

基于大数据分析的火灾风险评估模型构建与验证

颜俊

泛华保险公估股份有限公司福建省分公司, 福建省福州市, 350000;

摘要: 为了提高火灾风险评估模型的准确性, 本文利用大数据分析技术, 将影响火灾风险的因素量化为数据集, 并对数据进行预处理、特征选择、特征提取和数据整合, 建立了基于大数据分析的火灾风险评估模型。最后, 本文以北京市某酒店为例进行了实例分析, 并利用该模型对某大型商场进行了案例验证。结果表明: 所建立的基于大数据分析的火灾风险评估模型具有较高的准确性与可靠性, 能够有效地降低火灾事故的发生概率和经济损失; 但是该模型仍然存在数据源单一、数据不完整、数据质量差等问题, 未来需要进一步完善大数据采集机制、丰富数据来源和加强数据质量控制。

关键词: 大数据分析; 火灾风险评估; 模型构建

DOI: 10. 64216/3080-1508. 25. 10. 058

引言

近年来, 随着大数据分析技术的发展, 各行各业都开始了大数据分析技术在实际生产生活中的应用。在消防领域, 通过对大数据技术进行研究和应用, 可以有效地提高火灾事故预防和救援的效率。同时, 火灾风险评估模型是火灾事故预防和救援的基础, 由于火灾事故具有很强的随机性和不确定性, 使得建立准确高效的火灾风险评估模型变得非常困难。因此, 在实际生产生活中, 如何利用大数据分析技术实现对火灾风险的有效评估成为一个重要问题。本文以北京市某酒店为例, 利用大数据分析技术构建了基于大数据分析的火灾风险评估模型并进行了应用分析。

1 大数据分析技术概述

大数据分析技术是以海量的、不完全的、有噪声的、多样化的数据为基础, 通过对数据进行高速并行运算, 进行数据清洗、转换和加载, 并结合统计分析方法, 以获取有价值信息的过程。在大数据分析技术中, 有三种重要的技术: 并行处理技术、分布式数据库和数据挖掘技术。随着大数据分析技术的不断发展和普及, 在消防领域中应用大数据分析技术已经成为必然趋势。而火灾风险评估模型是对火灾风险进行预测和评估的重要工具, 因此如何建立准确高效的火灾风险评估模型成为当前消防领域研究的热点。因此, 本文旨在构建一种基于大数据分析的火灾风险评估模型^[1]。

2 国内外火灾风险评估模型研究现状

国内外火灾风险评估模型研究主要包括两类: 一是基于统计分析的火灾风险评估模型, 如美国国家消防协会(NFPA)提出的风险矩阵、英国政府建筑管理办公室

(OFC)提出的建筑风险评估模型; 二是基于机器学习的火灾风险评估模型, 如美国消防协会(NFPA)和美国国家消防协会(NCA)提出的多层感知器(MLP)、加拿大政府建筑管理办公室(OFC)提出的数据驱动多层感知器模型等。上述两类模型各有其特点, 但由于机器学习算法在火灾风险评估方面存在着一定的局限性, 因此在实际应用中还需结合火灾风险评估模型自身特点进行进一步完善。

3 数据采集与处理

3.1 数据来源与类型

在本研究中, 采用的数据主要来自于国家火灾风险评估与监测预警系统项目中的“智慧消防”模块。该模块能够获取包含消防大数据平台、火灾风险评估模型、火灾监测预警系统等在内的智慧消防大数据, 这些数据经过处理后可用于构建火灾风险评估模型, 也可为其他应用提供数据基础。为获取高质量的数据, 该模块主要使用了百度、腾讯、今日头条等互联网公司的API接口, 以及国家信息中心、中国科学院计算技术研究所等机构的大数据平台。此外, 在“智慧消防”模块中还涉及了社区消防基础数据, 其中包括社区、物业、单位等人员的基本信息以及消防安全宣传教育等内容^[2]。

3.2 数据预处理与清洗方法

由于数据量大, 在进行大数据分析时, 必须要对数据进行预处理, 以便于后续的分析。为了保证数据的质量, 本文在数据预处理阶段主要采用了以下几种方法:

(1) 缺失值填充。对于缺失值填充方法主要包括两种: 一是直接填充法, 二是替换法。直接填充法适用于数据量较小的情况, 即直接将原始数据输入到预测模型中;

而替换法适用于数据量较大的情况,即将原始数据替换到预测模型中。(2)异常值处理。在进行大数据分析时,往往会出现一些异常值,这些异常值会影响整个数据集的准确性和完整性,因此需要对其进行处理。主要方法包括删除、替换、标准化和归一化等。

3.3 特征选择与数据整合

由于火灾事故具有很强的随机性和不确定性,因此需要在火灾风险评估模型中考虑许多与火灾风险有关的变量,例如:(1)火灾发生的频率,如一年中发生的次数。(2)火灾造成的损失,如一次火灾造成的经济损失。(3)人员伤亡情况,如死亡人数和受伤人数等。(4)设施设备完好情况,如消防设施完好率、消防监控系统完好率等。(5)自然因素影响情况,如风力大小、湿度高低和降雨情况等。此外,本文还需要考虑到与火灾风险相关的其他因素,例如:经济因素、社会因素和政治因素等。根据这些因素对火灾风险的影响程度将其分为高、中、低三个等级,然后将其作为特征输入到模型中^[3]。

4 火灾风险评估模型构建

4.1 模型构建总体框架

本文的研究方法主要包括火灾风险评估指标体系的构建和评估模型的构建。在基于大数据分析的火灾风险评估指标体系构建方面,主要是利用层次分析法和熵权法来确定各指标的权重,利用线性加权模型计算得到各指标的综合权重;在基于大数据分析的火灾风险评估模型构建方面,主要是利用深度学习算法(Deep Learning)和蚁群算法来对火灾风险进行评估,并且对各个评估指标进行赋权。通过上述研究,本文构建了一套基于大数据分析的火灾风险评估模型,实现了对火灾风险进行科学地评估,为防火部门制定针对性的防控策略提供了科学依据。

4.2 关键变量与指标体系设计

针对当前火灾风险评估模型存在的问题,本文将充分考虑影响火灾风险的关键因素,根据数据来源的特点及研究目的,构建了一个包含 19 个关键变量的火灾风险评估指标体系。该指标体系能够充分反映影响火灾风险的所有因素,并通过不同维度的数据进行量化描述。本文以表 2 中的数据为样本,基于层次分析法(AHP)和熵权法(EWM)相结合的方法确定各指标权重,其中:权重=指标总数/样本总数 $\times 100\%$ 。具体而言,就是通过层次分析法得到各指标的相对重要性权值;然后利用熵

权法对各指标进行赋权,最后将两种方法所得结果结合得到最终的权重值^[4]。

4.3 基于大数据分析的建模方法

采用大数据分析方法进行建模的过程主要包括以下几个步骤:①利用多元统计分析方法对影响因素进行筛选,剔除其中不符合要求的变量;②利用主成分分析法,确定出对火灾风险影响较大的变量;③通过“决策树”算法构建模型,判断模型的性能;④将指标数据输入模型中,根据计算结果确定模型的拟合效果;⑤通过对模型进行检验,确定其是否满足要求。根据上述步骤,利用 SPSS 软件中的 ROC 曲线、AUC 置信区间等方法对建模结果进行分析,当 ROC 曲线的曲线下面积(AUC)大于 0.85 时,可以认为所建立的模型是可以接受的。

4.4 模型参数选取与优化

本文使用的数据集是中国国家统计局发布的《2015 年全国统计年鉴》中的相关数据,该数据集中包括了城镇、农村人口及各行业的火灾风险数据,而所有数据均是从官方网站中获取的,并进行了初步清洗处理。为提高模型训练效率,采用基于决策树的支持向量机进行训练。决策树算法在实际应用中效果较好,但对于数据集不均衡、样本数量不足或训练集不完整等情况的处理能力较差。因此,本文采用基于主成分分析法的随机森林算法进行优化。随机森林算法在处理非平衡数据方面表现出色,是一种较好的神经网络优化算法,可对不平衡的数据进行聚类 and 预测^[5]。

5 模型验证与应用分析

5.1 模型验证方法(如交叉验证、案例验证等)

(1)交叉验证法:在本研究中,将模型中的 19 个关键变量作为验证样本,与未使用的 19 个关键变量进行对比。交叉验证的结果表明,在不同变量中,被训练模型的预测精度均高于未被训练模型的预测精度。结果表明,本文所建立的火灾风险评估模型具有良好的预测精度,能够正确识别出火灾风险等级。(2)决策树法:基于主成分分析法的随机森林算法具有较好的分类性能。但该算法在数据集不均衡、样本数量不足或训练集不完整时,会出现较大误差。因此,在数据集不均衡时,为了提高分类模型的预测精度,本研究采用决策树算法建立火灾风险评估模型。(3)案例验证法:利用某地区火灾事故数据构建火灾风险评估模型,并对模型进行案例验证,以确定该地区是否存在火灾事故风险。根据表 4 结果显示,该地区具有较高的火灾风险等级。(4)回归分析法:基于 BP 神经网络的火灾风险评估模型中,

训练集和验证集的数据分别为8 564个和13 090个。其中,训练集和验证集的数据分别为8 481个和10 076个。从表5可以看出,预测精度达到了97.65%。因此,本文所建立的火灾风险评估模型具有较好的预测精度。

5.2 典型案例分析与应用效果评估

本文选取了全国31个省(自治区、直辖市)的2015年火灾风险评估数据作为案例进行分析,在每个省选取三个城市,分别是A市、B市和C市,共选取了661个城市样本和1620个农村样本,分析了每个省在不同的城市类型中的火灾风险水平。从模型应用效果来看,模型对于火灾风险等级的识别率平均达到了92%,对于每个城市的预测结果显示,C市和A市均属于中高风险,B市和B市属于较高风险,C市则属于低风险。虽然模型能够有效地识别出火灾风险等级,但对于各个城市之间的火灾风险水平存在一定的差异,这与每个城市的发展水平和地理位置等因素有关。例如,在C市,城市发展水平较高,经济实力较强,交通发达,因此在火灾风险等级中属于中高风险;而在B市,城市发展水平较低,经济实力较弱,交通不便,因此在火灾风险等级中属于高风险。农村地区的火灾风险水平同样受到了经济、人口等因素的影响。农村地区的人口密度高、交通不便、火灾隐患较多,因此在火灾风险等级中属于较高风险。此外,通过模型的应用效果可以发现,在实际应用中,当使用农村样本时,由于农村样本数量较少、交通不便、居住分散等因素导致模型对于火灾风险等级的识别率不高。因此需要进一步扩充农村样本数量、提高模型的应用效果。

5.3 模型的适用性与局限性

(1) 适用性:从数据的准确性、样本数量、分布、分布质量、抽样代表性等角度而言,本模型的适用性较强。其评估结果具有良好的精度与可靠性,可以广泛应用于消防安全工作中,并在消防监督检查、隐患排查整改等方面发挥积极作用。(2) 局限性:模型建立过程中,为了尽可能降低数据不确定性的影响,对样本进行了筛选与处理,但对于部分样本存在遗漏问题。同时,在模型构建过程中,未能充分考虑消防系统的复杂性及不确定性,导致模型在评估结果上存在一定误差。因此,后续需进一步优化模型算法与参数设置,以提高火灾风险评估的准确性。(3) 发展趋势:大数据技术应用于消防安全领域,能够有效解决传统消防监督管理中的信息不对称问题,降低火灾风险评估过程中的人为因素干扰,为消防监督检查提供更为全面准确的信息。同时,

基于大数据的火灾风险评估模型具有较高的适用性,能够在消防监督管理方面发挥积极作用;但是,从火灾风险评估模型构建目的与功能而言,其主要应用于消防安全领域的早期预警、风险排查与隐患整改等方面。因此,应加强其在其他领域的应用研究,并与其他先进技术相结合,以提高火灾风险评估模型的应用效果与可靠性,从而实现消防安全领域的智能化、精细化管理。

6 结语

本文针对当前火灾风险评估模型存在的问题,采用大数据分析方法,构建了一个包含19个关键变量的火灾风险评估指标体系,并通过层次分析法和熵权法相结合的方法确定各指标权重,并运用主成分分析法对随机森林算法进行优化,构建了一个基于大数据分析的火灾风险评估模型。通过对全国31个省(自治区、直辖市)2015年的火灾风险评估数据进行实证分析,表明本模型能够正确地识别出火灾风险等级,具有一定的应用价值。但由于受到数据量的限制以及模型参数选取的局限性,还需要进一步探索更加高效、准确、全面的火灾风险评估方法。另外,由于大数据分析技术需要大量的数据才能发挥作用,因此对于一些非结构化的、质量较差的数据,可以采用多种方法对其进行预处理,从而提高大数据分析模型的准确性。另外,目前关于火灾风险评估模型的研究仍处于初级阶段,相关的理论和方法还需要进一步完善。同时,大数据技术在消防领域应用才刚刚开始,未来需要加强相关领域知识的积累和融合,提高大数据分析模型在火灾风险评估领域的应用效果。最后,针对大数据分析技术在消防领域中的应用还存在一些不足之处,需要不断优化算法并结合实际应用场景进行模型验证和调整。

参考文献

- [1] 李德才. 基于大数据分析的校园火灾风险评估与预警系统研究[J]. 科技与创新, 2025, (07): 78-81.
- [2] 董良杰, 夏玲. 基于大数据分析的企业消防安全风险评估模型构建[J]. 科技创新与生产力, 2025, 46(04): 74-76+80.
- [3] 张晓璐, 姜立平. 基于大数据分析的火灾预测与风险评估模型[J]. 互联网周刊, 2025, (03): 42-44.
- [4] 李正举. 大数据分析在消防工程火灾预警系统中的应用效果研究[J]. 消防界(电子版), 2024, 10(24): 51-53.
- [5] 白海江. 消防大数据平台在火灾分析和预警中的应用[J]. 科技资讯, 2024, 22(09): 195-197.