

# 大数据时代下工程合同信息管理系统构建与索赔数据分析

梅 聪

民航机场建设集团华北有限公司 北京 101300

**摘要：**本文立足于大数据时代的技术背景，系统探讨了工程合同信息管理系统的构建路径与核心功能，并深入研究了如何利用大数据分析技术赋能工程索赔管理。文章首先剖析了传统合同管理在信息孤岛、流程滞后和风险被动等方面的痛点；其次，详细阐述了基于微服务架构的工程合同信息管理系统的设计原则、技术选型与功能模块，强调了其全生命周期管理能力；再次，聚焦于索赔这一核心环节，构建了包含数据采集、特征工程、模型构建与智能预警在内的索赔数据分析框架。研究表明，深度融合大数据技术的工程合同信息管理系统，不仅能显著提升管理效率与合规性，更能将索赔从“事后补救”转变为“事前预防”和“事中控制”的主动风险管理工具，为工程项目的成功交付提供坚实保障。

**关键词：**大数据；工程合同管理；信息系统；索赔分析

## 引言

工程合同是工程建设全过程的“法律基石”与“经济纽带”，其管理水平直接影响项目成本、进度、质量及效益。然而，当前我国工程合同管理普遍存在“重签订、轻履行”“重文本、轻数据”等问题，依赖纸质文档和人工台账，造成信息孤岛、协同困难，导致索赔时证据链不完整、数据支撑不足，易引发纠纷。据《2023年工程合同管理白皮书》显示，超60%项目涉及索赔，其中45%由工期延误引发，大型项目因此平均增加8%-12%的成本。与此同时，物联网设备、BIM数字孪生体及各类管理软件正持续生成涵盖人、机、料、法、环的海量实时与结构化数据。若能有效整合与挖掘这些数据，将推动合同管理向精细化、智能化转型。本文聚焦两大核心问题：一是如何构建面向大数据时代的高效工程合同信息管理系统；二是如何利用系统数据开展科学、精准的索赔分析，实现从被动应对到主动防控的转变。

## 1 传统工程合同管理的痛点与大数据赋能的价值

### 1.1 传统管理模式的主要痛点

(1) 信息孤岛严重：合同文本、往来函件、变更签证、进度报告、财务凭证等关键信息分散在不同部门、不同系统甚至不同物理载体上，缺乏统一的存储与关联，导致信息查询困难，难以形成完整的证据链。(2) 流程效率低下：合同审批、用印、归档等流程多依赖线下签字和纸质传递，周期长、易出错、难追溯。特别是在大型项目中，跨地域、多层级的审批更是耗时费力。(3) 风险预警滞后：风险识别主要依靠管理人员的经验

判断，缺乏基于数据的量化分析和实时监控。往往是问题已经发生并造成损失后，才启动应对措施，属于典型的“事后诸葛亮”。(4) 索赔依据薄弱：索赔的核心在于“有理、有据、有节”。但在传统模式下，承包商很难系统性地收集、整理和关联所有支持性数据（如天气记录、甲方指令、供应链中断证明等），导致索赔成功率不高，或即使成功也难以获得足额补偿。

### 1.2 大数据技术带来的赋能价值

大数据技术（Volume, Velocity, Variety, Veracity）为解决上述痛点提供了全新的思路和工具：(1) 打破信息壁垒：通过建立统一的数据中台，可以汇聚来自BIM、IoT、项目管理软件、财务系统等多源异构数据，实现合同全要素的数字化和结构化。(2) 实现智能风控：利用机器学习算法，可以对历史合同数据进行训练，自动识别高风险条款、异常履约行为和潜在索赔点，实现风险的早期预警和主动干预<sup>[1]</sup>。(3) 提升决策效率：通过对海量历史索赔案例的分析，可以提炼出不同事件（如设计变更、不可抗力）下的标准损失计算模型和责任判定规则，为当前索赔事件的处理提供数据驱动的决策支持。(4) 优化合同范本：基于对大量已执行合同的绩效分析，可以反向优化合同模板，剔除模糊不清、易引发争议的条款，嵌入更公平、更具操作性的权责约定。

## 2 大数据驱动的工程合同信息管理系统构建

### 2.1 系统设计原则

(1) 全生命周期管理：系统应覆盖合同从“意向-谈判-起草-审批-签署-履行-变更-结算-归档”的全过程，确

保每个环节都有迹可循。每个合同应拥有唯一的ID，作为贯穿所有关联系统的数据主键。(2)数据驱动：将数据作为核心资产，建立完善的数据治理体系，确保数据的真实性、完整性、一致性和及时性。所有关键业务动作(如审批、付款、变更)都必须在系统中留痕。(3)开放集成：采用标准化接口(API)，能够与BIM平台(如Autodesk BIM 360)、智慧工地系统(如广联达、明源云)、财务软件(如用友、金蝶)、OA系统等无缝对接，打破信息孤岛。(4)用户友好：界面设计简洁直观，操作流程符合用户习惯，降低使用门槛。例如，为一线项目经理提供移动端APP，支持现场拍照、语音录入事件详情并即时上传。

## 2.2 系统技术架构

推荐采用分层的微服务架构，以保证系统的灵活性、可扩展性和高可用性。

### 2.2.1 基础设施层

基于云计算平台(如阿里云、腾讯云)，提供弹性计算、存储和网络资源。采用容器化技术(如Docker, Kubernetes)部署应用，实现快速扩缩容。

### 2.2.2 数据层

一是关系型数据库(如MySQL, PostgreSQL)：用于存储高度结构化的数据，如合同基本信息(编号、名称、甲乙双方、金额、工期)、付款计划(期次、金额、日期、状态)、审批流配置等。二是非关系型数据库/数据湖(如MongoDB, Hadoop HDFS)：用于存储半结构化和非结构化数据，如PDF格式的合同全文、扫描的签证单、现场照片/视频、邮件原文等<sup>[2]</sup>。三是搜索引擎(如Elasticsearch)：为海量文档提供毫秒级的全文检索能力，方便用户快速查找相关合同或条款。

### 2.2.3 服务层(微服务)

将系统功能拆分为独立的服务，各服务通过RESTful API或消息队列(如Kafka)进行通信。合同服务负责合同的增删改查、版本管理； workflow服务提供可视化的BPMN流程引擎，处理复杂的审批逻辑；履约监控服务实时监控合同履行情况，与外部系统对接获取数据；索赔管理服务专门处理索赔事件的登记、跟踪和评估；数据分析服务提供数据统计、报表和高级分析(如机器学习模型预测)。

### 2.2.4 应用层

提供Web端(供管理人员使用)、移动端(供现场人员使用)等多种访问入口。

### 2.2.5 智能分析层

集成大数据处理引擎(如Apache Spark)进行离线批

处理，以及流处理引擎(如Flink)进行实时分析。机器学习平台(如TensorFlow Serving)用于部署和管理训练好的模型。

## 2.3 核心功能模块详解

### 2.3.1 合同全生命周期管理模块

(1)智能起草：内置由法务专家审核的合同模板库(如EPC总承包、施工总承包、专业分包等)。系统支持基于自然语言处理(NLP)的条款智能审查，例如，当用户输入“乙方应在合理时间内完成修复”，系统会高亮提示“‘合理时间’为模糊表述，建议明确为‘X个工作日内’”，并链接到公司内部的《标准措辞手册》<sup>[3]</sup>。

(2)在线审批：支持自定义审批流程。例如，对于500万以下的合同，只需项目经理和分公司经理两级审批；对于500万以上的合同，则需增加公司法务部和分管副总的审批节点。系统全程记录审批意见、时间和操作人，并支持电子签名(如e签宝)。(3)履行监控：系统自动抓取项目管理平台中的里程碑完成日期，并与合同约定的工期进行比对。一旦发现偏差超过预设阈值(如±3天)，系统会自动向相关责任人发送预警邮件或APP推送。

### 2.3.2 索赔管理专项模块

(1)索赔事件登记：提供标准化表单，强制要求用户选择索赔类型(工期/费用/两者兼有)、根本原因(业主方/第三方/不可抗力/自身原因)、关联的合同条款编号。系统会根据选择的原因，动态加载所需的证据清单。(2)证据链管理：支持上传各类证据，并自动提取元数据(如照片的拍摄时间、地点)。系统允许用户手动或自动(通过OCR识别)将证据与事件时间线进行关联。所有证据文件均通过哈希算法生成唯一指纹，确保证据的完整性和不可篡改性。(3)损失计算辅助：内置多种行业认可的索赔计算模型。例如，对于总部管理费(Head Office Overhead, HOOH)的计算，系统提供Eichleay公式、HUD公式等多种选项，并引导用户输入所需参数(如合同总工期、延误天数、合同总金额、同期总部总管理费等)，自动输出计算结果。

### 2.3.3 大数据分析预警模块

(1)风险画像：为每个合作方(供应商、分包商)建立动态风险评分卡，维度包括：历史合同履约率、平均付款周期、诉讼记录、财务健康度(通过爬取公开财报数据)等。评分低于阈值的合作方，在新合同审批时会触发特别审查。(2)智能预警：基于预设规则和机器学习模型，对潜在风险进行实时监控。例如，规则引擎可以设定：“若项目所在地未来7天预报有连续3天以上

暴雨 (> 50mm/天), 且当前处于土方开挖阶段, 则触发‘不可抗力-工期延误’一级预警。”(3) 知识图谱: 利用图数据库 (如Neo4j) 构建“合同-条款-事件-案例-法规”的知识图谱。当用户处理一个新的“设计变更”索赔时, 系统可以自动推荐历史上相似度最高的10个已完结案例, 并展示其最终的判赔结果和关键证据。

### 3 基于系统数据的索赔数据分析框架

一个完整的索赔数据分析流程可分为四个阶段:

#### 3.1 数据采集与清洗

(1) 内部数据: 从合同系统抽取合同文本、付款记录; 从项目管理系统抽取进度计划 (.xml格式的Primavera P6文件)、日报、周报; 从财务系统抽取成本明细。

(2) 外部数据: 通过API接入国家气象科学数据中心获取逐小时降雨量数据; 接入上海钢联 (Mysteel) 获取螺纹钢HRB400E的日度价格指数; 接入企查查API获取合作方的司法风险信息<sup>[4]</sup>。(3) 清洗: 处理缺失值 (如用前后均值填充)、异常值 (如剔除明显录入错误的负成本)、格式标准化 (如统一日期格式为YYYY-MM-DD)。

#### 3.2 特征工程

这是决定模型性能的关键步骤。(1) 原始特征: 延误天数、受影响工序数量、延误期间平均气温、钢材价格波动率等。(2) 衍生特征: 计算“延误天数占总工期的比例”、“关键路径上受影响工序的权重和”等更具代表性的指标。(3) 文本特征: 对合同条款、往来函件进行分词、去停用词, 并通过TF-IDF或Word2Vec将其向量化。

#### 3.3 模型构建与训练

(1) 分类任务 (是否会索赔): 采用XGBoost或随机森林算法, 因其对特征重要性有良好的解释性。模型输入为上述特征, 输出为“是/否”。(2) 回归任务 (索赔金额): 采用梯度提升回归树 (GBRT) 或神经网络。模型输入为特征, 输出为连续的金数额值。(3) 模型评估: 使用准确率 (Accuracy)、精确率 (Precision)、

召回率 (Recall) 评估分类模型; 使用均方根误差 (RMSE)、平均绝对误差 (MAE) 评估回归模型。

#### 3.4 结果应用与反馈

将模型输出封装成API, 供前端应用调用。例如, 当项目经理登记一个事件后, 系统后台立即调用分类模型, 若预测索赔概率 > 80%, 则自动升级为高优先级事件, 并通知索赔专员介入。所有模型预测结果和最终的实际行动处理结果都会被记录下来, 用于下一轮的模型迭代。

### 4 结语

大数据时代为工程合同管理带来了前所未有的机遇。构建一个集成了大数据、云计算、人工智能等先进技术的工程合同信息管理系统, 不仅是提升管理效率的工具, 更是企业核心竞争力的体现。通过将索赔管理深度融入系统, 并利用数据分析技术对其进行赋能, 可以有效地将风险管理关口前移, 变被动应对为主动预防和精准控制。本文提出的系统构建框架和索赔数据分析方法, 特别是通过引入具体的技术栈 (如Spark, Neo4j)、数据源 (如Mysteel, 气象API)、计算模型 (如Eichleay公式), 为工程企业实现合同管理的数字化转型提供了极具操作性的路径。面对挑战, 唯有积极拥抱技术变革, 深化数据应用, 才能在激烈的市场竞争中立于不败之地, 确保每一个工程项目都能高质量、高效益地顺利完成。

### 参考文献

- [1] 郑忠钦. 基于大数据的合同管理系统V1.0. 重庆市, 家无忧(重庆)网络科技有限公司, 2022-04-19.
- [2] 李奎光. 浅谈大数据一体化下的合同全流程管理[J]. 财经界, 2020, (34): 88-89.
- [3] 李晓波, 刘勇. 建设管理中合同文档关键信息智能抽取技术研究及应用[J]. 长江信息通信, 2025, 38(02): 158-160+175.
- [4] 李晓波. 建设管理中合同关键信息智能提取技术研究及应用[D]. 三峡大学, 2024. DOI: 10.27270/d.cnki.gsxau. 2024.000726.