

# 基于知识驱动的建筑风环境灾害评估平台 研究与实现

闵妍涛<sup>1</sup>, 冷烁<sup>2</sup>, 胡振中<sup>1</sup>, 林佳瑞<sup>2</sup>, 于言滔<sup>3</sup>

- (1. 清华大学国际研究生院, 广东 深圳 518000;
2. 清华大学土木工程系, 北京 100086;
3. 香港科技大学, 香港 999077)

**【摘要】**建筑风灾的灾前预防评估和灾后应对对于防灾减灾工作而言相当重要。现今主流的试验、计算方法往往结果精确但各方面的成本较高。频发的风灾提供了大量的数据。本课题通过收集既往风灾数据, 使用机器学习、近邻检索、知识图谱等技术, 实现建筑风灾致损情况预测、相似案例检索、知识自动问答功能, 构建建筑风环境灾害评估平台, 实现对建筑风环境致损的快速预测判断, 并能提供相关知识, 为制定风灾应对策略提供参考。

**【关键词】**建筑风灾致损; 机器学习; 近邻检索; 知识图谱;

## 1 前言

我国是世界上遭遇台风袭击最多的国家之一, 随着人口和社会财富的增长, 风灾应对的重要性愈发凸显。

当前已有多种技术可用以获取建筑风环境信息, 如现场实测、风洞试验和计算机数值模拟等<sup>[1]</sup>。基于试验或计算结果, 结合结构抗力模型等知识, 可以实现建筑在特定风环境下破坏情况的预测评估。这些评估方案结果精确, 但往往需要较高的试验或计算成本和对专业知识的深入了解。

我国风灾频发的同时产生了大量数据, 其中隐含着信息, 经过技术手段加工, 可以获得规律和知识。涉及的技术包括深度学习、进化算法等。目前已有类似研究将人工智能技术应用于这一领域, 该类方法无需对结构及风场理论知识的深入了解, 也可免于繁重的计算程序, 是一种提供建筑风灾致损评估的快速、简易的途径。

本课题的研究内容以风灾致损建筑案例为基础, 应用多种算法构建建筑风环境灾害评估平台, 包括风灾致损预测、相似案例推荐和自动问答三个功能模块, 三者相互配合, 实现预测评估流程的简化, 并为制定应对策略提供参考。

## 2 相关研究综述

在风灾结构易损性领域, 除了主流的工程和经验方法, 采用人工智能技术的易损性评

估模型也得到应用，如 Sandri P.<sup>[2]</sup>使用人工合成的数据对人工神经网络进行训练，并发现模型能够有效习得输入特征与建筑物损坏等级之间的关系。

除上述技术手段，机器学习等技术也同样可用于获取相关判断。机器学习基于数据建立模型，描述其间隐含规律，从而对问题做出合适的决策。其中，采用有标签样本进行训练的监督学习<sup>[3]</sup>可以获得较高质量的分类或回归模型，从而实现相关预测和推断。监督学习算法包括支持向量机、随机森林等。

与机器学习直接描述数据中隐含规律的概念不同，近邻检索借由相似度，通过“类比推理”给出参考或作出决策。除了逐一比对查询点和样本点的传统方法，目前已有多种近似最近邻搜索方案以少量准确率换取次线性时间复杂度，提升检索效率。HNSW<sup>[4]</sup>是目前最常用的基于图的方案，它借鉴跳表的思路，通过分层结构降低了时间复杂度，但这类方案也具有训练、存储成本高的缺点。

随着信息爆炸式增长，传统搜索引擎力有不逮，自动问答系统应运而生，其本质是一种信息检索系统。现有的问答系统包括社区问答、检索式问答和知识库问答。知识库问答优势在于可以实现基于图的深层逻辑推理，而不止于浅层的语义匹配。其核心为将用户输入的自然语言语句转化为结构化查询语句，包括传统的 lambda 范式<sup>[5]</sup>及深度学习方方法<sup>[6]</sup>等。

### 3 评估平台构建流程

参考业务应用的三层架构，构建建筑风环境灾害评估平台，包括用户交互层、业务功能层、数据层。据此提出建筑风环境灾害评估的技术路线，如图 1 所示。

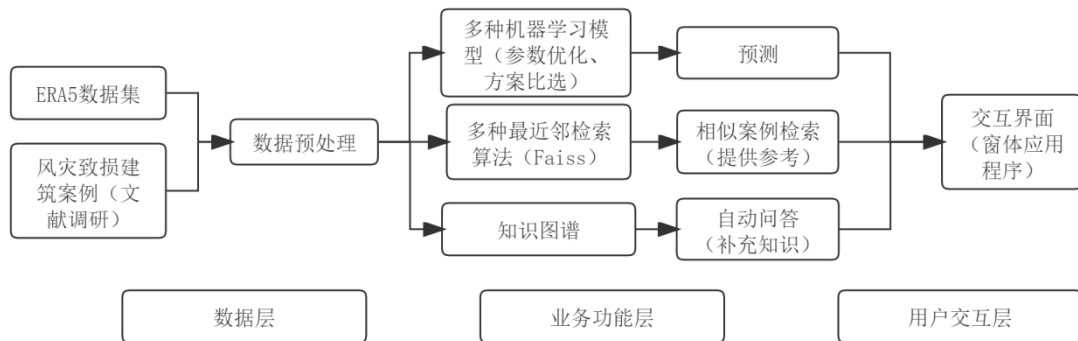


图 1 技术路线

### 4 知识驱动的建筑风环境灾害评估

收集风灾致损建筑案例及风环境数据并进行预处理，将非结构化的文本转化为特征明确的结构化数据，以此为数据基础构建三个功能模块。

#### 4.1 数据收集与处理

本研究数据收集处理工作主要有两部分：

- （1）风灾致损建筑案例

通过中文数据库及搜索引擎等，以“风灾”“建筑”等为关键词进行检索，收集信息相对完整的风灾致损建筑案例，共获得 340 条。然后对案例进行三方面的处理：特征选取，数据预处理及破坏等级划分。

台风灾害包含三个要素：孕灾环境、致灾因子以及承灾体<sup>[7]</sup>。本研究依据这一概念以及信息完整度，确定描述案例的特征集{孕灾环境：风速；致灾因子；承灾体：结构类型，使用年限，规范程度（建筑本身特征），屋盖承重/非承重构件损伤，竖向承重/非承重结构损伤，门窗、玻璃幕墙和内部构件损伤（建筑破坏特征）}。

进行特征选取后，进行三种方式的数据预处理。数据转换针对分类变量，基于字典将其转化为定距型数值变量。数据清理即处理缺失值。对于定距型数值变量，取该特征存在值的平均值以插补。数据不平衡是指训练集中各类样本比例失衡。由于案例数量有限，采用过采样的方式进行处理，并注意在复制样本时添加随机扰动。最后根据《灾损建（构）筑物处理技术规范》CECS 269<sup>[8]</sup>等多部规范人工标注破坏等级，以量化建筑破坏现象。

## （2）ERA5 数据集

从欧洲中期天气预报中心下载 ERA5 数据集。数据采用经纬网格形式，分辨率为 0.25°，时间间隔为 1 小时。本研究选择地表 10m 风东西向分量（下称 u）、地表 10m 风南北向分量（下称 v）进行下载。数据处理包括两步：首先将 u、v 分量合成为地表 10m 高度处水平风速；然后排序得出区域内最大 10m 水平风，代表特征集中描述风环境的“风速”特征。

## 4.2 建筑风灾致损评估框架

为了快速获取对于单体建筑在一定风环境下可能破坏情况的了解，基于结构化数据构建建筑风灾致损评估框架。这一框架包括下述两个功能模块，因其间关联紧密。

### （1）基于机器学习的建筑风灾致损预测

这一模块可以快速、大致地评估判断建筑风灾致损情况，其业务逻辑如下图 2 所示。

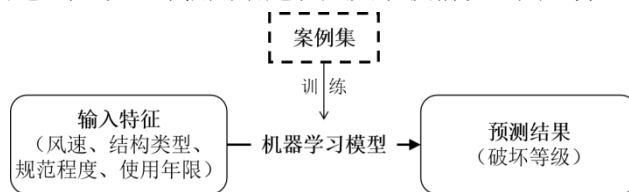


图 2 风灾致损预测模块业务逻辑

获取上图中四个维度的输入特征后，使用机器学习模型获得对其破坏等级的预测。本研究采用监督学习算法构建预测模型。进行数据 Z-score 标准化后，采用 sklearn 库及 xgboost 库构建算法模型，包括随机森林、XGBoost、Adaboost、Bagging、支持向量机。以最小化  $(1-R^2)$  为目标，使用 hyperopt 进行超参数优化。对各模型优化后的指标进行比较，选出相对最优模型用于预测。综合模型稳定性，随机森林模型相对最优，其指标如表 1 所示。

表 1 随机森林模型优化后指标

MSE	MAE	R <sup>2</sup>	accuracy
0.686	0.420	0.618	0.709

(2) 基于近邻检索的相似案例推荐

由于机器学习模型只能获得对破坏等级的预测，既不够准确，也无法获知破坏细节，故构建相似案例推荐模块，提供情景相似的案例以供补充参考，其业务逻辑如图 3 所示。

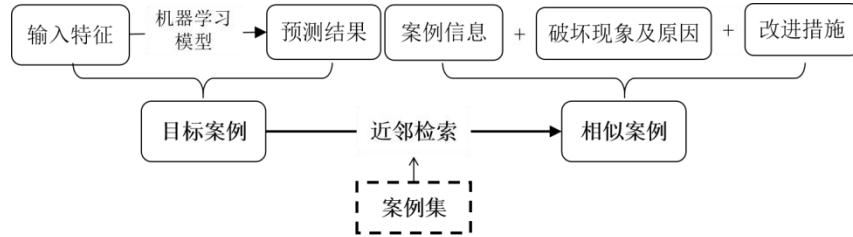


图 3 相似案例推荐模块业务逻辑

用户输入建筑及风环境特征后，经过预测及近邻检索可得相似案例。其中近邻检索的距离度量依赖向量相似度实现。使用 Faiss 构建检索算法，实现的 4 种检索算法为欧氏距离、余弦相似度两种相似度测度和暴力检索、HNSW 检索两种检索方式的组合。获取 50 条测例，使用 4 种算法分别进行相似案例检索，表 2 为某一测例检索结果。比较不同算法检索结果，发现距离度量方式影响获取相似案例的特征，但各有其合理性，可互为补充；综合准确度和效率，HNSW 优于暴力检索。

表 2 检索测例

输入测例	算法	相似案例
{24.73,砌体结构,农村自建,一般,轻微损坏}	欧式距离	{26.16,砌体结构,农村自建,一般,中度损坏}
	×HNSW 检索	坡屋面损坏;部分檩条损坏 屋面抗风能力较弱

4.4 基于知识图谱的自动问答系统

通过前两个模块了解的是某一建筑的情况，所知也仅限于此，因此构建基于知识图谱的自动问答系统，提供综合性知识，作为评估框架的背景性补充。

(1) 知识图谱构建

知识图谱以结构化的方式表示现实事物概念及其间联系，其本质是一种语义网络。逻辑上，知识图谱可分为数据层和模式层<sup>[9]</sup>。本研究采取自上而下的方式<sup>[10]</sup>构建知识图谱，即先设计模式层，再据此抽取知识形成数据层。模式层设计参考台风灾害三要素和事件传统表示模型，据此定义 6 类实体、11 种属性及 7 种关系。所得模式层如图 4 所示。

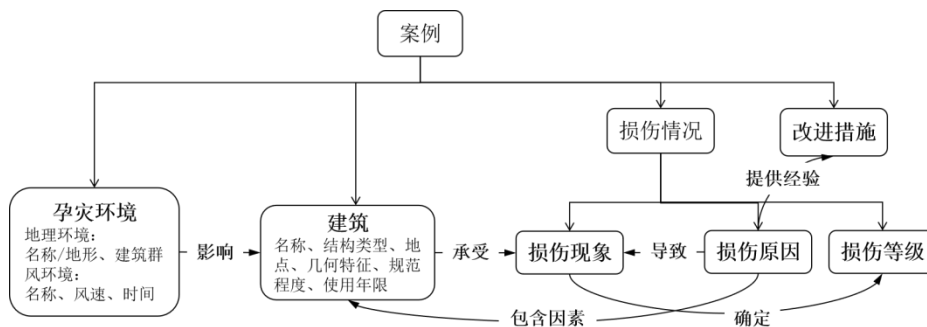


图4 知识图谱

然后基于模式层，从已有结构化数据中提取三元组，使用 Neo4j 图数据库进行存储。

### (2) 自动问答系统

基于知识图谱构建自动问答系统，通过三个步骤实现：问句语义解析、信息检索与答案生成。问句语义解析可拆分为三个子任务：首先根据实体生成领域词典，使用 Aho - Corasick 匹配算法[78]实现自然语言问句中命名实体的识别；然后根据现有数据及实际需求，确定 4 类问题及相应关键词集，对问句进行关键词匹配，确定问句类型；再根据问题类别以及命名实体生成相应的查询语句，此即 Neo4j 相应的 Cypher 查询语句。然后使用查询语句在图数据库中进行检索，并根据需要对返回数据做进一步处理。最后，将处理所得套入与问题类型对应的答句模板，生成自然语言答案返回给用户。4 类问题说明如表 2 所示。

表2 4类问题

问题类型	问题示例	类别 1	具体说明
1 comparison	xx 结构和 xx 结构相比，哪种抗风能力更好？	关键词集	{“相比”...“较”}
2 reasons	xx 结构破坏，可能有什么原因？		'根据统计，{0}的抗风
3 improvements	xx 结构的加固措施	答句模板	性能更好'.format (self.
4 phenomena	xx 结构可能会出现什么样的破坏？		Compare (answers))

注：Compare 函数返回的是某两类结构根据破坏等级在统计意义上的比较结果

## 5 建筑风环境灾害评估平台设计与实现

以格式化数据为基础，整合三种功能，设计建筑风环境灾害评估平台的整体架构，包含三层：用户交互层、业务功能层和数据层。交互层收集用户指令并进行反馈，是系统的对外窗口，基于 Visual Studio 2012 构建 WinForm 窗体小程序予以实现，如图 5 所示。业务功能层接受交互层传来的信息和指令，调用相应功能模块对从数据层获取的数据进行处理，具体实现过程见第 4 节。数据层收集、处理并存储相关数据，是整个架构的数据基础。



图 5 运行案例

具体运行案例如图 5 所示。用户输入的案例特征及预测结果如图。据此进行相似案例检索，获取的相似案例为{钢结构，一般，较正规，24.22，严重损坏}；损伤情况为“屋面坍塌;屋架与竖向承重结构连接处坍塌”；损伤原因及相应改进措施如图。可见预测结果及相似案例较为一致，符合预期。输入问题及获得回答如图，以获取钢结构风灾预防的相关经验知识。该回答为检索钢砌体结构案例改进措施所得。这一类问答主要存在的问题是未能对知识进行整合，形成体系，仅能列举部分案例作为参考。

## 6 结语

本研究运用文献调查法收集整理风灾致损建筑案例，并经过数据处理将非结构化的文本信息处理为结构化数据。在此基础上，整合机器学习、近邻检索、自动问答等技术，基于 Visual Studio 开发了建筑风环境灾害评估平台，通过损伤情况预测、相似案例推荐、相关知识问答三个功能模块，实现对于风灾环境下建筑状况的快速分析和评价。

关于本平台的进一步改善，主要考虑下述两方面。一是将平台嵌入 BIM 模型的二次开发中。这一做法有两点好处，一是方便进行案例数据的管理，既易形成结构化数据，也可进行全生命周期内的数据更新，因而可以获得对案例更准确的特征描述；二是可以通过点击拾取或从数据库中直接导入建筑特征，无需用户输入。另一方面的改进主要在于知识图谱。其构建仅使用风灾致损建筑案例数据，而未引入其他相关知识，因此局限了后续可实现功能。可考虑进一步完善知识图谱，采用深度学习等技术构建自动问答系统，拓宽服务范围，提高处理能力。

## 参 考 文 献

- [1] Sandri P. An artificial neural network for wind-induced damage potential to nonengineered buildings[M].

Texas Tech University, 1996.

- [2] 张润, 王永滨. 机器学习及其算法和发展研究[J]. 中国传媒大学学报(自然科学版), 2016, 23(02): 10-18+24.
- [3] Qiu-Zhen L I, Bai X Q, Li-Xia L I, et al. Hierarchical navigable small world graph algorithms based on quantization coding[J]. Computer Engineering & Science, 2019.
- [4] Zettlemoyer L S, Collins M. Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars[J]. arXiv preprint arXiv:1207.1420, 2012.
- [5] 周博通, 孙承杰, 林磊, 刘秉权. 基于 LSTM 的大规模知识库自动问答[J]. 北京大学学报(自然科学版), 2018, 54(02): 286-292.
- [6] 史培军. 三论灾害研究的理论与实践[J]. 自然灾害学报, 2002(03): 1-9.
- [7] 叶观宝. CECS 269-2010, 灾损建(构)筑物处理技术规范[S]. 北京: 中国建筑工业出版社出版, 2012.
- [8] 徐增林, 盛泳潘, 贺丽荣, 王雅芳. 知识图谱技术综述[J]. 电子科技大学学报, 2016, 45(04): 589-606.
- [9] 刘峤, 李杨, 段宏, 刘瑶, 秦志光. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(03): 582-600.
- [10] 王培凤, 李莉. 基于 Aho-Corasick 算法的多模式匹配算法研究[J]. 计算机应用研究, 2011, 28(04): 1251-1253+1259.
- [11] Kim J Y, Yu E, Kim D Y, et al. Calibration of analytical models to assess wind-induced acceleration responses of tall buildings in serviceability level[J]. Engineering Structures, 2009, 31(9): 2086-2096.