

# 基于大数据分析的危化品储存安全隐患预测模型构建

侯凯凯(齐成(山东)石化集团有限公司, 山东 东营 257300)

**摘要:**危险化学品储存安全至关重要,关乎人民生命财产与环境安全。本文针对危化品储存安全隐患,提出基于大数据分析的预测模型构建方法。通过全面收集储存设备运行、环境参数、管理记录等多源数据,经清洗、预处理后,运用特征工程提取关键特征。采用机器学习算法如随机森林、支持向量机等构建预测模型,并利用交叉验证和网格搜索优化超参数。实验结果表明,该模型能有效预测危化品储存安全隐患,为预防事故提供有力支持,对提升危化品储存安全管理具有重要意义。

**关键词:**大数据分析;危化品储存;安全隐患;预测模型

中图分类号:X937 文献标识码:A 文章编号:1674-5167(2025)013-0119-03

## Construction of a Predictive Model for Safety Hazards in Hazardous Chemical Storage Based on Big Data Analysis

Hou Kaikai(Qicheng(Shandong) Petrochemical Group Co., Ltd., Dongying Shandong 257300, China)

**Abstract:**The safety of hazardous chemical storage is of paramount importance, as it concerns the safety of people's lives, property, and the environment. This paper proposes a method for constructing a predictive model based on big data analysis to address safety hazards in hazardous chemical storage. By comprehensively collecting multi-source data such as storage equipment operation, environmental parameters, and management records, the data is cleaned and preprocessed, and key features are extracted through feature engineering. Machine learning algorithms such as random forest and support vector machines are used to build the predictive model, and hyperparameters are optimized using cross-validation and grid search. Experimental results show that the model can effectively predict safety hazards in hazardous chemical storage, providing strong support for accident prevention and significantly improving the safety management level of hazardous chemical storage.

**Keywords:**Big data analysis; hazardous chemical storage; safety hazards; predictive model

危险化学品(简称危化品)在工业生产、日常生活等领域应用广泛,然而其具有易燃、易爆、有毒、腐蚀等特性,储存过程中存在诸多安全风险。一旦发生事故,将造成严重的人员伤亡、财产损失和环境污染。例如,某化工园区的危化品仓库曾因管理不善,发生爆炸事故,导致周边区域遭受巨大破坏,众多人员伤亡,经济损失惨重。因此,加强危化品储存安全管理,提前预测安全隐患,对于保障社会安全稳定发展至关重要。

## 1 危化品储存安全隐患分析

### 1.1 储存设备故障

储存设备是危化品储存的关键载体,其故障可能直接引发安全事故。例如,储罐、管道等设备长期使用,可能出现腐蚀、磨损等问题,导致危化品泄漏。以某石油化工企业为例,由于储罐腐蚀未及时发现和处理,发生了危化品泄漏事故,对周边环境造成了严重污染。阀门、泵等设备的故障也可能影响危化品的正常储存和输送,如阀门密封不严导致泄漏,泵故障造成压力异常等。

### 1.2 环境因素影响

环境因素对危化品储存安全有着重要影响。温度、

湿度的变化可能导致危化品性质改变,增加安全风险。高温环境下,易燃液体的蒸发速度加快,易形成爆炸性混合气体;湿度较大时,某些危化品可能发生潮解、变质等反应。例如,在夏季高温时段,曾有企业因危化品仓库温度控制不当,引发火灾事故。此外,通风不良会导致仓库内有害气体积聚,达到爆炸极限,遇火源即可能引发爆炸。

### 1.3 人为管理失误

人为管理失误是危化品储存安全隐患的重要来源。人员操作不规范,如违规装卸、搬运危化品,可能导致容器破裂、泄漏。某企业在装卸过程中,因工作人员未按操作规程作业,致使危化品容器倾倒,造成泄漏事故。安全管理制度不完善或执行不到位,如巡检不及时、记录不准确等,也无法及时发现和处理安全隐患。部分企业对员工的安全培训不足,员工安全意识淡薄,缺乏应对突发事故的能力,进一步增加了事故发生的风险。

## 2 大数据分析在危化品储存安全管理中的应用优势

### 2.1 数据整合与处理能力

大数据技术能够整合来自不同来源、不同格式的危化品储存相关数据,包括设备传感器数据、环境监

测数据、管理记录数据等。通过高效的数据清洗、转换和集成技术，将这些数据统一格式，为后续分析提供高质量的数据基础。

例如，利用大数据平台可以实时采集和整合危化品仓库中各类传感器上传的温度、压力、液位等数据，以及设备运行状态数据，打破数据孤岛，实现数据的全面共享和协同分析。

## 2.2 挖掘潜在规律与关联

大数据分析具备强大的挖掘能力，能够从海量数据中发现潜在的规律和关联。通过数据挖掘算法，可以分析危化品储存过程中各种因素之间的关系，如设备故障与环境参数、操作行为之间的关联。例如，通过对历史数据的分析，发现当仓库温度超过一定阈值且通风不良时，设备故障发生率显著增加，从而为制定针对性的安全管理措施提供依据。

## 2.3 实时监测与预警

基于大数据分析的危化品储存安全管理系统能够实现实时监测和预警。通过实时采集和分析数据，系统可以及时发现异常情况，并发出预警信号。例如，当监测到储罐压力异常升高或气体泄漏浓度超标时，系统立即向管理人员发送预警信息，以便及时采取措施，避免事故发生。这种实时监测和预警功能大大提高了安全管理的及时性和有效性。

# 3 危化品储存安全隐患预测模型构建

## 3.1 数据收集与预处理

### 3.1.1 多源数据收集

全面收集危化品储存相关的多源数据。从储存设备传感器获取温度、压力、液位、振动等实时运行数据，这些数据能够直观反映设备的工作状态。收集仓库环境监测数据，包括温度、湿度、空气质量、通风情况等，以了解环境因素对危化品储存的影响。同时，整理安全管理记录数据，如人员操作记录、巡检记录、设备维护记录等，这些数据包含了人为管理方面的信息。此外，还可收集历史事故数据，分析事故发生的原因、经过和后果，为模型训练提供参考。

### 3.1.2 数据清洗与预处理

对收集到的数据进行清洗，去除重复、错误和缺失的数据。对于缺失值，根据数据特点采用合适的方法进行填补，如均值填充、中位数填充或基于模型的预测填充。对数据进行标准化处理，将不同量级和单位的数据转换为统一的标准尺度，以提高模型的训练效率和准确性。例如，对温度、压力等数据进行归一化处理，使其取值范围在 [0,1] 之间。同时，对类别型数据进行编码，如将设备状态（正常、故障）编码为数字，以便模型能够处理。

## 3.2 特征工程

### 3.2.1 特征提取

从预处理后的数据中提取关键特征。对于设备运行数据，提取特征如设备运行时间、温度变化率、压力波动幅度等，这些特征能够反映设备的运行稳定性和潜在故障风险。从环境数据中提取特征，如平均温度、湿度变化范围、通风时长等，用于评估环境因素对危化品储存的影响程度。在管理数据方面，提取特征如巡检间隔时间、违规操作次数、安全培训时长等，以衡量人为管理的质量和水平。

### 3.2.2 特征选择

采用特征选择算法，从提取的众多特征中筛选出对安全隐患预测最具影响力的特征。常用的特征选择方法包括过滤法、包装法和嵌入法。过滤法通过计算特征与目标变量之间的相关性、信息增益等指标，对特征进行排序，选择排名靠前的特征。例如，使用皮尔逊相关系数计算特征与安全隐患发生概率之间的相关性，去除相关性较低的特征。包装法将特征选择看作一个搜索问题，通过训练模型来评估不同特征子集的性能，选择性能最优的特征子集。嵌入法则在模型训练过程中自动选择重要的特征，如决策树算法在构建决策树时会根据特征的重要性进行分裂。通过特征选择，可以减少数据维度，降低模型复杂度，提高预测模型的准确性和泛化能力。

## 3.3 模型选择与训练

### 3.3.1 机器学习算法选择

在危化品储存安全隐患预测模型构建中，选择合适的机器学习算法至关重要。随机森林算法是一种常用的集成学习算法，它通过构建多个决策树，并将决策树的预测结果进行综合，来提高预测的准确性和稳定性。随机森林能够处理高维数据，对噪声和缺失值具有较强的鲁棒性，且不易过拟合。支持向量机算法则在小样本、非线性分类问题上表现出色，它通过寻找一个最优的分类超平面，将不同类别的数据分开。在危化品储存安全隐患预测中，安全隐患与各种特征之间可能存在复杂的非线性关系，支持向量机能够有效地捕捉这种关系。此外，逻辑回归算法也可用于预测安全隐患发生的概率，它简单高效，可解释性强，能够清晰地展示各个特征对安全隐患的影响程度。

### 3.3.2 模型训练与优化

使用训练数据集对选定的机器学习算法进行模型训练。在训练过程中，通过交叉验证的方法来评估模型的性能。交叉验证将训练数据集划分为多个子集，每次使用其中一个子集作为验证集，其余子集作为训练集，多次训练和验证模型，然后取平均性能指标作

为模型的最终性能评估。例如，采用十折交叉验证，将数据集平均分成十份，依次用其中一份作为验证集，其余九份作为训练集，重复训练和验证十次，最后将十次的性能指标求平均。同时，利用网格搜索等方法对模型的超参数进行优化。网格搜索通过在指定的超参数空间中进行穷举搜索，尝试不同的超参数组合，选择使模型性能最优的超参数设置。

### 3.4 模型评估与验证

#### 3.4.1 评估指标选择

为了准确评估危化品储存安全隐患预测模型的性能，选择合适的评估指标至关重要。准确率是评估模型预测准确性的常用指标，它表示预测正确的样本数占总样本数的比例。在危化品储存安全隐患预测中，准确率高意味着模型能够准确地判断出安全隐患是否存在。召回率则反映了模型对实际存在安全隐患的样本的覆盖程度，即实际存在安全隐患且被模型正确预测出来的样本数占实际存在安全隐患样本总数的比例。F1值是综合考虑准确率和召回率的指标，它能够更全面地评估模型的性能。

#### 3.4.2 模型验证方法

采用多种方法对构建的预测模型进行验证。使用独立的测试数据集对模型进行验证，将模型在训练过程中未见过的数据输入模型，观察模型的预测结果与实际情况的符合程度。如果模型在测试数据集上表现良好，说明模型具有较好的泛化能力，能够对新的数据进行准确预测。同时，可以采用时间序列验证的方法，按照时间顺序将数据划分为训练集和测试集，先使用较早时间的数据进行模型训练，然后用较晚时间的数据进行测试，以验证模型在不同时间段的预测性能。

### 4 案例分析

#### 4.1 数据收集与处理过程

以某大型危化品储存企业为例，该企业拥有多个危化品仓库和储存设施。收集了该企业近三年的储存设备运行数据、仓库环境数据、安全管理记录数据以及历史事故数据。在数据收集过程中，通过在储存设备上安装传感器，实时采集设备的温度、压力、液位等运行数据；利用环境监测系统收集仓库内的温度、湿度、空气质量等环境数据；从企业的安全管理信息系统中获取人员操作记录、巡检记录、设备维护记录等管理数据。对收集到的数据进行清洗，发现并纠正了部分错误数据，填补了缺失值。例如，对于设备运行数据中的少量缺失温度值，采用线性插值法进行了填补。然后对数据进行标准化处理，将所有数据转换为统一的尺度，以便后续分析。

### 4.2 模型构建与训练结果

基于处理后的数据，采用随机森林算法构建危化品储存安全隐患预测模型。在特征工程阶段，提取了包括设备运行稳定性、环境参数变化、管理操作规范性等多个方面的特征。通过特征选择，筛选出了对安全隐患预测影响较大的特征，如设备温度变化率、仓库通风不良时长、违规操作次数等。在模型训练过程中，使用十折交叉验证对模型进行评估，并利用网格搜索对随机森林的超参数进行优化。经过多次实验，确定了最优的超参数组合：决策树数量为100，最大深度为8，最小样本分裂数为5。训练后的模型在训练数据集上的准确率达到了90%，召回率为85%，F1值为87.5%。

### 4.3 模型预测性能评估

使用独立的测试数据集对构建的随机森林模型进行预测性能评估。测试结果显示，模型在测试数据集上的准确率为88%，召回率为83%，F1值为85.4%。与其他常用的机器学习模型（如支持向量机、逻辑回归）相比，随机森林模型在准确率、召回率和F1值等指标上均表现更优。同时，通过时间序列验证发现，模型在不同时间段的预测性能较为稳定，能够有效地预测危化品储存过程中的安全隐患。例如，在对近期发生的一起安全隐患事件的预测中，模型提前发出了预警，为企业及时采取措施避免事故发生提供了有力支持。

### 5 结论

本文通过对危化品储存安全隐患的分析，阐述了大数据分析在危化品储存安全管理中的应用优势，并构建了基于大数据分析的危化品储存安全隐患预测模型。通过多源数据收集与预处理、特征工程、模型选择与训练以及模型评估与验证等一系列步骤，实现了对危化品储存安全隐患的有效预测。案例分析表明，所构建的随机森林模型在预测危化品储存安全隐患方面具有较高的准确率、召回率和F1值，能够为企业的安全管理提供科学依据，帮助企业提前发现和处理安全隐患，降低事故发生的风险。

#### 参考文献：

- [1] 高舒珊, 李燕勇, 陈俍, 等. 危化品全生命周期安全管控模式在危化品使用单位中的研究应用 [J]. 当代化工, 2019, 48(1):5.
- [2] 曹国龙, 庞小辉, 崔聪聪. 化工企业危化品储存安全管理探究 [J]. 中国化工贸易, 2023:190-192.
- [3] 周琳. 我国危险化学品事故的安全管理体系原因研究 [D]. 中国矿业大学(北京), 2025.