守模型中,堆芯高度的让步比均小于 1,表明此预测变量与核泄漏事故发生风险之间可能存在负相关的关系。

另一个重要的影响因素是人为因素。在所有模型中,人为因素与核泄漏事故发生风险均存在非常明显的正相关关系:在越有可能发生人为操作失误的样本中,发生核泄漏事故的风险越大。人为因素也可能是主导所有其他因素并成为影响核泄漏事故发生的唯一因素。尽管在样本核反应堆中,有一些未发生过核泄漏事故的样本也存在人为操作失误的历史记录,但这并不妨碍人为操作失误可以直接导致核泄漏事故发生这一结论。同时,基于 LRM 发现人为操作这一预测变量的让步比大于 1, P 值较小,说明在本研究中,人为因素可覆盖其他所有预测变量的趋势并有可以单独主导核泄漏事故发生的可能性。即使增大样本大小,人为因素与核泄漏事故之间联系的强度不会改变,依然可能是影响核泄漏事故发生的最重要因素。

4.2 模型不确定性分析

文中模型潜在的不确定性主要来自样本数据,样本数据的大小偏小。虽然历史上发生核泄漏事故的次数较多,但关于这些核泄漏事故的详细历史记录较少,因此极难囊括所有的核泄漏事故并把它们全部都用于此研究样本中。尤其在

LRM 中,较小的样本数量可能会在构建该模型时造成误差,因此 LRM 的精确度会随着样本数量的增加而变得更为精确。另外,在样本反应堆中,采用现阶段不同运行状态的核反应堆。这些不同运行状态的核反应堆样本会在一定程度上给模型结果带来不确定性。例如,将正在安全运行的核反应堆样本完全归为安全的无核泄漏事故发生的核反应堆样本是不严谨的。核泄漏事故是一种一次性事件,不能被看作反应堆内部的自有特性,因此不会发生核泄漏事故与目前还没有发生核泄漏事故是不能完全等价的。也就是说,对于这些目前还没有发生核泄漏事故的“干净”的核反应堆样本,其在未来依然有发生核泄漏事故的风险,尽管概率很低。从另外一个角度来诠释这一样本不确定性,发生核泄漏事故可被视作为一个核反应堆的内在“基因”,一些样本核反应堆已经展示出了此“基因”,另外一些样本核反应堆在现阶段还没有展示出此“基因”,但并不能完全排除其没有易发生核泄漏事故的“基因”。然而,本文研究的是评估一些列核反应堆的内在、外在属性与其发生核泄漏事故“基因”之间的内在联系,却使用这些样本的外在表现来代表其内在发生核泄漏事故“基因”,这必然会给模型结果带来不确定性。此外,本文对预测变量之间存在的关联性进行探讨,以日本的核反应堆为例,由于日本倾向于建造沸水型反应堆 (boiling water reactor, BWR),采取的来自日本的核反应堆样本倾向为 BWR,这些样本均有很相似的燃料类型、冷却液类型以及慢化剂类型等特性。由于日本国的国土面积较小,在靠近海岸线的位置建造核反应堆是不得已选择,会大大增加核反应堆受到海啸或者地震影响而发生核泄漏事故的概率,从而在一些样本核反应堆中,可以很明显得到预测变量之间的关联性。预测变量之间的关联性如图 9 所示。

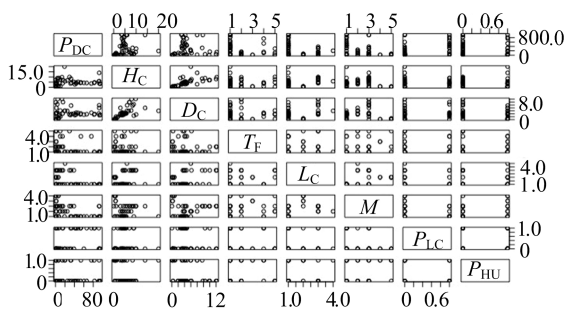


图 9 预测变量之间的相关性

根据以上分析,本文所建立的 LRM 和与

LDM 均由相同的一组数据训练和测试,LRM 和与 LDM 中的预测准确性就是样本数据的训练误差。因此对于这 2 种模型,当其被用于预测一组新的数据时,模型所产生的真实的预测准确性要低于上述所报道的预测准确性。此外本文所建立的 SVM 的预测误差要大于 LRM 和 LDM,说明这 2 种模型在预测新的数据时可能会产生大于本文所报道的预测误差的结果。

4.3 模型对比

本文 6 种模型的预测准确率对比见表 13 所列。对比自由模型与保守模型发现,保守模型的预测准确率要大于自由模型的,除了 LRM 的情况。对比不同模型的预测能力发现,LRM 的预测准确率要强于其他 2 种判别模型。总体来看,3 种不同模型类别均具有基本预测核反应堆发生核泄漏事故风险的能力。

表 13 模型的预测准确率对比 %

模型类别	自由模型	保守模型
LRM	92	88
LDM	86	88
SVM	68	84

另外一种直观地对比 3 种不同类型模型的预测能力的方法就是通过比较 ROC 曲线的 AUC。3 种自由模型与保守模型的 ROC 曲线对比分别如图 10、图 11 所示。

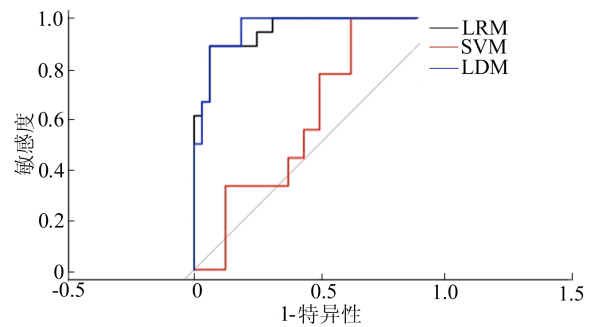


图 10 自由模型的 ROC 曲线对比

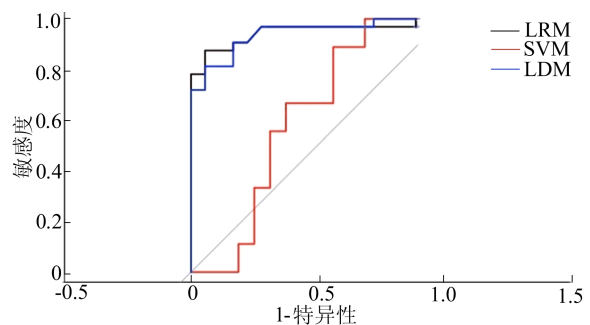


图 11 保守模型的 ROC 曲线对比

由图 10、图 11 可知,LRM 和 LDM 的预测能力均比 SVM 更精确,然而 SVM 有最接近真实值的预测准确性,因此也是解决这一类问题非常好的模型类别,尤其考虑到该风险评估系统的非线性特征。

5 结 论

本文通过构建概率统计分析模型探究核反应堆中的内部和外部因素与核泄漏事故之间的关联性,并对核泄漏风险的影响因素进行评估。通过对 3 种不同模型研究并预测核反应堆发生核泄漏事故风险的安全评估能力。利用样本中的训练数据组与检测数据组构建 3 种不同概率统计分析模型并检测其预测能力。LRM 中自由模型的预测准确率为 92%,保守模型的预测准确率为 88%;LDM 中自由模型的预测准确率为 86%,保守模型的预测准确率为 88%;SVM 中自由模型的预测准确率为 68%,其保守模型的预测准确率为 84%。

通过构建并比较 3 种不同类型的概率统计分析模型发现,核反应堆内部和外部因素均会影响核泄漏事故发生的安全风险。但在这些影响因素中,外部因素比内部因素更具有支配地位,并在整个评估体系中起主要作用。此外,虽然 LRM 和 LDM 在预测计算方面相较于 SVM 更加精准,但其仍然没有 SVM 在这个风险评估体系中更加适用,原因是该风险评估体系是非线性的。

[参 考 文 献]

- [1] FURUKI G, IMOTO J, OCHIAI A, et al. Caesium-rich micro-particles, a window into the meltdown events at the Fukushima Daiichi Nuclear Power Plant [J]. *Scientific Reports*, 2017, 7(1):42731.
- [2] ZYGA L. How does a nuclear meltdown work? [EB/OL]. [2020-01-22]. <https://phys.org/news/2011-03-nuclear-meltdown-video.html>.
- [3] MAHAFFEY J. Atomic accidents: a history of nuclear meltdowns and disasters from the Ozark Mountains to Fukushima[M]. [S. l.]: Open Road Media, 2014.
- [4] ASSOCIATION W N. Reactor database 2016, World Nuclear Association [EB/OL]. [2020-03-20]. <https://www.world-nuclear.org/information-library/facts-and-figures/reactor-database.aspx>.
- [5] IAEA. Research reactor data base (RRDB) [EB/OL]. [2020-01-22]. <https://www.iaea.org/OurWork/ST/NE/NEFW/Technical-Areas/RRS/databases.html>.
- [6] NEWS K, NEWS K. Japan fault under Tsuruga nuclear plant could trigger M7. 4 quake research [EB/OL]. [2020-01-22]. <http://mdn.mainichi.jp/mdnnews/news/20120306p2g00m0dm081000c.html>.
- [7] WALKER J S. Three Mile Island: A nuclear crisis in historical perspective [M]. California: University of California Press, 2004.
- [8] HOSMER D W, LEMESBOW S. Goodness of fit tests for the multiple logistic regression model [J]. *Communications in Statistics-Theory and Methods*, 1980, 9(10):1043-1069.
- [9] HARREL J F E. The practical value of logistic regression [C]//Proceedings of the 10th Annual SAS Users Group International Conference. [S. l. : s. n.], 1985: 1031-1036.
- [10] PLACKETT R L. Karl Pearson and the chi-squared test [J]. *International Statistical Review/Revue Internationale de Statistique*, 1983, 51(1):59-72.
- [11] MCHUGH M L. The chi-square test of independence [J]. *Biochemia Medica*, 2013, 23(2):143-149.
- [12] SHARPE D. Chi-square test is statistically significant; now what? [J]. *Practical Assessment, Research, and Evaluation*, 2015, 20(1):8.
- [13] NAGELKERKE N J D. A note on a general definition of the coefficient of determination [J]. *Biometrika*, 1991, 78(3):691-692.
- [14] TU J V. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes [J]. *Journal of clinical epidemiology*, 1996, 49(11):1225-1231.
- [15] GAO J, QI H, XIA X, et al. Linear discriminant model for information retrieval [C]//Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. [S. l. : s. n.], 2005:290-297.
- [16] MORAIS C L M, LIMA K M G, MARTIN F L. Uncertainty estimation and misclassification probability for classification models based on discriminant analysis and support vector machines [J]. *Analytica Chimica Acta*, 2019, 1063:40-46.
- [17] SANCHEZ A V D. Advanced support vector machines and kernel methods [J]. *Neurocomputing*, 2003, 55(1/2):5-20.
- [18] AN S, LIU W, VENKATESH S. Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression [J]. *Pattern Recognition*, 2007, 40(8):2154-2162.
- [19] GREINER M, PFEIFFER D, SMITH R D. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests [J]. *Preventive Veterinary Medicine*, 2000, 45(1/2):23-41.
- [20] HANLEY J A, MCNEIL B J. The meaning and use of the area under a receiver operating characteristic (ROC) curve [J]. *Radiology*, 1982, 143(1):29-36.
- [21] BRADLEY A P. The use of the area under the ROC curve in the evaluation of machine learning algorithms [J]. *Pattern Recognition*, 1997, 30(7):1145-1159.
- [22] FARAGGI D, REISER B. Estimation of the area under the ROC curve [J]. *Statistics in Medicine*, 2002, 21(20):3093-3106.

(责任编辑 闫杏丽)