



同濟大學

TONGJI UNIVERSITY

硕士学位论文

(学术学位)

基于自然语言处理（NLP）的暖通规  
范语义解析方法研究

姓名：齐梓轩

学号：2130276

学院：机械与能源工程学院

学科门类：工学

一级学科：土木工程

二级学科：供热、供燃气、通风及空调工程

研究方向：建筑智能化

指导教师：许鹏教授

联合培养单位：

二〇二四年三月

---



同濟大學  
TONGJI UNIVERSITY

A thesis submitted to

Tongji University in conformity with the requirements for  
the degree of Master of Engineering

## **Advancing Fully Automated Compliance Checking with Large Language Models**

Candidate: Zixuan Qi

Student Number: 2130276

School/Department: School of Mechanical  
Engineering

Categories: Engineering

First-level Discipline/Degree: Civil Engineering

Second-level Discipline/Degree's Field: Heating,  
Gas Supply, Ventilation, and Air Conditioning Engineering

Research Fields: Building Automation

Supervisor: Prof. Peng Xu

Associate Supervisor:

Joint Training Institution:

---

基于自然语言处理(NLP)的暖通规范语义解析方法研究

齐梓轩

同济大学



## 摘要

传统的建筑 BIM 模型审查方法主要依赖于人工审核，不仅耗时长、效率低，而且还易受到个人经验和主观判断的影响，导致漏检和误判。因此，建筑自动合规性检查（Automated Compliance Checking, ACC）成为了解决这一问题的重要研究方向。然而，以往的建筑自动合规性检查流程过分依赖于基于规则的自然语言处理范式，导致条款解析方式过于复杂且条款覆盖面依然较窄。除此以外，还需要大量人力介入进行标注、规则撰写和维护工作，使得该系统的应用难以推广。

然而，近来以强大的上下文学习和推理能力而被广为讨论的大语言模型为解决此类问题提供了契机。本课题提出了一个全新的模块化可拓展应用框架，尝试将大语言模型引入全自动建筑合规性检查流程中。考虑到现有规范并非完全适合建筑自动合规性检查，本课题首先根据《标准化工作导则 第 1 部分：标准化文件的结构和起草规则》（GB/T 1.1-2020）和建筑规范特点，定义了四种条款类型；然后从《建筑设计防火规范》和《公共建筑节能设计标准》规范中经过文本预处理获得了 1097 例条款，并在标注后使用 BERT 模型微调训练得到条款分类模型；接着，基于此研究方向中数据集难以互通问题，提出了一种普适的条款语义结构并微调训练得到词性自动标注模型；同时，本课题对当下主流大语言模型在建筑自动合规性检查任务中的表现进行试验测试，通过具体的评价标准得出此任务下最佳模型和提示策略组合；最后，本课题针对 Revit Model Checker 插件开发了一套全自动 BIM 模型检查工具，并选取了三栋体量各异的 BIM 建筑模型对该系统进行了测试，验证了该方法的可行性。

结果表明，在良好的模型和提示策略组合下，大语言模型能够以非常可靠的解析结果取代此前基于规则进行条款解析的方式的同时，极大地减缓前序工作的错误累积问题。同时该系统可以在没有人工维护的情况下以高度鲁棒且成本低廉的方式完成建筑 BIM 模型检查，为建筑合规性检查在土建行业中的应用起到了很大的推进作用。

**关键词：**建筑自动合规性检查、大语言模型、提示工程、建筑信息模型

---

## ABSTRACT

Traditional methods of Building Information Modeling (BIM) model review rely heavily on manual inspection, which is not only time-consuming and inefficient but also prone to the influence of individual experience and subjective judgment, leading to omissions and misjudgments. Therefore, Automated Compliance Checking (ACC) has emerged as an important research direction to address this issue. However, previous processes of automated compliance checking in architecture overly relied on rule-based natural language processing paradigms, resulting in overly complex term parsing methods and narrow coverage of terms. Additionally, a large amount of human intervention is required for annotation, rule writing, and maintenance, making it difficult to widely apply such systems.

However, recent discussions have highlighted the opportunity presented by powerful large language models with strong contextual learning and reasoning capabilities to address such issues. This study proposes a novel modular and scalable application framework, attempting to integrate large language models into fully automated architectural compliance checking processes. Considering that existing standards are not entirely suitable for automated architectural compliance checking, this study first defines four types of clauses based on the "Guidelines for Standardization Work Part 1: Structure and Drafting Rules of Standards" (GB/T 1.1-2020) and architectural specification characteristics. Then, 1097 clauses were obtained from the "Building Design Fire Protection Specification" and "Public Building Energy Conservation Design Standards" specifications after text preprocessing, and a clause classification model was trained by fine-tuning the BERT model after annotation. Furthermore, considering the interoperability issue of datasets in this research direction, a universal clause semantic structure is proposed and a part-of-speech automatic annotation model is fine-tuned. Meanwhile, this study experimentally tests the performance of mainstream large language models in architectural compliance checking tasks, and determines the best model and prompt strategy combination for this task through specific evaluation criteria. Finally, this study develops a set of fully automatic BIM model checking tools for the Revit Model Checker plugin, and tests the system on three BIM architectural models of varying sizes to validate the feasibility of

this approach.

The results indicate that, with a good model and prompt strategy combination, large language models can reliably replace the previous rule-based term parsing methods while greatly alleviating the problem of error accumulation in pre-processing work. Additionally, the system can complete architectural BIM model checks in a highly robust and cost-effective manner without manual maintenance, thus significantly advancing the application of architectural compliance checks in the construction industry.

**Key Words:** Automated Compliance Checking, Large Language Models, Prompt Engineering, Building Information Model (BIM)



## 目录

摘要.....	I
ABSTRACT.....	II
第 1 章 绪论.....	1
1.1 研究背景和意义.....	1
1.2 研究现状.....	2
1.2.1 建筑自动合规性检查.....	2
1.2.2 建筑规范规则提取.....	3
1.2.3 大语言模型和提示工程.....	4
1.3 研究内容和技术路线.....	8
第 2 章 规范条款自动分类.....	13
2.1 概述.....	13
2.2 条款分类标准.....	13
2.3 规范预处理和条款分类数据集制作.....	15
2.3.1 规范预处理.....	16
2.3.2 条款分类数据集制作.....	17
2.4 基于深度学习的条款分类模型.....	19
2.4.1 模型训练流程.....	19
2.4.2 模型优化实验.....	22
2.5 本章小结.....	25
第 3 章 条款词性标注方法.....	27
3.1 概述.....	27
3.2 条款结构定义及数据集构建.....	27
3.2.1 语义结构定义.....	27
3.2.2 词性标签数据集制作.....	31
3.3 基于深度学习的词性标注模型.....	31
3.3.1 BIO 标签格式.....	31
3.3.2 模型训练和参数优化.....	32
3.4 本章小结.....	38
第 4 章 基于大语言模型提示工程的条款解析方法.....	39
4.1 概述.....	39

4.2 针对暖通规范解析的提示词构建.....	39
4.2.1 通用提示词构建.....	39
4.2.2 思维链构建.....	42
4.2.3 动态案例构建.....	45
4.3 大语言模型、提示词组合实验.....	47
4.4 条款解析结果和实验分析.....	51
4.4.1 解析结果评价指标.....	51
4.4.2 实验评价结果分析.....	53
4.5 本章小结.....	64
第 5 章 基于设计说明的模型参数强化方法.....	65
5.1 概述.....	65
5.2 设计说明预处理.....	65
5.3 提示词构建.....	66
5.4 设计说明解析结果.....	69
5.5 本章小结.....	71
第 6 章 自动合规性检查代码生成方法.....	73
6.1 概述.....	73
6.2 软件：Revit Model Checker.....	73
6.3 自动检查代码生成.....	76
6.4 本章小结.....	79
第 7 章 BIM 模型规范自动审查验证与分析.....	81
7.1 概述.....	81
7.2 规范解析结果.....	82
7.3 自动审查结果.....	86
7.4 本章小结.....	90
第 8 章 结论与展望.....	91
8.1 研究总结.....	91
8.2 未来展望.....	92
附录 A：条款解析提示组件（JSON）.....	95
附录 B：案例提示库（JSON）.....	101
附录 C：设计说明参数提取提示组件（JSON）.....	111
参考文献.....	113
致谢.....	119
个人简历、在读期间发表的学术成果.....	121





# 第 1 章 绪论

## 1.1 研究背景和意义

近年来，随着国家基础设施建设项目的持续推进和规模的不断扩大，各地的建设项目呈现出了蓬勃发展的趋势。与此同时，政府为了促进工程项目的快速实施，不断地精简和改革审批流程，显著缩短了审批时间。然而，这一措施在一定程度上加剧了施工图审查的时间压力，进而挑战了审图工作的严谨性和全面性，导致漏审现象时有发生。尽管建筑信息模型（Building Information Modeling, BIM）已成为行业的重要工具，其合规性检查仍大多依赖于传统的人工方式，比如三维浏览或内部漫游。而传统的手动检查流程不仅效率低下、成本高昂，还易于出错<sup>[1]</sup>。尤其对于大型 BIM 模型，漏检的风险几乎无法避免。此外，审图人员在审查过程中过分依赖于个人经验和对规范的熟悉度，这不可避免地引入了主观因素，降低了审查的客观性和准确性。即便经过大量时间和资源培养出的经验丰富的审图人员，也难以保证在工作中免于犯错。此外，即使 BIM 模型存在明显的不合规问题，审图人员也无法保证百分之百发现，因为人类不擅长高度集中和机械性地进行穷尽检查，这更多是机器的长处。因此，当下迫切需要探索一种方法，通过机器自动化来完成合规性检查或者部分复杂审查工作，以解决当前审图工作面临的挑战。

同时，建筑设计在其整个生命周期中发挥着至关重要的作用。设计阶段是确定建筑形态和功能的关键时期，审图则是确保设计方案符合相关法规和标准的重要环节。然而，设计与审图之间的关系并非总是顺畅，常见的问题包括审图合格率低，以及审图返修所需的时间成本和人力成本较高。这主要源于审图过程中所面临的时间与人力资源的双重压力，以及普遍缺乏专业的审图技能。若能开发出一种工具，使设计师能够在较短时间内独立完成模型检查并得到及时的反馈，将极大程度上提升整体设计效率。

审图自动化的发展可追溯至不同的历史阶段。随着 BIM 技术的兴起，建筑设计中的信息量显著增加，促使人们对审图效率的提升寄予了更大的期待。在现有技术方案的大背景下，审图自动化的方法主要包括硬编码、半自动以及完全自动化方法。其中，完全自动化方法基于白箱框架，但在具体实施中采用黑箱操作，如在词性标注和条款语义解析环节中使用引入机器学习方法。

尽管当前的全自动方法在理论技术上已经较为成熟，但在实际应用中仍存在

一些瓶颈。首先，即使是先进的条款解析方法，也无法避免手工逻辑编码操作的大量使用。这主要因为当下的自动条款解析步骤难以完全模拟人类的复杂思维过程，在审图中经常需要人工介入进行灵活判断。此外，依赖于条款语料词性标签训练的黑盒模型因缺乏统一的数据集和可扩展的规则，导致不同研究之间的标签设置各不相同，进而影响模型训练的效率 and 可迁移性。在语义解析部分，尽管简化标签旨在结构化表达，但众多元素组合的解析规则和迭代算法难以全面覆盖所有情况，这一条款解析环节往往构成自动合规性审查系统的短板。

最后，模型参数方面的挑战主要包括参数缺失或混乱、设计说明中省略的间接参数以及对审图可能产生重要影响的经验判断或超纲内容等问题。

综上所述，当前建筑自动合规性检查的主要框架在实际应用中存在一系列挑战，这要求进一步研究和开发更加智慧和高效的模型检查自动化方法，以更好地应对复杂多变的审图需求，从而提高建筑设计和审图工作的整体质量及效率。

## 1.2 研究现状

### 1.2.1 建筑自动合规性检查

建筑合规性检查利用系统化方法对设计进行分析，而无需更改设计本身。该方法依赖于评估设计模型的属性参数及其相互之间的联系<sup>[2]</sup>。在建筑的整个生命周期里，设计需满足众多法规、标准和要求的制约<sup>[3]</sup>。在目前的实践中，从业者进行合规性检查，目的是检测外观设计和现有记录中的空间违规行为，方法是手动浏览冗长的文本法规，根据他们的知识和经验对其进行主观解释，并根据它们检查海量和异构的效用数据。这种做法既不有效，也不可持续，这归因于公用事业规范文件的庞大规模和异质性，以及解释对人类知识和主观判断的严重依赖——不同的解释者甚至可能从同一条款中产生不同的含义<sup>[4]</sup>。

为应对增加的建筑许可需求，科研工作者与行业专家开始探索新的检查方法，旨在实现检查过程的无纸化、自动化和标准化<sup>[5]</sup>。自 1966 年 Fenves 首次尝试应用决策表来自动化建筑合规性检查以来<sup>[6]</sup>，80 年代见证了对此类系统的广泛探索，包括 SICAD 和 SPEX 等系统<sup>[7]</sup>。但这一时期由于技术限制，这些系统多停留在概念验证阶段<sup>[8][9]</sup>。然而，这些早期探索为后续发展的自动规范检查系统奠定了基础，如 21 世纪初的 Baider 和 DesignCheck 等<sup>[10]</sup>。此外，80 年代和 90 年代还见证了一些聚焦于特定领域规范的专家系统的出现<sup>[11]</sup>。

随着 2000 年以来建筑信息模型(BIM)的崛起，为土木工程领域的多项任务，如结构分析<sup>[12-14]</sup>、消防安全评估<sup>[15]</sup>、能源建模<sup>[16-18]</sup>、施工管理<sup>[19,20]</sup>以及自动化

[21,22]带来了新的方法。BIM 为合规性检查系统提供了革新的工作方式，通过提供可靠的数字建筑信息[23]，如 SMC 最初作为 BIM 验证工具，后来增加了合规性检查功能[2]。一些国家和地区政府，如新加坡和英国，开始要求将 BIM 模型而不是传统纸质图纸用于规划审查，以此推动 BIM 在行业中的应用[24]。

建筑合规性检查的过程通常包括规则解析、模型强化、规则执行和检查报告四个阶段。当前研究主要集中在“模型强化”与“条款解析”两大领域。“模型强化”旨在从建筑模型中提取执行检查逻辑所需的具体数值或属性。理想情况下，所有必要参数都能在模型中直接找到匹配。但现实中，由于模型可能不完整或标准化程度不一，需要对模型进行预处理，比如淘汰不必要的信息以及进行参数推理以补充缺失信息。“条款解析”则聚焦于将领域专家的知识转换为计算机能理解的标准逻辑，从而实现检查任务的自动化。尽管已有研究尝试探索不同的自动化或半自动化规则解释方法，但这些方法在自动化水平和应用范围上仍有局限。

## 1.2.2 建筑规范规则提取

随着 BIM 作为建筑的可靠数字表示，规范合规性检查系统在过去二十年中取得了巨大进步。然而，它们离实际的广泛部署还有很长的路要走。在许多当前的自动化规范合规系统中，规范信息提取和信息转换依赖于领域专家的手工作，将建筑规范转换为计算机可处理的格式。

一些早期的努力采用产生规则来表示建筑法规。产生规则通常采用“if <条件> then <操作>”的形式。Ma 等人[25]设计了一个数学矩阵表示方法，将土木工程师关于基于几何特征和成对关系对桥梁构件进行分类的知识封装起来。当模型缺乏对象分类信息时，数学矩阵表示方法通过数学计算实现语义丰富化。Fenves[6]提出了一种决策表方法，用产生规则简洁明确地表示规则。决策表指示适用于特定情况的条件以及基于条件值的适当操作。尽管表示了单个规则条款，但这种方法的主要缺陷是决策表中没有表示规则之间的关系[26]。类似地，Tan 等人[27]通过使用新的决策表来增强了这种表示方法的简洁性和表达能力，以表示信封设计规定。一组参数（例如，位置、类型）从规则中提取出来，作为决策表的子标题。还包括参考索引，链接到原始规则和交叉引用。然而，通常只有具有非常相似的条件和操作集的规则才能以这种方式表示。此外，这种方法仍然无法显示规则之间的逻辑关系。后来，最初的决策表得到了改进，有人提出了一个 SASE（标准分析、合成和评估）模型[28]以解决这个问题。这是一种四级表示方法，包括一个组织网络（建筑法规组织）、一个信息网络（规定之间的依赖关系），其中包括个体规定级别（决策表）和规定中引用的数据项。这也是规则引擎的独立表示，因

此非程序员可以轻松创建和修改。然而,该模型的主要问题在于对数据项的适用性和过于复杂的优先关系<sup>[29]</sup>。

一些研究采用基于逻辑的方法使用参数表来表示建筑法规。商业软件如 Solibri Model Checker (SMC)<sup>[30]</sup>就是使用这种方法的一个例子,其中解释规则由软件工程师编写成计算机代码。因此,规则嵌入到规则引擎中,没有单独表示。类似的方法也在多个学术研究中看到。例如, Lee<sup>[31]</sup>采用参数表来表示美国法院的流通规则。Soliman-Junior 等人<sup>[32]</sup>使用 SMC 检查英国医院规则。为了决定是否使用 SMC 检查特定规则,他们提出了一个框架来根据规则的性质(即定性、定量和模糊)和转化为逻辑规则的可能性对规则进行分类<sup>[33]</sup>。基于规则的 NLP 往往表现出更好的文本处理性能(在精度和召回率方面),但需要更多的人力。

尽管使用参数表可以简洁地表示非常复杂的规则,但这种方法在表达能力有限和维护困难方面存在问题,因为它们是硬编码的<sup>[34]</sup>。它们也因为对用户而言是“黑匣子”而受到批评,缺乏透明度可能导致可信度问题<sup>[35]</sup>。

基于 NLP 使用黑盒算法根据给定训练文本的文本特征来训练文本处理模型<sup>[36]</sup>。Zhang 和 El-Gohary 提出了一种自动规则转换方法,包括(1)信息提取,即识别相关句子中的单词和短语,并通过词性标注、编纂者等预先定义的信息标签;(2)信息转换,即通过基于正则表达式的映射规则和冲突解决规则将提取的信息实例转换为逻辑子句<sup>[37]</sup>。此外,通过将信息提取和转换集成到一个统一的系统中,他们实现了完全自动化的规则检查<sup>[1]</sup>。之后他们还提出了一种基于本体的信息提取方法,以支持完全自动化的建筑节能合规检查<sup>[38]</sup>。在他们的研究中,模式匹配方法通过本体、预处理和许多其他领域特定技术的增强,关注建筑能源领域,提高了准确性。Xu 和 Cai 提出了一种基于语义框架的信息提取方法,关注领域和词汇语义,以支持自动合规性检查,该方法的特点是词汇语义框架的丰富和与领域语义框架的映射。随后,采用基于本体和规则的 NLP 框架对地下公用设施文本规则进行自动化解释,对模式匹配规则进行编码进行信息提取,提取的信息元素通过本体映射到其语义对应关系,最后转化为道义逻辑(DL)子句<sup>[39]</sup>。

### 1.2.3 大语言模型和提示工程

#### 1) 语言模型

语言建模领域的研究经历了四个主要阶段的发展,如图 1,展示了该领域跨越数十年的进步与创新:

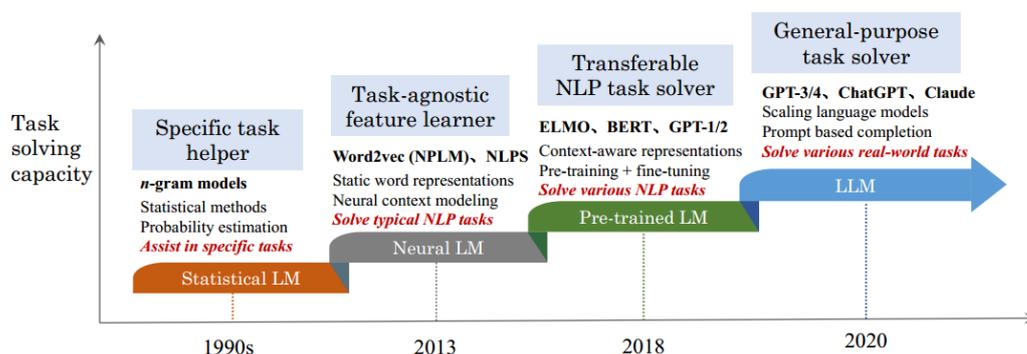


图 1 主流语言模型发展时间轴

a. 统计式语言模型 (Statistical Language Models, SLMs) : 在 20 世纪 90 年代, 基于统计学习方法兴起的 SLMs<sup>[40-42]</sup>开始发展。其核心理念是利用基于马尔可夫链假设的文字预测机制, 例如依据前文来预测接下来的单词。这种方法采用固定上下文长度  $n$  的  $n$ -gram 模型, 如 bigram 和 trigram 等。SLMs 已经在信息检索(IR)<sup>[43]</sup>和自然语言处理(NLP)<sup>[44-46]</sup>等领域找到了广泛的应用, 显著提升了这些任务的性能。然而, SLMs 面临的一个主要挑战是维度灾难: 由于需要估计庞大数量的转移概率, 高阶模型的精确估计变得异常困难。因此, 为了解决数据稀疏性问题, 引入了特别设计的平滑策略, 例如 backoff 和 Good-Turing 估计方法<sup>[47,48]</sup>。

b. 神经网络语言模型 (Neural Language Models, NLMs) : NLMs<sup>[49-51]</sup>利用神经网络架构来表现词序的概率, 采用了如多层感知机(MLP)和递归神经网络(RNN)等结构。特别值得注意的是, 文献<sup>[49]</sup>中的研究提出了分布式词表示的概念, 并创建了基于上下文特征聚合的词预测机制 (即分布式词向量)。NLMs 通过提出一种通用的神经网络方法, 延伸了对 NLP 任务的有效特征学习, 构建了端对端的统一解决方案<sup>[52]</sup>。此外, word2vec 技术<sup>[53,54]</sup>, 通过建立一个简化的浅层神经网络以学习分布式词表示, 证明了其在多种 NLP 任务中的有效性。这些研究标志着利用语言模型进行特征学习的开端 (超越了单一的词序列建模)。

c. 预训练语言模型 (Pre-trained Language Models, PLMs) : ELMO<sup>[55]</sup>作为一种早期尝试, 通过预先训练一个双向 LSTM 网络 (而非固定的词表示模型), 然后根据下游任务对网络进行微调, 来捕捉上下文相关的词表示。进一步地, BERT<sup>[56]</sup>基于可并行化的 Transformers 架构和自注意机制, 在大规模无标签文本库上预训练了双向语言模型及其特殊设计的预训练任务。这些先进的上下文敏感词表示有效地作为通用语义特征, 极大地推动了 NLP 任务性能的提升。由此引发的大量研究围绕“预训练和微调”的学习范式展开, 探索了 PLMs 的不同架构<sup>[57,58]</sup> (如

GPT-2<sup>[59]</sup>和 BART<sup>[57]</sup>) 和改进的预训练策略<sup>[60-62]</sup>。在这一范式下，通常需要对 PLMs 进行特定任务的微调。

d. 大语言模型 (Large Language Models, LLMs): 研究表明, 通过扩展 PLMs (例如, 增加模型大小或数据量), 通常能够提升下游任务的模型性能 (即遵循缩放法则<sup>[63]</sup>)。通过训练参数更多的 PaLMs (例如, 参数达到 175 billion 的 GPT-3 和 540 billion 的 PaLM), 研究者探索性能上的极限。尽管这种扩展主要在模型规模上进行 (保持相同的架构和预训练任务), 但这些大型 PLMs 与其较小的对应版本 (例如, 具有 330 million 参数的 BERT 和 1.5 billion 参数的 GPT-2) 展现出不同的行为特性, 在处理一系列复杂任务时表现出了令人惊讶的能力, 被称作应急能力<sup>[64]</sup>。例如, GPT-3 能够通过少量的上下文提示直接学习解决任务, 而 GPT-2 则在这方面表现不佳。因此, 这些大型 PLMs 被广泛称为“大型语言模型 (MLMs)”<sup>[65,66]</sup>引发越来越多的研究兴趣。一个显著的应用实例是 ChatGPT, 它使用 GPT 系列中的大型语言模型处理人机对话, 展现出了接近人类的交互能力。图 2 大语言模型发展时间轴 为近年来大语言模型的推出时间轴。

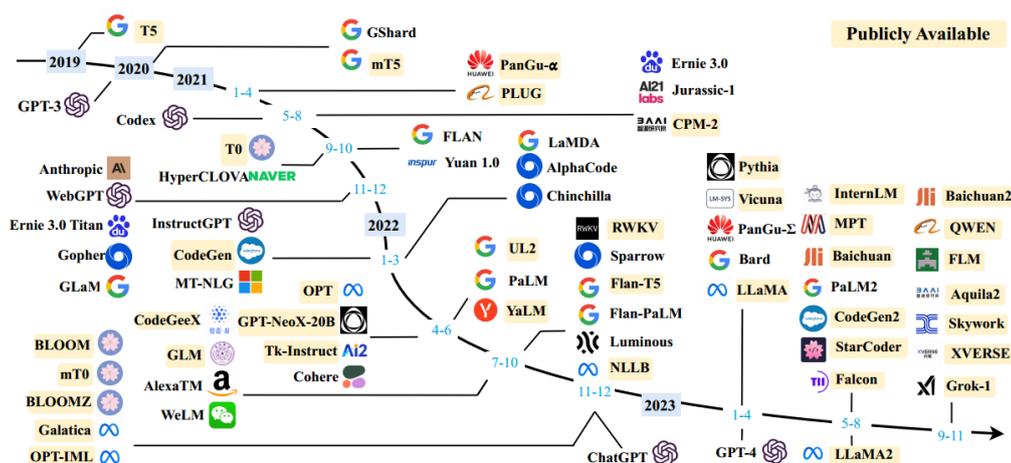


图 2 大语言模型发展时间轴

大语言模型与预训练模型之间的差异体现在以下三个核心方面。首先, 大语言模型展示了一种惊人的“涌现”能力, 涌现能力被通常定义为“在小型模型中不存在但在大型模型中产生的能力”。<sup>[64]</sup>这是区别 LLM 与先前 PLM 的最显著特征之一。这种涌现能力成为了语言模型在执行复杂任务时表现卓越的关键因素, 进而极大地增强了人工智能算法的力量和效率。其次, 大语言模型的涌现将会根本性地改变人们开发和应用人工智能算法的方式。区别于较小规模的预训练模型, 接入大型语言模型的主要途径是通过提示式接口, 如 GPT-4 API。因此,

用户必须掌握大语言模型的运作原理，并能够以大语言模型能理解的方式来构建并形式化他们的任务。第三，大语言模型的发展模糊了研究与工程之间的界限。训练大型语言模型要求研究人员在处理大型数据集和运用分布式并行训练技术方面拥有深厚的实操经验。为了构建功能强大的大语言模型，研究团队不得不面对极为复杂的工程挑战，这要求他们与工程师密切合作或直接承担工程师的角色。

## 2) 提示工程

大语言模型有着规模极大的参数量，需要更大量且内容广泛的训练数据，例如：GPT-3 有 1750 亿参数，在训练的过程中，批数量最大可达到 320 万 token，预训练规模使用 3000 万亿 token 的数据。这些资源对于个人用户来说通常是不可获得或难以承担的<sup>[67]</sup>。因此，使用大语言模型的方法——提示工程被学界提出。

在实际应用中，提示是模型的输入，其工程设计可能导致显著的输出差异<sup>[68]</sup>。修改结构（例如，改变长度、实例排列方式）和内容（例如，措辞、插图选择、指示）可能会对模型生成的输出产生显著影响<sup>[69]</sup>。如图 3，原输入为询问 GPT-3.5 在某一种情景下房间内的风口数量，该问题包含了分类讨论以及数学计算，不能够直接从原文中得到答案，需要模型有一定的推理能力。因此，在第一轮对话中模型回答出了错误答案（错误原因是没有考虑到回风口的数量）。在后一轮对话中，测试加入了进一步提示，该提示引导语言模型不仅仅给出答案，而是生成一系列解决问题的中间步骤，从而得到了正确的答案。

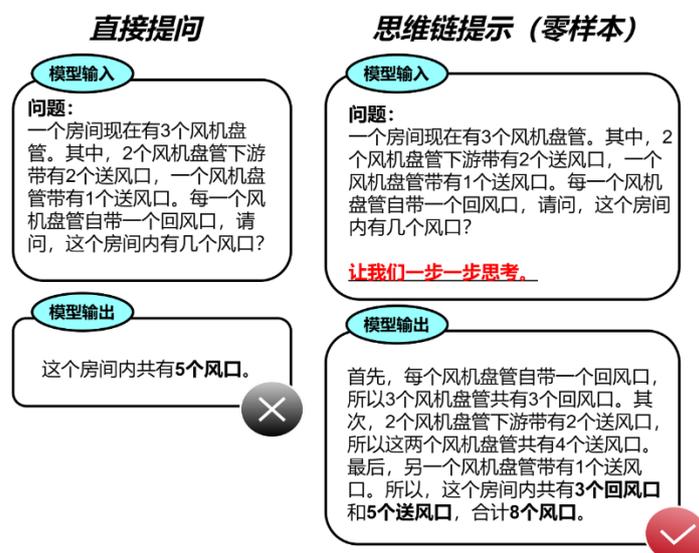


图 3 提示案例

研究表明，提示中包含的措辞和示例序列均被观察到对模型行为产生重大影响

响<sup>[69,70]</sup>。一种典型的提示方法是将任务描述和(或)示范(demonstration)以自然语言文本的形式表达的上下文学习(in-context learning, ICL)<sup>[71,72]</sup>。多项研究表明, ICL 的有效性在很大程度上受到示范设计的影响<sup>[69,73,74]</sup>, 不同的示范对于 ICL 的性能影响非常大。因此, 选择一个能够有效发挥大模型的 ICL 能力的样例子集很重要。关于示范选择的主要方法有两种, 被称作启发式方法和基于大语言模型的方法。

(1) 启发式的方法: 由于其简单性和低成本, 现有工作广泛采用启发式方法来选择示范。一些研究采用基于 k-NN 的检索器来选择与查询语义相关的样例<sup>[75,76]</sup>。然而, 他们只是针对每一个样例进行单独选择, 而不是对整个样例集合进行评估。为了解决这个问题, 基于多样性的选择策略被提出来选择对于特定任务最具代表性的样例集合<sup>[77,78]</sup>。此外文献<sup>[79]</sup>在选择样例时同时考虑了相关性和多样性。

(2) 基于大语言模型的方法: 另一部分工作利用大语言模型来选择示例。例如, LLM 可以直接根据添加样例后的性能提升<sup>[80]</sup>评估每个样例的信息量, 从而进行选择。此外 EPR<sup>[81]</sup>提出了一种两阶段检索方法, 首先使用无监督方法(例如 BM25)召回相似的样例, 然后使用密集检索器(使用 LLM 标记的正负样例训练)对它们进行排名。作为一种替代方法, 可以将示范选择任务建模为一个强化学习的问题, 其中 LLM 作为奖励函数, 为训练策略模型提供反馈<sup>[82]</sup>。因为 LLM 在文本标注方面表现良好, 最近的一些研究用 LLM 本身作为没有人工干预的示范生成器<sup>[83]</sup>

此外, 采用思维链提示(chain-of thought prompting)<sup>[75]</sup>可以通过将一系列中间推理步骤加入提示中来增强 ICL。思维链(Chain-of-Thought, CoT)<sup>[66]</sup>是一种改进的提示策略, 旨在提高 LLM 在复杂推理任务中的性能, 例如算术推理<sup>[84-86]</sup>, 常识推理<sup>[87,88]</sup>和符号推理<sup>[66]</sup>。不同于 ICI 中仅使用输入输出对来构造提示, CoT 可以导出最终输出的中间推理步骤纳入提示中。

### 1.3 研究内容和技术路线

针对当前 ACC 框架下实现完全自动化的规则解析方式依然需要大量人工操作现状, 以及 BIM 模型参数缺失导致自动检查范围严重受限的问题, 同时考虑到大语言模型正在引领全新 AI 时代的趋势, 本文拟采用大语言模型、深度学习、NLP 等技术, 提出一种全新的规则自动解析方法并构建基于大语言模型的建筑自动合规性检查的全流程可拓展应用框架。本研究主要的研究内容包括:

(1) 基于此研究方向中词性标注数据集难以互通问题, 提出了一种普适的条

款语义结构；

- (2) 将最先进的大语言模型应用于暖通领域的规范解析，建立并验证基于大语言模型的建筑自动合规性检查流程框架。
- (3) 对主流大语言模型在自动合规性检查任务中的表现进行试验测试，通过具体的评价标准得出此任务下的普适结论以及多个大语言模型+提示策略的推荐组合。
- (4) 基于 Revit Model Checker 软件开发出一套少人力工作的全自动建筑合规性检查系统。为设计师提供更为全面、可解释性更强的模型检查报表。

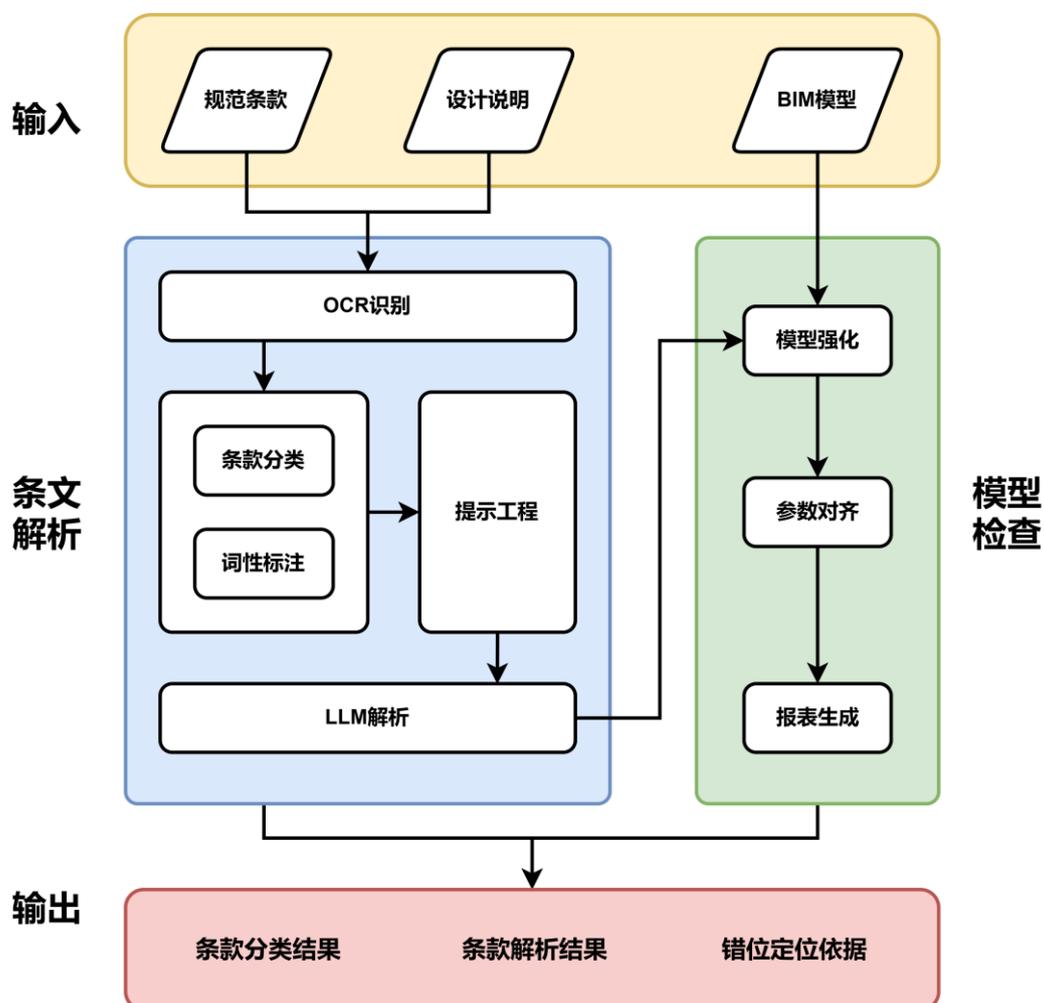


图 4 系统功能设计图

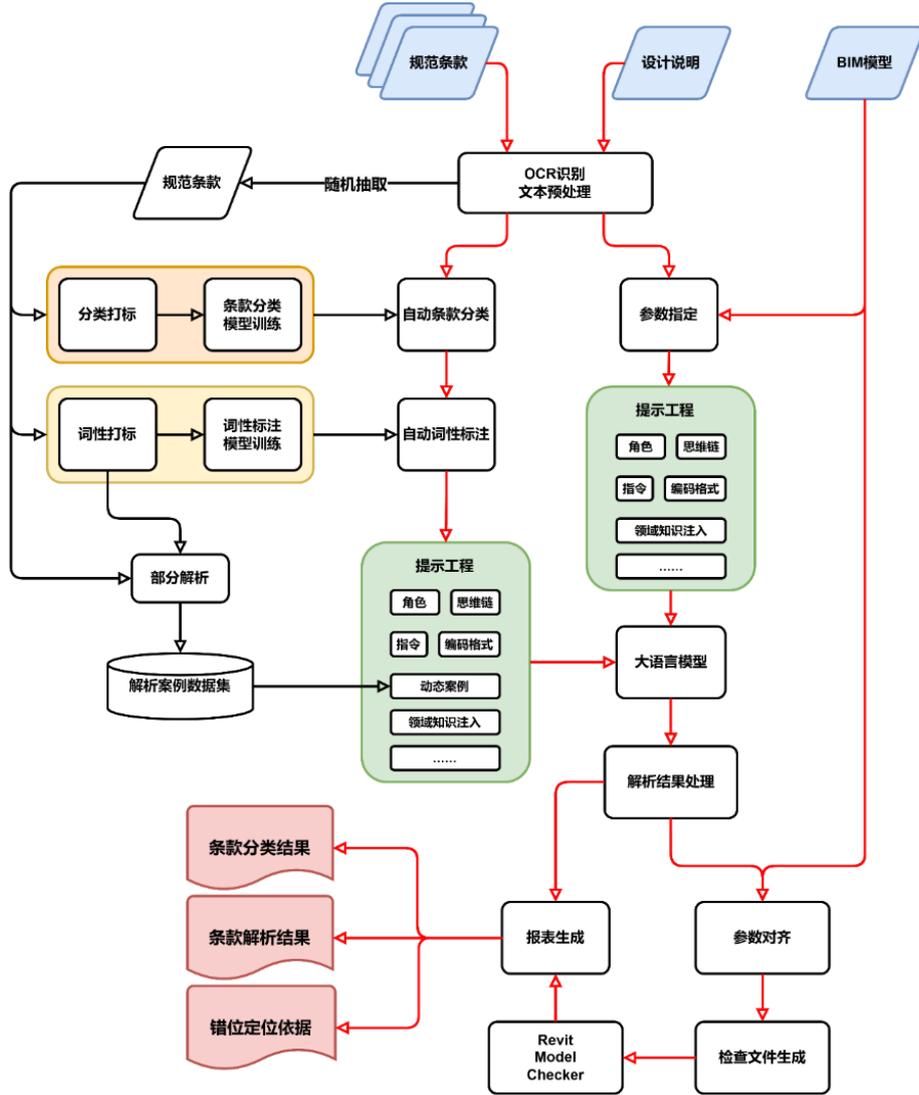


图 5 技术路线图

如图 4 系统功能图所示,该系统中分为条文解析和模型检查两大板块。对于设计师手中的规范通常是以 pdf 格式储存的自然语言条款文本。因此,在导入后需要通过光学字符识别(OCR)转化成文字,并进行一定程度的清洗后导入后续的流程;接下来,对于不同类型的条款通过训练好的条款分类模型进行自动分类,然后通过词性标注模型进行条款内不同实体的自动标注;之后根据条款文本和提示库中内容进行组合形成最终提示;输入到大语言模型实现建筑条款的自然语言本文转结构化文本。根据结构化表达中的语义元素与 Revit 模型中的对应参数和设计说明中提取的项目设计参数进行参数对齐——例如“防火门”和“平推门”分别都对应 Revit API 中对应变量名“OST\_Doors”(因为条款表达的多样性,所以在自动检查代码生成时进行语义元素和模型构件之间的命名对齐是十分必要的。)最后利用 Autodesk 官方推出的合规性检查插件 Revit Model Checker 运

行检查，生成最终报表。

如图 5 所示，该技术路线涉及到 3 类自然语言任务模型，其中包括条款分类模型和词性标注模型。这两个模型的主要目的是为条款解析提供明确的显式提示，以提高大语言模型的条款解析鲁棒性和准确性：

**(1) 条款分类模型：**属于“句子分类”任务，旨在识别规范文本中的条款，特别是那些包含显式或隐式比较的条款。这一任务源于自动合规性检查任务的最终目标，即对设计师所产生的模型和文本进行参数的定性和定量检查。因此，条款分类模型实质上是一种“过滤”和“比较”的行为。若存在无法进行比较的参数，则无法对其进行检查。这可能是由于参数无法获取，或者不存在“比较”的情况。在这种情况下，需要专家介入，或根据第二步的分词任务所提取出的程度副词来判断模型的合规性。即：本研究仅考虑存在进行比较的参数的条款进行模型检查自动化的研究。

**(2) 词性标注模型：**属于“句内分词”任务，其主要目的是为文本中的每个词汇赋予相应的词性标签。在这项任务中，特别关注的是提取“比较”(cmp)词性标签，作为判断条款强弱的依据。这个任务相对来说较为简单，传统的预训练模型（如 BERT）完全可以胜任。因此，将词性标注任务与分词任务一起交给大语言模型处理可能会增加其任务负担并分散其注意力，从而导致解析效果的下降。然而，词性标注模型的作用在于减轻大语言模型在解析过程中的负担，并提供额外的提示信息。

**(3) 大语言模型：**属于“句子解析”任务，其主要目的是将以自然语言形式呈现的条款转换成结构化、便于自动化生成代码的结构化（如表格、树形拓扑等）语句。在这个部分中，提示技巧和组成的意义在于为大语言模型提供知识、额外信息和清晰指令，以期望其能够在短时间内完成大量质量匹配人类的结构化解析任务。通过多大语言模型对比验证，我们可以评估模型在解析任务中的性能，并产出实验结果评分结果。跨模型自治性验证能够验证模型在不同条件下的稳定性和可靠性，从而产出可信的解析结果。在设计说明解析方面，本研究将自然语言写成的设计说明与模型参数相结合，以应对质量参差不齐的建筑系统模型。这种综合方法能够提高模型的适用性和准确性，使其能够更好地解析设计说明中的关键信息，为自动合规性检查提供支持。

通过“条款解析”流程，得到待检查条款的“结构化”表达后，则需要执行“模型检查”步骤。当前“模型检查”部分的方法主要包括三种主流途径。首先是采用 IFC 数据结构结合 MVDXML 的方式。IFC 是 Industry Foundation Classes 的缩写，是建筑信息模型（BIM）领域的一个国际标准，用于在建筑、土木工程和建筑管理等领域中描述和交换建筑信息模型数据。IFC 定义了一组数据模型，

用于描述建筑和基础设施的物理和逻辑特性，包括几何形状、构件属性、空间关系等。MVDXML 是 Model View Definition XML 的缩写，是一种 XML 格式的文件，用于描述特定的 IFC 模型视图定义（MVD）。MVD 是 IFC 模型的特定视图或子集，定义了特定用途或需求下所需的 IFC 实体和属性。MVDXML 文件指定了模型检查所需的准确的 IFC 实体、属性和关系，以及相应的规则和要求。虽然这一方法是开源的，但操作繁琐冗余，要求设计师对 IFC 数据结构有较为熟练的掌握。其次是采用 RDF 格式结合 SPARQL 查询语句的方法。RDF 是 Resource Description Framework 的缩写，是一种用于描述网络资源的语义模型。RDF 采用三元组（主语、谓词、宾语）来表示信息，其中主语和宾语是资源，谓词表示资源之间的关系。RDF 通常以图形结构表示数据，是语义网的基础之一。SPARQL 是 SPARQL Protocol and RDF Query Language 的缩写，是一种用于查询 RDF 数据的查询语言。SPARQL 允许用户以图形模式的形式查询 RDF 数据，支持模式匹配、聚合查询、路径查询等复杂查询操作。但这种方法实施过程则需要基于模型数据结构对所提出的本体框架进行数据映射，从头构建该规范和领域内的知识体系，如果仅仅针对模型检查这一操作，则整体流程显得过于臃肿。最后本研究将采用 Revit Model Checker 插件结合 BIM 模型的方式进行模型检查，虽然该方法要求用户持有正版 Revit 许可证，但介于当前设计院普遍会采购对应的 BIM 正版软件，且操作相对简便，易于没有编程经验的建筑设计师上手，并且该插件预置最终报表生成和即时的模型可视化功能，可以更加快速且直观地让设计师得到检查反馈，减少软件的切换，从而提高工作效率。

然而，Revit Model Checker 所提供的报表内容并不能完全满足设计师的检查需求。例如，在输入的多个条款中，存在着定义型条款而非属性型条款。其中定义型条款如“独栋建筑面积小于或等于 300 m<sup>2</sup> 的建筑，应为乙类公共建筑。”这类条款并不存在对模型的任何实际约束，所以并不能作为该模型是否满足条款约束的检查条款。因此，该软件的最终检查报表应对输入的条款进行分类统计，并进行标明。同样，在可执行的条款中，为便于在执行本规范条文时区别对待，对要求严格程度不同的用词例如条款的程度词。例如：“表示很严格，非这样做不可的：正面词采用“必须”，反面词采用“严禁”；表示严格，在正常情况下均应这样做的：正面词采用“应”，反面词采用“不应”或“不得”；表示允许稍有选择，在条件许可时首先应这样做的：正面词采用“宜”，反面词采用“不宜”；表示有选择，在一定条件下可以这样做的采用“可”等”<sup>[89]</sup>。此程度副词对于设计师尤其是最终审查时应当被加以重视。所以，在最终报表中应加以分类和强调，从而给予设计师更全面、可靠的检查和评估结果。最终在不同体量、结构的实际 BIM 建筑模型中进行自动合规性检查验证。

## 第 2 章 规范条款自动分类

### 2.1 概述

在建筑自动合规性检查中,并非所有的规范条文都适合用于模型检查。例如,《公共节能设计标准》<sup>[89]</sup>中的第 1.1 条指出:“为贯彻国家有关法律法规和方针政策,改善公共建筑的室内环境,提高能源利用效率,促进可再生能源的建筑应用,降低建筑能耗,制定本标准。”这类规范旨在引导设计师进行设计,而不涉及具体的检查对象。同样,还有一些条文涉及到具体操作步骤、公式等,这些内容难以从设计模型文件中提取,例如 4.2.1 条的第一款:“有可供利用的废热或工业余热的区域,热源宜采用废热或工业余热。当废热或工业余热的温度较高、经技术经济论证合理时,冷源宜采用吸收式冷水机组。”虽然该条款包括了被检查对象,但内容不够具体,甚至模糊,需要专家介入。这些条文并不适合用于自动合规性检查。因此,在进一步的条款解析前,本研究需要对规范中的条款进行自动分类,以识别适合用于 BIM 模型执行自动检查的条款。

句子分类任务作为自然语言处理的经典任务,在预训练阶段已达到较好的分类效果。与大语言模型不同,预训练模型体积较小,通常为几个 GB 以内,且训练成本低、推理计算时间短,非常适合本地部署。因此,本研究采用基于深度学习的预训练模型,并结合领域任务微调,训练出条款分类模型。该任务属于监督学习,需要研究者根据任务目标定义标签类型,并对条款内容进行少量手动标注。因此,本章首先介绍了本研究采用的两种文本规范数据集收集方法;其次对识别出的文字部分进行预处理,例如删除多余空格、换行符等;然后对现有的规范条款分类标准进行简化分类,使其更适应自动合规性检查的任务范畴,并制作对应数据集;最后根据分类结果进行基于深度学习的预训练文本分类模型训练,实现条款类型的自动识别。

本章的第 2.2 节将阐述暖通规范条款的分类标准,第 2.3 节将说明条文的预处理范式以及条款分类数据集的制作过程,第 2.4 节将对基于深度学习的条款分类模型的训练过程进行详细描述,并在第 2.5 节对条款分类的结果进行系统性的分析。

### 2.2 条款分类标准

根据《标准化工作导则 第 1 部分：标准化文件的结构和起草规则》（GB/T 1.1-2020）<sup>[90]</sup>建筑领域的规范条款通常被分为：

- ✧ 术语标准：界定特定领域或学科中使用的概念的指称及其定义的标准；
- ✧ 符号标准：界定特定领域或学科中使用的符号的表现形式及其含义或名称的标准；
- ✧ 分类标准：基于诸如来源、构成、性能或用途等相似特性对产品、过程或服务进行有规律的划分、排列或者确立分类体系的标准；
- ✧ 试验标准：在适合指定目的的精密度范围内和给定环境下，全面描述试验活动以及得出结论的方式的标准；
- ✧ 规范标准：为产品、过程或服务规定需要满足的要求并且描述用于判定该要求是否得到满足的证实方法的标准；
- ✧ 规程标准：为活动的过程规定明确的程序并且描述用于判定该程序是否得到履行的追溯证实方法的标准；
- ✧ 指南标准：以适当的背景知识提供某主题的普遍性、原则性、方向性的指导，或者同时给出相关建议或信息的标准。

在自动合规性检查中，审查的对象实际上是 BIM 模型文件中的数据。因此，在审查的最后一步，需要对某一类构件中的某一个属性参数与规范中的参数进行比对。术语、符号和分类内容通常位于具体构件之前，起到限定主语的作用。例如，在绝大部分标准中都包含的“术语”章节，以及《公共建筑节能设计标准》（GB50189-2015）中对于甲类、乙类寒冷地区公共建筑的描述，都属于这一类情况。规范标准在条文中通常会赋予具体的数值，这与自动合规性检查的执行流程相符合。然而，试验标准、规程标准、指南标准等部分通常描述了某一类设计过程，这些过程难以在最终设计的 BIM 模型中体现。

基于以上分析，本研究将规范分为以下四个部分：

#### ✧ 专家型条款

定义：这类条款包含了规范所涵盖的前缀指导性思想和操作知识性文本，或者包含了常规 BIM 模型中难以定义参数。通常涉及工程实践经验，需要专家进行人工介入。

举例：

- (1) 为贯彻国家有关法律法规和方针政策，改善公共建筑的室内环境，提高能源利用效率，促进可再生能源的建筑应用，降低建筑能耗，制定本标准。
- (2) 当利用通风可以排除室内的余热、余湿或其它污染物时，宜采用自然通风、机械通风或复合通风的通风方式。

#### ◇ 定义型条款

定义：该类型条文包含本规范内容中可能需要用到的专业名词知识性文本。

举例：

- (1) 建筑体形系数 **shape factor** 建筑物与室外空气直接接触的外表面积与其所包围的体积的比值，外表面积不包括地面和不供暖楼梯间内墙的面积。
- (1) 公共建筑分类应符合下列规定：
  - 1) 单栋建筑面积大于 300 m<sup>2</sup> 的建筑，或单栋建筑面积小于或等于 300 m<sup>2</sup> 但总建筑面积大于 1000 m<sup>2</sup> 的建筑群，应为甲类公共建筑；
  - 2) 单栋建筑面积小于或等于 300 m<sup>2</sup> 的建筑，应为乙类公共建筑。

#### ◇ 属性型条款

定义：该类型条文包含常规 BIM 模型中所定义的模型参数，比对运算方式直接方便。

举例：

- (1) 甲类公共建筑单一立面窗墙面积比小于 0.40 时，透光材料的可见光透射比不应小于 0.60；甲类公共建筑单一立面窗墙面积比大于等于 0.40 时，透光材料的可见光透射比不应小于 0.40。
- (2) 空气调节系统采用上送风气流组织形式时，宜加大夏季设计送风温差，并应符合下列规定：
  - 1) 送风高度小于或等于 5m 时，送风温差不宜小于 5℃；
  - 2) 送风高度大于 5m 时，送风温差不宜小于 10℃。

#### ◇ 间接型条款

定义：该这类条款包含了常规 BIM 模型中未定义的参数，但所需对比参数可以进行进一步运算或其他方式获取，通常需要进行模型数据增强。

举例：

- (1) 各城市的建筑热工设计分区应按表 3.1.2 确定。
- (2) 屋面、外墙和地下室的热桥部位的内表面温度不应低于室内空气露点温度。
- (3) 建筑空间高度大于等于 10m、且体积大于 10000m<sup>3</sup> 时，宜采用辐射供暖供冷或分层空气调节系统。

## 2.3 规范预处理和条款分类数据集制作

## 2.3.1 规范预处理

### 1) 光学字符识别（OCR）

设计师通常保有建筑规范的 PDF 版本，而现存的大部分规范均为不含有文本的影印版本。近年来，随着计算机视觉、自然语言理解、知识图谱等人工智能技术在 OCR 中的实际使用，OCR 的性能和效率都得到了很大提升。通过深度学习的自适应学习驱动方式，能够更好地应对传统 OCR 产生的一些问题，简化参数预处理的流程，实现端到端的处理，提高 OCR 识别率。目前，基于人工智能技术的 OCR 在简体印刷文字方面的识别率已达 98% 以上。

本节使用了 Github 上的第三方库 CnSTD 以及 CnOCR 完成基于视觉的 OCR 识别任务 (<https://github.com/breezedeus/CnSTD/tree/master>)。首先利用 CnSTD 完成场景识别，生成文字框并接入 CnOCR 完成框内文字识别，获取其内容并转化输出为纯文本.txt 格式。值得注意的是，目前尚未接入图片解析以及表格识别，这一点可以在未来进行模块化补充。。

### 2) 规范收集

通常情况下，设计师可以通过多个规范收集网站对建筑规范进行收集，这对于创建数据集来说较为有利，因为它们通常经过人工校对，可以规避图片转文字过程中因图片质量和文字排版而造成的识别错误。然而，网页通常会夹杂广告以及按钮中的文字，因此在收集之后需要进行一定程度的文字清洗工作，如图 6 所示。本节使用 Python 编写爬虫自动化脚本，在开放式规范收集网站“搜建筑” (<https://www.soujianzhu.cn/>) 上收集暖通相关的规范内容，并将其转化为纯文本格式。最后，通过正则表达式进行文本清洗工作，包括但不限于：

1. 删除文本中掺杂的<img><table>等非规范文本内容代码；
2. 删除网页中掺杂的广告、按钮等文本内容如：“搜索”、“联系我们”、“点击展开条文说明”等冗余符号；
3. 删除文本中多余的换行符、缩进符如：“\n”、“\t”；
4. 删除文本中的重复章节名称；
5. 分别将章、节、条前的空行统一设置为三行、两行、一行，并进行切割存储；
6. 为减轻训练计算成本，将对长文段进行切割，以“；”、“。”作为条文内的分割标志。

▼ 点击展开条文说明  
 1.0.7 本规范虽涉及面广，但也很难把各类建筑、设备的防火内容和性能要求、试验方法等全部包括其中，仅对普遍性的建筑防火问题和建筑的基本消防安全需求作了规定。设计采用的产品材料要符合国家有关产品和材料标准的规定，采取的防火技术和措施还要符合国家其他有工程技术标准的规定。

2术语、符号  
 2.1 术语  
 2.1.1 高层建筑 high-rise building 建筑高度大于27m的住宅建筑和建筑高度大于24m的非单层厂房、仓库和其他民用建筑。注：建筑高度的计算应符合本规范附录A的规定。  
 2.1.2 裙房 podium 在高层建筑主体投影范围外，与建筑主体相连且建筑高度不大于24m的附属建筑。  
 2.1.3 重要公共建筑 important public building 发生火灾可能造成重大人员伤亡、财产损失和严重社会影响的公共建筑。  
 2.1.4 商业服务网点 commercial facilities 设置在住宅建筑的首层或首层及二层，每个分隔单元建筑面积不大于300m<sup>2</sup>的商店、邮政所、储蓄所、理发店等小型营业性用房。  
 2.1.5 高架仓库 high rack storage 货架高度大于7m且采用机械化操作或自动化控制的货架仓库。

a 预处理前

b 预处理后

图 6 预处理结果预览

### 2.3.2 条款分类数据集制作

基于 2.2 节所阐述的分类标准，本课题将现有的规范分为“专家型”、“定义型”、“属性型”、“间接型”四类。首先将现有的《建筑设计防火规范》(GB50016-2014)、《公共建筑节能设计标准》(GB50189-2015)所收集到的 1097 例条款转为 JSON 格式导入到开源标注工具 doccano (<https://github.com/doccano/doccano>) 通过网页端进行人工标注。标注的过程以及输入输出格式定义如图 7-7 所示。并最终按照 6:2:2 进行分割创建训练集、测试集、验证集。

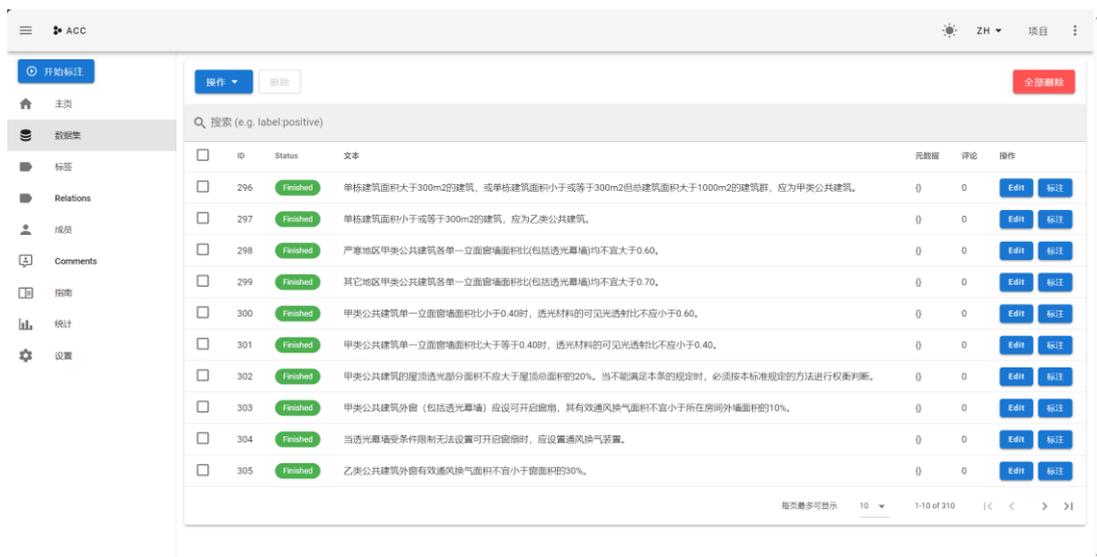


图 7 Doccano 标注页面

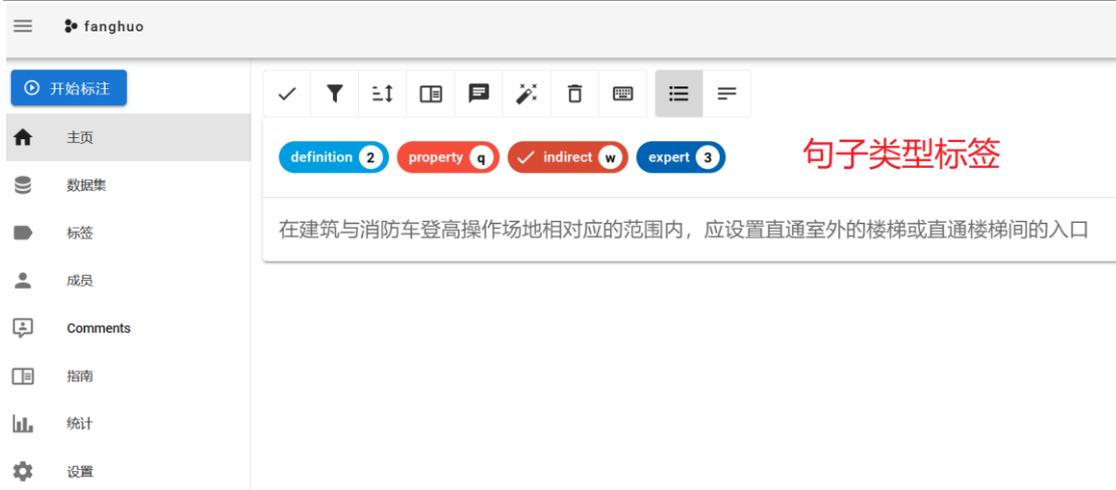


图 8 待标注条款输入条文、类型标签

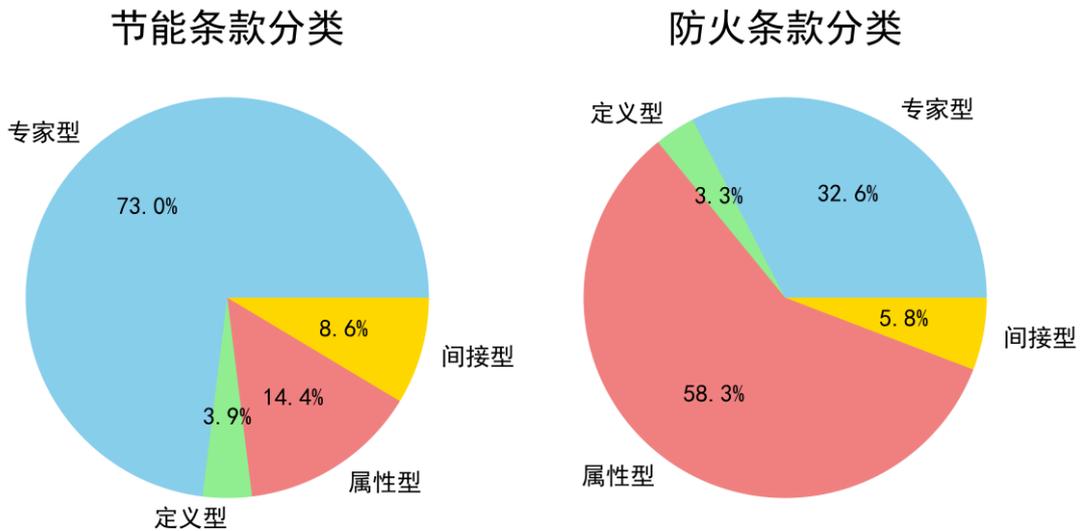


图 9 条款类型分布

如图 9 所示，在《建筑设计防火规范》（GB50016-2014）中，属性型和间接型条款的比例高达 64.1%，而相比之下，《公共建筑节能设计标准》（GB50189-2015）中这两类条款的比例不足 25%。这种结果反映了这两本规范的内容特点。防火规范中的约束条款通常定义更为明确简洁，对于构建的要求往往能具体到具体属性对应的数据要求，例如：“甲类厂房与人员密集场所的防火间距不应小于 50m，与明火或散发火花地点的防火间距不应小于 30m。”而在《公共建筑节能设计标准》（GB50189-2015）中，我们常常可以见到相对模糊的表达，例如：“建筑设计应充分利用天然采光。天然采光不能满足照明要求的场所，宜采用导光、反

光等装置将自然光引入室内。”这种类型的条文包含了规范的前缀指导性思想和操作知识性文本，或者包含了常规 BIM 模型中难以定义的参数，通常涉及工程实践经验，需要专家进行人工介入。这种特征在一定程度上反映了不同规范的信息密集度和可执行性程度。对于一个规范来说，其行文和标准越明确，可进行自动检查的部分比例就会更高。

## 2.4 基于深度学习的条款分类模型

### 2.4.1 模型训练流程

在大型语言模型出现之前，已经有大量先进的预训练语言模型，以 BERT 为代表，被用于句子分类。然而，在暖通乃至建筑领域，这一技术的应用始终未能被广泛推广。中文语料始终只占模型训练语料的一小部分，因此许多模型在英文的自然语言处理任务中表现良好，但在中文语料中表现不尽如人意。因此，出现了增强训练的预训练模型，如 BERT-chinese-base、bge-reranker-large、distilBERT-base-multilingual-cased 等模型。由于在中文语料少样本标注的迁移学习中，BERT-chinese-base 表现出色，被研究人员广泛接受并应用在相关研究中，因此本文选择 BERT-chinese-base 作为条款分类的基础模型。

BERT 是一种基于 Transformer 架构的大型语言模型，其核心创新在于采用双向训练方法。这意味着它在处理每个词时考虑上下文中的所有词。BERT 模型由多层 Transformer 编码器组成，每层包含两个主要部分：多头自注意力机制和前馈神经网络。自注意力机制使模型能够关注输入序列中不同位置的词，而前馈神经网络则对这些信息进行进一步处理。此外，BERT 在预训练阶段使用了两种任务：遮蔽语言模型（MLM）和下一句预测（NSP）。在 MLM 训练中，BERT 随机遮蔽输入句子中的某些词，并预测这些遮蔽词。而 NSP 任务则是学习判断一对句子是否在自然语境中连续出现。这有助于理解句子间的关系，对于自然语言处理任务至关重要。

在条款分类任务中，BERT 通过微调来适应不同场景下的特定任务。这意味着在预训练的基础上，BERT 使用特定任务的数据集进行额外训练。在微调过程中，模型的大部分权重保持不变，只有最后几层进行调整以适应特定任务的需求。如图 10，左图为预训练过程，决定了基础模型各单位的参数。而在右图的微调阶段，针对不同的任务，如多体自然语言推理（Multi-Genre Natural Language Inference, MNLI）：给定一对句子，目标是预测第二个句子相对于第一个句子是

包含，矛盾还是中立；命名实体识别（Name Entity Recognition, NER）：所谓实体识别就是你将想要获取到的实体类型，从一句话里面挑出来的过程，例如“同济大学”在一句话中就是一个实体，但在输入端，这四个字是被切分开的；斯坦福问答数据集（Stanford Question Answering Dataset, SQuAD）是 10 万个问题/答案对的集合。给定一个问题以及维基百科中包含答案的段落，任务是预测段落中的答案文本范围，即右图中训练的输出（Start, End Span）。图 10 为本章所涉及的句子分类任务的输入和输出，输入包括一个完整的句子，[CLS]标志着一个句子的开始，Tok 表示 token，是指对文本进行分词或分解后得到的基本单元，通常表示文本中的一个单词或一个子词；输出则为整个句子所属的分类标签（Class Label）。

这种结构和训练方法使 BERT 能够有效理解和处理语言数据，并在多种自然语言处理任务上展现出卓越性能。

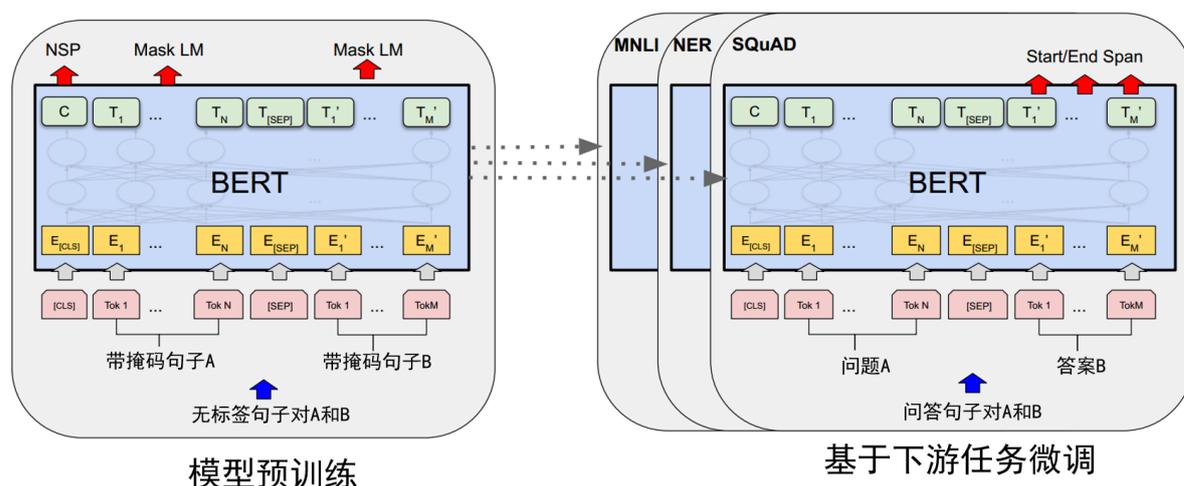


图 10 BERT 模型训练原理

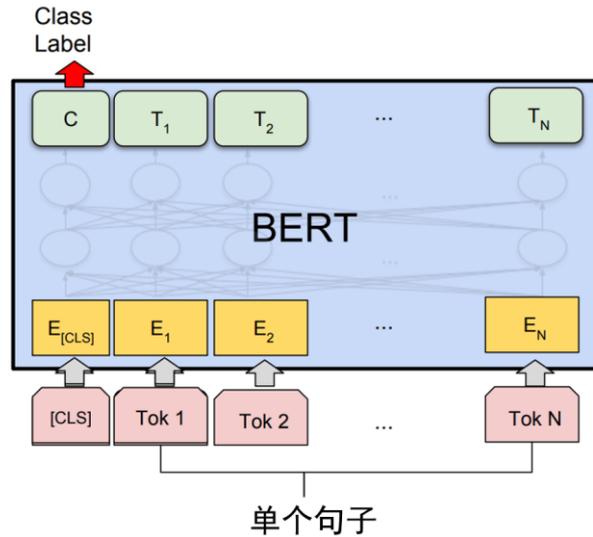


图 11 句子分类微调模型输入&amp;输出

因此，本文根据 2.3 节的数据集分类结果在 BERT 模型的顶层添加一个分类全连接层。通过在机器学习中特别是在多类分类问题中常用的 Softmax 函数完成分类计算。它将一个含任意实数的  $K$  维向量  $z$  转换成另一个  $K$  维实向量  $\sigma(z)$ ，其中每一个元素的范围都在  $(0,1)$  之间，并且所有元素的和为 1。该函数的形式通常用于将分类模型的原始输出，称为 logits，转换为概率分布。对于向量  $z$  中的每个元素  $z_i$ ，softmax 函数的输出通过以下公式计算：

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (2.1)$$

式中：

$e$ ：自然对数  $e$ ；

$z_j$ ：分类  $z$  中的第  $i$  个元素

分类模型训练采用相关任务常用的交叉熵损失作为 loss 函数，用于分类任务中衡量模型预测概率分布与真实标签分布之间的差异。公式中的变量代表以下含义：

$$L = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(p_{ic}) \quad (2.2)$$

式中：

$L$ ：损失函数的平均值；

$N$ : 条款总数, 在这里是  $x$ ;

$L_i$ : 第  $i$  个样本的损失;

$M$ : 分类的类别数, 在这里是 4;

$y_{ic}$ : 一个指示器变量, 如果样本  $y_i$  属于类别  $c$ , 则  $y_{ic}=1$ , 反之则  $y_{ic}=0$ ;

当一个 epoch 的数据训练完成后, 模型导入测试集以评估模型的性能。自然语言处理的分类任务中, 研究人员主要关注分类的准确度 (Accuracy)、召回率 (Recall)、F1 分数 (F1 score) 三大指标:

$$Precision = \frac{TP}{TP + FP} \quad (2.3)$$

$$Recall = \frac{TP}{TP + FN} \quad (2.4)$$

$$F1 \text{ Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.5)$$

如式 2.3, 精确度是正确预测为正例 (True Positives, TP) 的数量除以所有被预测为正例的数量, 后者包括正确预测的正例和错误预测为正例的案例 (False Positives, FP)。式 2.4 中, 召回率是正确预测的正例除以所有实际正例的数量, 后者是正确预测的正例与错误预测为负例 (False Negatives, FN) 之和。F1 分数 (式 2.5) 是精确度和召回率的调和平均数。其中, 精确率是正确预测的正例 (True Positives, TP) 除以所有预测为正例的数量 (TP 和 False Positives, FP 之和)。F1 分数在精确率和召回率之间取得平衡, 在分类标签分布不均衡时非常有用。

## 2.4.2 模型优化实验

本研究使用 python Transformer 库中所提供的 BERT-base-chinese 作为深度学习模型 (<https://huggingface.co/BERT-base-chinese>), BERT-bas-chinese 模型包含 12 个 Transformer 层, 每层有 12 个注意力头, 隐藏层的大小是 768, 约为 110M (百万) 个参数。模型可以处理最长为 512 个 token 的序列。由于中文是一个字符级别的语言, BERT 对中文进行了字级别的处理。在微调 BERT-base-chinese 模型时, 本研究需要设定以下实验参数:

- (1) 标签数量: 4, 根据 2.2 节的分类标准, 将输入的专家型、定义型、属性型、间接型分别生成  $1 \times 4$  的 one hot 向量;
- (2) 句子最大长度: 根据数据集条款长度统计, 句子最大长度设定在 125。此参数对于模型进行 embedding 的过程非常关键 (结果见图 13)。目前 BERT 支持最长 512 个 token。一般来说, 序列越长计算成本越高。本研

究根据现有训练集统计（如图 12），将此参数设定为 125，即：在训练过程中出现大于 125token 的条款时，该条款将会被截断；

- (3) **epoch**: 通常 BERT 的微调 epoch 数会设置在 30 以内。过多的 epoch 可能导致过拟合，尤其是在样本量相对较少的情况下；
- (4) **batch size**: 根据硬件资源的限制，常见的 batch size 设置为 16, 32 或 64；小批量通常有助于泛化，但太小可能会导致训练不稳定。本课题优化实验将设定为 2、4、8、16、32、64；
- (5) **学习率**: 对于 BERT 微调，学习率通常设置在到  $5e-5$  到  $3e-4$  之间。可以从较小的学习率开始尝试，如果模型收敛太慢，可以适当增加。本课题优化实验将初始学习率分别设定为  $1e-6$ 、 $3e-5$ 、 $5e-5$ 、 $7e-5$ 、 $9e-5$ 、 $1e-4$ 、 $1e-3$ ；

在模型训练过程中，本研究选择了 AdamW 作为优化器，这是对传统 Adam 优化器的一个改进版本，它采用了权重衰减策略以防止过拟合。本课题对模型的不同参数进行了分组以应用不同的权重衰减率：对于大多数参数，优化器应用 0.01 作为权重衰减率；而对于不应该进行权重衰减的参数，如偏置项和归一化层的参数，权重衰减率将设置为 0。这样的分组策略有助于训练更细致地控制正则化过程，以避免某些不应受到罚的参数受到不必要的约束。

本课题使用一张 NVIDIA GeForce GTX1080ti12G 显卡以及 Intel(R) Core(TM) i7-8700K CPU 3.70GHz 进行模型训练。

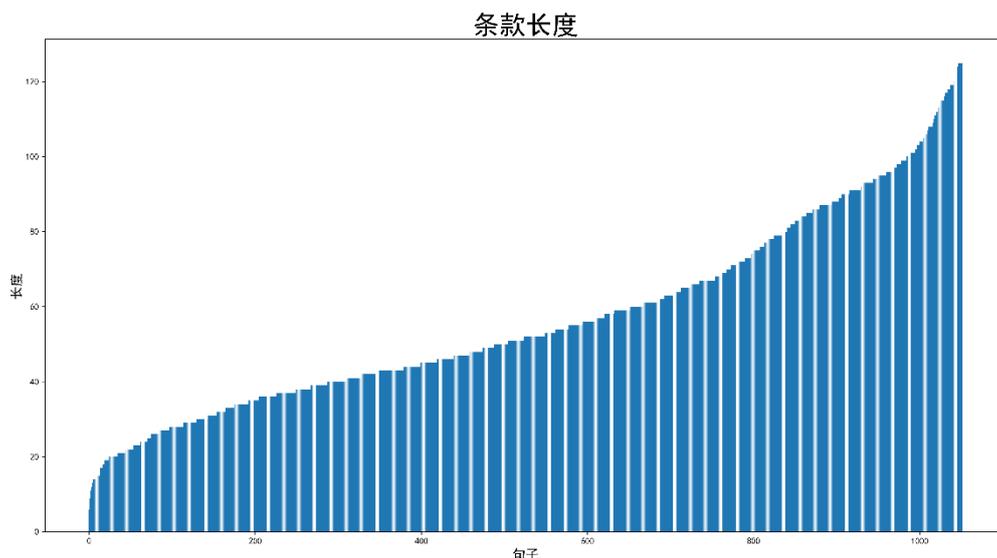
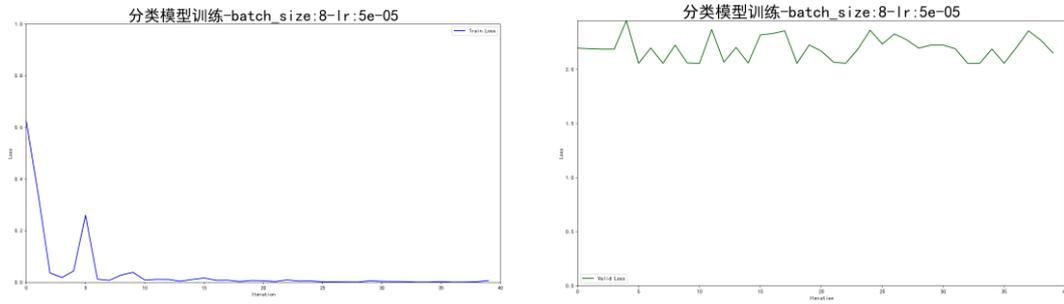


图 12 条款长度排序结果





a.分类模型训练曲线

b.分类模型验证曲线

图 14 最佳学习率下分类模型曲线 (batch size=8, lr=5e-5)

## 2.5 本章小结

不同建筑规范条款所规定的侧重点变化多样,并不是所有的规范都适合进行机械式的自动合规性检查。本章尝试把规范分类这一以往需要大量前序人工介入的工作纳入自动化流程,从而提高模型审查的效率和系统的实用性。首先本章基于专业知识将现有的规范类型分为四类(专家型、定义型、属性型和间接型),并考虑到条款文本格式使用使用 OCR 以及数据清洗手段将 pdf 格式的规范转化成.txt 条文;导入到 doccano 标注平台对条款数据集进行分类标签标注,并对《建筑设计防火规范》(GB50016-2014)和《公共建筑节能设计标准》(GB50189-2015)进行了标签统计和分析,指出了两类标准在内部条文风格上的不同;接着使用 BERT 模型对条款分类训练集进行训练,期间本研究针对模型超参基于以往研究的训练经验进行了批大小、学习率的优化实验,最后选择了批大小=8、学习率=5e-5、epoch=20 时得到的条款分类模型。



## 第3章 条款词性标注方法

### 3.1 概述

根据第二章的条款分类结果，本章将对属性型和间接型条款进行自动词性标注研究。通过利用 doccano 标注工具对条款进行少量词性标注，并结合 BERT 模型进行暖通条款的命名实体识别和词性标注，以为后续大模型的条款解析环节中提供更多思维链参考特征。首先，在第 3.2 节中，本章提出了一种简化的语义结构，以便更广泛地应用于整个建筑工程施工领域。这种语义结构能够有效缓解相关工作数据集之间的封闭性，并促进建筑工程施工行业数据集之间的交流与共享。接下来，在第 3.3 节中，详细描述了模型任务和模型设置，为后续的词性标注研究提供了基础和指导。最后，在第 3.4 节中，对词性标注的结果进行了深入分析和讨论。

### 3.2 条款结构定义及数据集构建

#### 3.2.1 语义结构定义

几乎此前所有采用深度学习模型的全自动建筑自动合规性检查系统研究均遵循了：1. 定义条款内部语义结构；2. 人工标注少量条款；3. 进行监督训练；4. 基于规则的条款解析的流程框架。第一步的必要性在于减轻第四步基于规则的条款解析的难度。例如，在普遍被业内认可的 Penn TreeBank 中，提出了 36 种用词性。如果使用 BIO 标注格式（见 3.3 节），那么标签数量则会上升到 108 个，这样庞大的标签种类会使得后续解析部分的工作难以进行；而如果标签种类过于稀少，例如只保留其中的主语、谓语、宾语。则在解析的过程中会因为标签过于笼统缺乏特征而无法正确地对条款进行解析。更重要的是，由于语言、专业等原因，行业内始终没有较为统一的可拓展的规范语义结构。这直接导致了四个问题：

- (1) 各类建筑自动合规性检查研究中经过标注的条款数据集无法进行互通；
- (2) 条款数据集之间没有互相认可的检验指标；
- (3) 元素定义过于关注条款内部，忽略了子句之间的关系（如：条件、引用等）。直接导致了即使是最先进的 ACC 系统在复杂句和长难句上的表现都不尽如人意；

(4) 任何一个研究者一旦决定添加某一个标签或尝试一个新的数据结构表达形式（如：图结构），往往需要从头进行耗时费力的人工打标工作。

以图 15<sup>[91]</sup>为例，该研究使用了 `sobj`、`obj`、`prop`、`cmp`、`ARprop`、`Rprop` 作为条款中的基础元素，形成树状结构 `Rctree`。

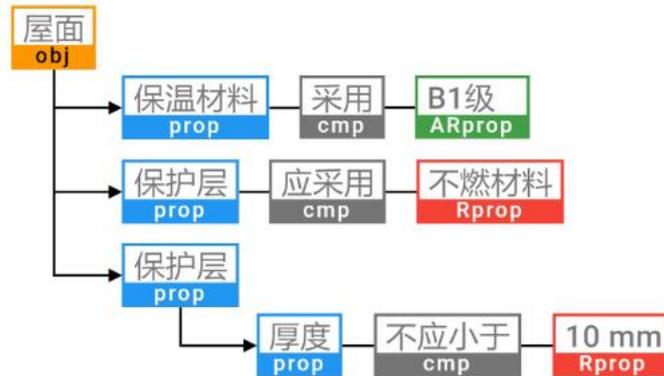


图 15 Rctree 示意图

可以看出，若是基于规则的条款解析方法，每一个元素都有存在的必要。但 这些元素大都是为了条款的内部解析服务，只有 `ARprop`（在该研究中表示条件从 句中的属性）是作为子句之间的关系标签，指引子句之间的关系。在大语言模型 中的“涌现”现象被观察到后，人们发现大语言模型强大的理解和推理泛化能力 可以在不加以微调训练以及人工标签的情况下很好地完成绝大部分常规 `NLP` 任 务。当前大语言模型在信息提取（`Information Extraction`）任务中体现出很大的潜 力，通过大语言模型解析条款相对于以往专注于标签之间的连接规则的方法所需 要的标签会更少甚至是不需要标签。或者说，在大语言模型进行解析的过程中， 句子中所出现的标签会更加偏向于句子内子句之间的联系，且所需标注的标签层 级将会更高、更易扩展、更易被业内认可统一，从而推进该研究方向数据集的流 通。

由于本研究中基于条款解析步骤所连接的后续操作是通过条件（`if-then`）语 句自动生成 `Revit` 自带插件 `Revit Model Checker` 的输入 `xml` 文件。所以，对于条 款解析步骤输出的要求应当是清晰地输出条款内：

- (1) 子句间的关系（或与非）
- (2) 子句内元素的包含关系、元素值

对于足够强大的大语言模型来说（如：`GPT4`），可以在完全没有标签提示的 情况下完成条款中的语义元素解析并输出足够可靠的结构化表达，如图 16 中的 右侧案例所示，`GPT4` 能够在没有标签作为提示的情况下正确地进行分词、理解

并有序地进行语义元素的连接；同样的条款对于同样以理解、推理能力而被人熟知的 GPT3.5 来说，结果却经常出现错误甚至是缺失的情况，如图 16 左侧的案例，GPT3.5 在分词任务中就出现了较为明显的错误，例如没有将“单一立面”和“窗墙面积比”分开，比较词和值也混在了一起，导致后续的元素连接的结果无法使用；除此以外，GPT3.5 错误地将主句中的“严寒地区”归纳成了条件从句中的内容。而经过标注的条款语句则通过提前进行分词并提供每个分词的显示类型帮助 GPT3.5 更好地完成了解析任务，即：标签可以通过向大语言模型提供额外知识或提示的方法帮助大语言模型解析出更完整、更正确的条款结构性表达。



图 16 通过分词标签来弥补模型推理能力差距案例

从上述的内容可以归纳出，在现有的大语言模型中，可以通过添加显式标签的形式来提高模型在条款解析任务中的表现。这样的帮助主要是通过预先分词和先验知识达成的，提前分词可以确保大语言模型更加专注与语义元素之间的先后排序以及逻辑连接问题，标签知识则为逻辑连接提供外界知识，从而确保逻辑链接正确性。由于条款解析任务可以被分为子句内部的逻辑解析以及子句之间的逻辑解析，而解析任务的主要实施者则为大语言模型。因此，本研究认为在后续的工作中可以将标签的设定基于两个解析环节进行设计，摒弃以往多而杂的标签设定，转向较为简约且更为上层且通用的标签设计，从而可以进一步解决该研究方向数据集的互通问题。

因此，在大语言模型解析条款的环境下，本研究提出了更为简化的 obj、cmp、ctg、qtg、value 五类标签，如表 2：

表 2 标签说明

标签	说明
obj	条款中出现的建筑元素，通常是一个建筑构件，元素之间的相互连接可能是并列也可能是从属游关系
cmp	比较谓词，表示比较的判断词，被比较的对象可以是数值也可以是 obj
value	具体数值，表示某个构建属性的值
qtg	属于 trigger 的一种，是一种表示引用的触发标签
ctg	属于 trigger 的一种，是一种表示条件的触发标签

其中：

1. **obj(object)**: 表示条款中出现的语义元素，通常是一个建筑构件，元素之间存在父级和属性关系。例如：“窗户的反射率”中，存在“窗户”和“反射率”两个语义元素，“反射率”是“窗户”的属性。元素之间的连接可能是并列也可能是上下游关系，比如：窗户[obj]的长[obj]和宽[obj]；也可能是存在包含关系，比如：窗户[obj]边框[obj]的厚度[obj]，这里的厚度和边框就是从属关系而不是并列关系。

2. **cmp(compare)**: 是比较谓词，例如：“应当满足”、“不应大于”、“不应小于”，表示比较和判断。在本研究中，一个待执行条款必须存在至少一个比较词。

3. **qtg(quoted trigger)**: 属于触发标签的一种，是一种表示引用的触发标签。比如：“应当满足[cmp]表 1-1 的要求[qtg]；应当满足[cmp]《某个规范》的要求[qtg]”。此类标签则偏向于帮助大语言模型进行条款中条件子句的辨别，即：`quoted_sentence` 的判别。

4. **ctg: conditinal trigger**, 属于触发标签的一种，是一种表示条件的触发标签。比如：“当 xxxx 时[ctg]，在 xxxx 条件下[ctg]”。此类标签则偏向于帮助大语言模型进行条款中条件子句的辨别，即：`conditional_sentence` 的判别。

5. **value**: 代表具体值，比如：“甲类[fobj]商业建筑[fobj]窗户[obj]的反射率[prop]应当大于[cmp]0.6[value]”。其中，value 的值包括了数值和单位。因为一般来说，单位（如果存在）总是紧跟着数值。所以，提取之后只需要根据语义元素

的数值类型进行进一步提取即可，无需再次将单位作为单独的标签，从而减少人工标注数量同时减少大语言模型解析难度。

### 3.2.2 词性标签数据集制作

基于上节所阐述的分类标准，本课题将条款内的语义元素分为“obj”、“cmp”、“value”、“ctg”、“qtg”五类。首先将 2.3 节中所收集到的 1097 例条款转为 JSON 格式导入到开源标注工具 doccano (<https://github.com/doccano/doccano>) 通过网页端进行人工打标形成“黄金标准”。打标的过程以及输入输出格式定义如图 17 所示。并最终按照 6:2:2 进行分割创建训练集、测试集、验证集。

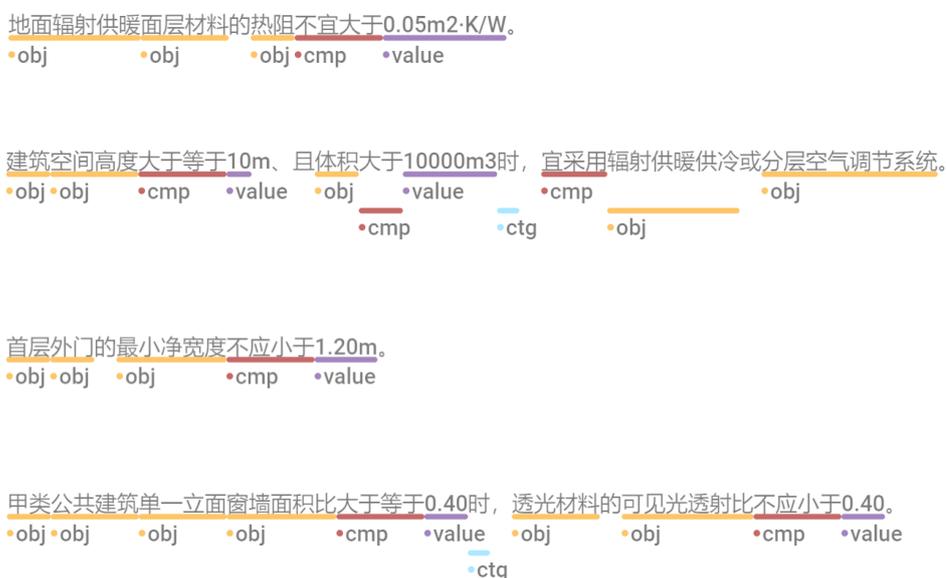


图 17 词性标注过程（部分）

## 3.3 基于深度学习的词性标注模型

### 3.3.1 BIO 标签格式

BIO (Beginning, Inside, Outside) 格式是一种常见的序列标注方法，尽管它最初是为了命名实体识别 (Named Entity Recognition, NER) 任务设计的，例如：

“我”“来”“自”“同”“济”“大”“学”这一句话中，“同济大学”就是其中具有特定意义的实体。NER 的目的是从文本中识别出具有特定意义的实体，如人名、地名、机构名、时间表达式等，并将这些实体归类到预定义的类别中。但也可以

用于其他序列标注任务，包括词性标注（Part-of-Speech, POS）。在 BIO 格式中，每个 token（通常是一个词或字）被标记为下列之一：

- **B (Beginning)**：表示一个实体或词性类型的开始。
- **I (Inside)**：表示属于同一个实体或词性类型，并且是该实体或类型的内部部分。
- **O (Outside)**：表示不属于当前考虑的任何实体或词性类型。

尽管词性标注通常不需要复杂的内外部分标记，因为大多数词性都是单个词而非词组或短语。但在某些情况下，BIO 格式可能被用于更复杂的 POS 标注任务，尤其是当词性标注涉及到词组或需要识别特定短语结构时。例如：建筑领域的复合名词（如天氟地水、天圆地方）、固定短语（如窗墙比）或专有名词（如 FCU）。等可能需要更细致的划分。BIO 格式提供了一种标记不仅单个词性，还标记词组边界的灵活方式。在处理复杂的语言结构，如嵌套或交叠的实体时，BIO 格式可以提供更多的信息。因此，相较于 3.2 节中提及的五类标签，在执行词性标注训练时，被扩展成了 11 类标签，如表 3 所示。

表 3 标注模型训练标签

标签	B	I	O
obj	B-obj	I-obj	
cmp	B- cmp	I- cmp	
value	B- value	I- value	O
qtg	B- qtg	I- qtg	
ctg	B- ctg	I- ctg	

### 3.3.2 模型训练和参数优化

在中文句子分类和词性标注两种方向的微调中，主要的微调区别在于任务的性质和标签的不同，因此需要针对不同的任务做出相应的调整。从任务性质来说，

句子分类任务是一个文本分类任务，旨在将中文句子分为不同的类别或标签。在句子分类任务中，整个句子作为一个输入序列被输入到 BERT 模型中，模型输出一个向量，然后通过一个分类器层（通常是一个全连接层）对句子进行分类，通常使用单一的标签编码方案（例如单标签分类或多标签分类），每个句子只对应一个或多个分类标签。词性标注任务是一个序列标注任务，旨在为句子中的每个词语标注其相应的词性。常用的损失函数包括交叉熵损失函数等，用于衡量模型输出和真实标签之间的差异。最后通常使用准确率（Accuracy）等指标来评估模型的性能。

在词性标注任务中，句子中的每个词语及其上下文被分别输入到 BERT 模型中，模型对每个词语的隐藏状态进行预测，并输出一个对应于每个词语的词性标签，通常使用序列标注的标签编码方案，将每个词语的词性标签作为一个序列，并与输入句子的词语一一对应。常用的损失函数包括序列标注任务专用的损失函数，如 CRF（Conditional Random Field）损失函数，用于衡量模型对整个句子序列标签的预测准确性。最后使用序列标注任务专用的评估指标，如准确率（Accuracy）、召回率（Recall）、F1 分数等，用于评估模型对整个句子序列标签的预测准确性，计算公式已在 2.4.1 节列出。

如

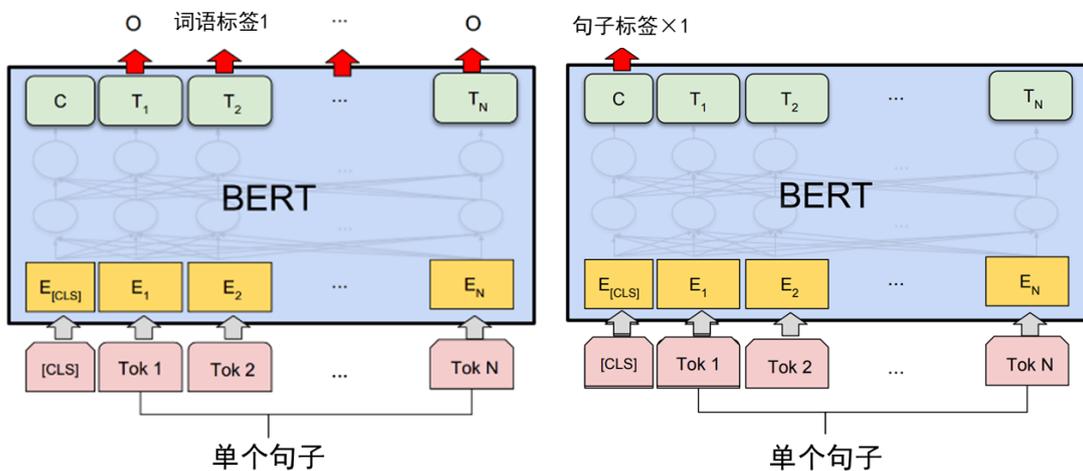


图 18，这两种任务的主要区别在于它们的输出结构和训练目标。词性标注是一种序列标注任务，侧重于理解文本中每个单词的语法角色，旨在为给定文本中的每个词单独进行 softmax 计算，分配一个正确的词性标签，损失是基于每个 token 预测计算并累加；而句子分类任务侧重于捕捉整个句子的总体意义或属性，将整个句子作为一个单位，对其进行分类，如情感分析、主题识别等，训练中的损失则是基于整个句子的单个预测。由于这些差异，模型的训练策略和细节需要根据具体任务进行调整，以确保模型能够准确地学习和预测。

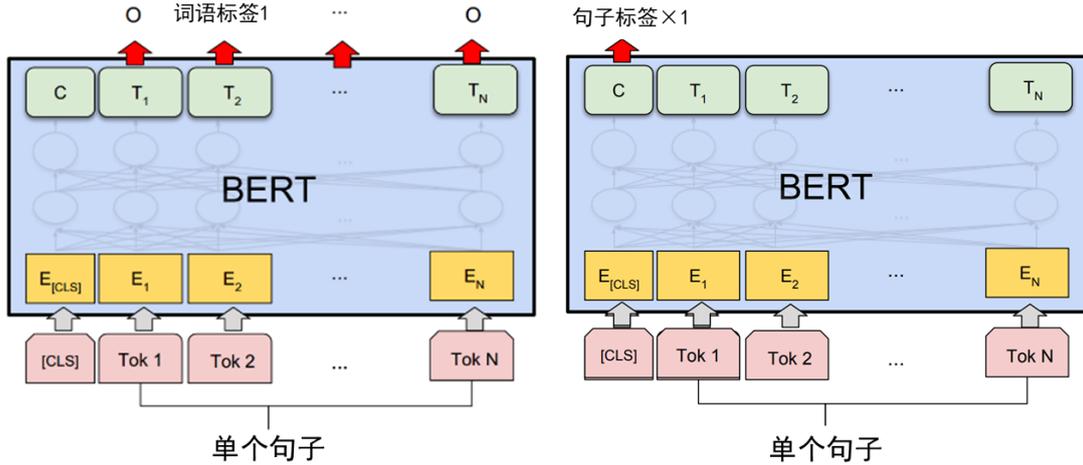


图 18 区别词性标注模型、句子分类模型

如 2.4 节所述，BERT-base-chinese 模型已经在大规模的中文语料库上进行了预训练，这意味着它已经学会了大量的中文语言知识和语境信息。通过在 BERT-base-chinese 上进行微调，可以利用迁移学习的优势。即使是对于那些在预训练过程中未见过的特定领域的的数据，模型也能快速适应新的任务，这对于词性标注尤其有用。由于 BERT-base-chinese 是一个在开源社区非常流行的模型，它有一个广泛的用户社区和大量的实现资源。这意味着在模型开发和调试过程中可以获得更多的支持和共享的知识，所以本研究使用 python Transformer 库中所提供的 BERT-base-chinese 作为深度学习模型，并在其基础上对暖通规范条款结合 3.2 节中所确定的十一类 BIO 格式标签进行微调训练。

在微调 BERT-base-chinese 模型时，本研究需要设定以下实验参数：

- (1) 标签数量：11，根据 3.3.1 节的分类标准，相较于 3.2 节中提及的五类标签，在执行词性标注训练时，被扩展成了 11 类标签；
- (2) 句子最大长度：根据数据集条款长度统计，句子最大长度设定在 125。此参数对于模型进行 embedding 的过程非常关键。目前 BERT 支持最长 512 个 token。一般来说，序列越长计算成本越高，在训练过程中出现大于 125token 的条款时，该条款将会被截断
- (3) epoch：通常 BERT 的微调 epoch 数会设置在 30 以内。过多的 epoch 可能导致过拟合，尤其是在样本量相对较少的情况下；
- (4) batch size：根据硬件资源的限制，常见的 batch size 设置为 16, 32 或 64；小批量通常有助于泛化，但太小可能会导致训练不稳定。本课题优化实验将设定为 2、4、8、16、32、64；
- (5) 学习率：对于 BERT 模型微调，开源社区通常将学习率设置在  $5e-5$  到  $3e-4$  之间。本研究可以从较小的学习率开始尝试，如果模型收敛太慢，

可以适当增加。因此，本课题优化实验将初始学习率分别设定为  $1e-6$ 、 $3e-5$ 、 $5e-5$ 、 $7e-5$ 、 $9e-5$ 、 $1e-4$ 、 $1e-3$ ；

在模型训练过程中，本研究选择了 AdamW 作为优化器，这是对传统 Adam 优化器的一个改进版本，它采用了权重衰减策略以防止过拟合。本课题对模型的不同参数进行了分组以应用不同的权重衰减率：对于大多数参数，优化器应用 0.01 作为权重衰减率；而对于不应该进行权重衰减的参数，如偏置项和归一化层的参数，权重衰减率将设置为 0。这样的分组策略有助于训练更细致地控制正则化过程，以避免某些不应受到罚的参数受到不必要的约束。

本课题使用一张 NVIDIA GeForce GTX1080ti12G 显卡以及 Intel(R) Core(TM) i7-8700K CPU 3.70GHz 进行模型训练。如表 4 所示，随着 batch 数量上升，训练时间逐渐减少，并在 batch size=16 之后下降变缓。通常情况下，较大的批量大小可以更有效地利用硬件资源，如 GPU 或 TPU，因为可以并行处理更多的样本。这意味着随着批量大小的增加，每个训练步骤所需的时间会减少，因为可以在同一时间处理更多的数据。因为所训练的模型在最终 F1 分数上相差并不大，所以出于训练时间考虑，本研究选择 batch size=16 作为进一步探究学习率的参数。

对于最佳学习率来说，随着 batch 数量上升，最佳学习率逐渐变高。因为较大的批量大小会导致梯度更新的方差较大，这可能需要更大的学习率来进行参数更新。因此，随着批量大小的增加，最佳学习率通常也会相应地增加。本研究最终选择平均验证 loss 最为稳定的  $lr=5e-5$  版本模型作为自动标注的最终模型。

表 4 不同 batch 下最佳表现模型

Batch size	最佳学习率	验证集 F1 分数	训练时间 (s)
2	$3e-5$	0.960	486
4	$3e-5$	0.959	318
8	$5e-5$	0.960	228
16	$5e-5$	0.962	192
32	$9e-5$	0.960	162

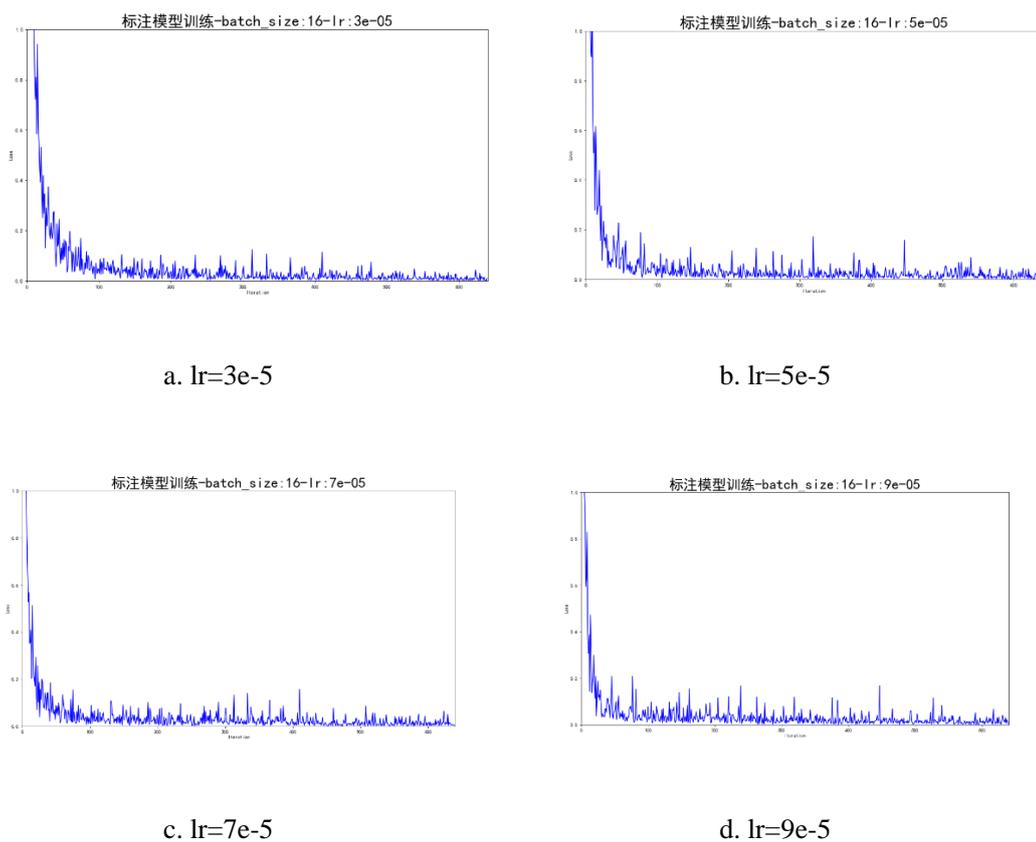
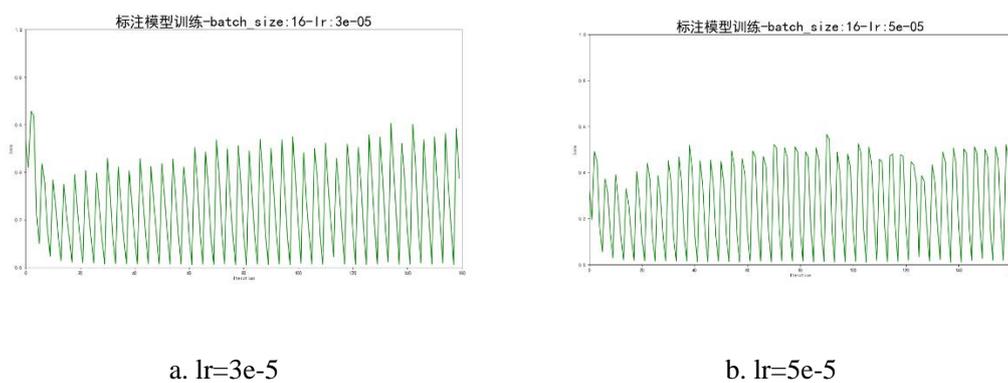


图 19 batch size=16 时，不同 lr 训练曲线



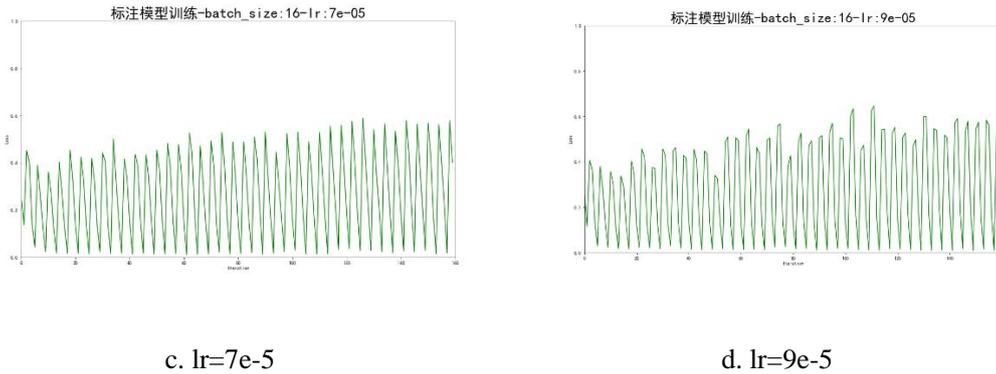


图 20 batch size=16 时，不同 lr 验证曲线

如表 5 所展示的数据，本研究按照既定的 11 个标签进行评估，平均而言，准确度、召回率和 F1 分数均达到了 0.962、0.963 和 0.962 的高标准，整体表现优异。在这些标签中，大多数的评分都超过了 0.9 的高分阈值，唯有 **B-qtg** 和 **I-qtg** 两个标签的表现较为不尽人意，其准确度分别仅有 0.667 和 0.545。这一现象可以追溯至“qtg”标签的验证样本相对于其他标签明显较少的原因。特别是在处理引用类的条款时，通常会省略引用的确切词汇。例如，条款“在名义工况和规定条件下，锅炉的热效率不应低于表 4.2.5 的数值。”中的“表 4.2.5”即被标记为引用触发标签（qtg）。然而，由于其包含的数字信息（如 4.2.5）容易给模型的训练带来挑战；同理，在“乙类公共建筑外窗有效通风换气面积不宜小于窗面积的 30%。”中，“窗面积的 30%”也作为一种隐式引用，被标记为引用触发标签。这种标注方式同样增加了对该类标签训练难度的提高。此外，相较于条件标签，引用标签在筛选的条款数据集中的出现频率较低，这也进一步影响了模型对这些特定标签的预测性能。

表 5 最佳模型标签预测表现

标签	准确度	召回率	F1 分数	验证数量
B-obj	0.949	0.987	0.967	375
I-obj	0.975	0.993	0.984	1689
B-cmp	0.976	0.960	0.968	126
I-cmp	0.985	0.995	0.990	202

B-qtg	0.667	0.167	0.267	12
I-qtg	0.545	0.429	0.480	28
B-ctg	0.916	0.970	0.942	101
I-ctg	0.905	0.974	0.938	78
B-value	0.962	0.962	0.962	131
I-value	0.962	0.966	0.964	444
O	0.959	0.838	0.895	365
加权平均/合计	0.962	0.963	0.962	3633

### 3.4 本章小结

建筑规范条款是一种基于某些被广泛被接受的自然语言行文范式写成的，相对于日常用语，条款则显得更加规范、更少歧义。复杂的语义结构可以引出更复杂且覆盖面更广的语言表达模式，但是因为本文的最终目的并不是尝试去寻找这些显式的表达模式，而是使用大语言模型试图像人一样去对条款进行解析，而对于人来说，一些处于上层的简单标签可以帮助减少人类在此类任务中因为对专业词汇不熟悉而出现断句错误以及条件子句之间的逻辑连接错误的概率。

因此，本研究首先利用专业经验尝试对现有规范的范式进行较为普适的语义元素标签和框架定义，推进相关领域数据集的互通；再使用 doccano 平台对数据集进行 BIO 格式标注，导入到 BERT 词性标注模型进行训练；期间本研究针对模型超参基于以往研究的训练经验进行了批大小、学习率的优化实验，最后选择了批大小=16、学习率=5e-5、epoch=20 时得到的词性标注模型。该模型在词性标注测试集的平均准确率、召回率和 F1 分数分别达到 96.2%、96.3% 和 96.2%。

## 第 4 章 基于大语言模型提示工程的条款解析方法

### 4.1 概述

在第三章中，本研究将规范中的条款进行了自动标注，从而获得了条款行文中的元素级表达。而条款整体的结构性表达则需要理解条款内部元素之间的逻辑关系，将不同的元素进行有序的连接，这便是条款解析部分需要解决的问题。此前研究专注于基于规则或形式文法的方式将元素之间的关系进行预先定义从而完成条款元素关系的解析，似乎在不断尝试以白箱的方式对语言的边界进行框定并向内发散，从而使得整个解析任务变得复杂无比以至于难以推进。而现有的大语言模型则完全具备类似人类的“黑箱”语言能力，似乎是解决自动合规性检查任务中“条款解析”问题的终极途径。而现有大语言模型虽然具备着极强的语言能力，但有时候却需要通过一定的特殊“提示”才能发挥出完整的理解、推理以及生成能力。

因此，本章以“提示工程”入手，在 4.2 节阐述了现有提示工程中常用的技巧，并详细说明了如何将这些技巧转换成对应条款解析任务中所需要的提示词的过程。而后在 4.3 节对于现有的多种大模型进行了列举和介绍，并将 4.2 节中的不同提示词进行排列组合后带入到不同大模型中进行组合实验。最后在 4.4 节对不同提示组合、不同大语言模型在不同类型条款数据解析实验中的表现进行对比分析。

### 4.2 针对暖通规范解析的提示词构建

#### 4.2.1 通用提示词构建

在大语言模型接收输入提示时，首先将输入文本转换为模型可以理解和处理的 token。这些 token 经过转换器层的管理，该层捕获它们之间的关系和上下文。在这些层中，注意机制根据标记之间的相关性和上下文为它们分配不同的权重。在注意力机制的处理之后，模型形成了输入数据的内部表示，称为中间表示。然后，这些表示被解码回人类可读的文本<sup>[72]</sup>。

这个过程的一个重要方面是随机性函数。该函数受到两个主要参数的影响，即温度和 top-k 采样。首先，温度参数平衡了输出中的随机性和确定性。较高的

温度值会导致更多的随机输出，而较低的温度值则使输出更具确定性。其次，top-k 采样将模型的选择限制在每一步输出生成的最可能的前 k 个标记范围内。最后一个阶段是输出生成，模型在这个阶段生成最终的文本。<sup>[92]</sup>

提示词可以包含以下任意要素<sup>[93]</sup>：

1. 角色指派：让大语言模型以某一角色进行任务的执行；
2. 指令提示：希望模型执行的特定任务或指令；
3. 上下文信息：包含外部信息或额外的上下文信息，引导语言模型更好地响应；
4. 领域知识：大语言模型是通用模型，可以通过输入领域知识让模型进行上下文学习，从而让模型以所提供的规则和知识框架下执行任务；
5. 任务案例：大语言模型可以通过既定格式的输入案例理解并模仿输入和输出的范式进行工作；
6. 输入数据：用户输入的内容或问题；
7. 输出指示：指定输出的类型或格式；

本节首先介绍了大语言模型开发中常用的提示技巧，由此形成此次课题的基础静态指令。接着对于条款解析任务中所需要的案例以及特点结合更多逐渐涌现出的提示技巧进行上下文和动态案例提示词构建，如图 21；最后对于输入和输出格式进行定义，形成最终提示词。

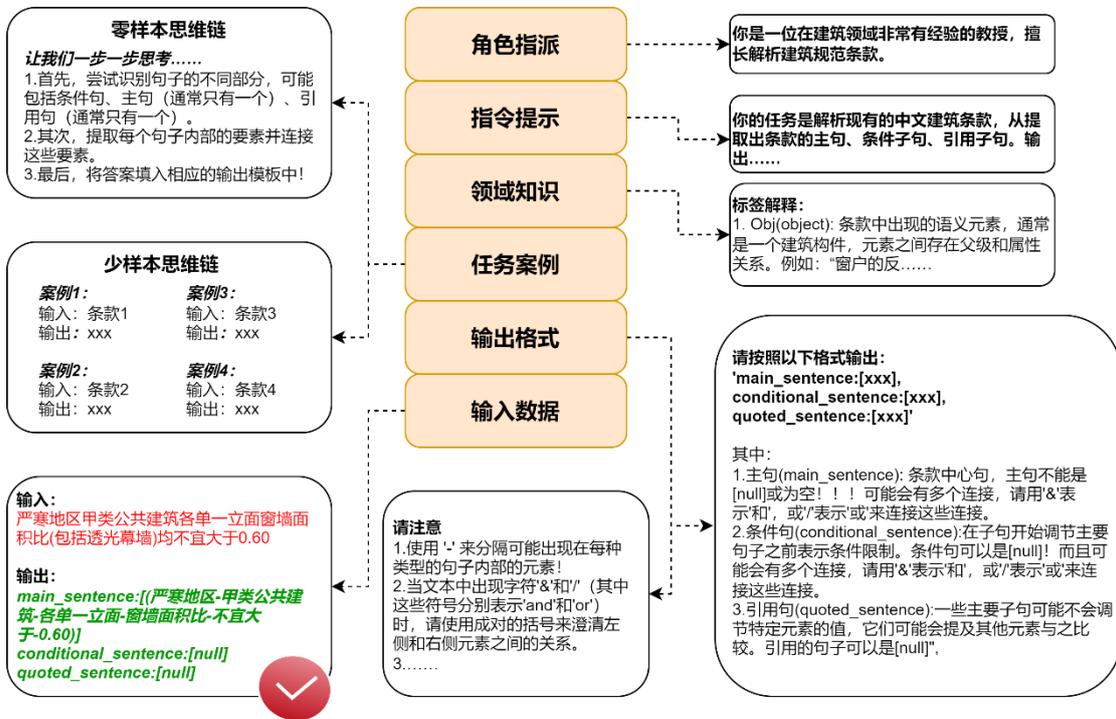


图 21 提示词构建概览

注意，以下的所有演示案例均在 openai playground 的“聊天模式”中进行。  
(<https://platform.openai.com/playground?mode=chat>)

### 1) 角色指派和指令提示

角色提示是工程中的一种基本方法，其目的是为大语言模型赋予特定角色，如一个有帮助的助手或知识渊博的专家<sup>[94]</sup>。通过为大语言模型分配角色，我们为其提供了一定的上下文，这有助于模型更好地理解问题。通过更好地理解问题，大语言模型通常能够提供更好的答案。在各类大语言模型平台中，“你是一个非常有经验的……”的指令已经成为“系统提示”的代名词。

角色指派的机制在于为大语言模型提供了明确的行为框架。通过设定特定的角色，大语言模型能够更好地理解和适应任务类型。例如，当大语言模型被指定为“研究助理”时，它倾向于进行深入的学术研究和分析，而作为“创意作家”时，则会激发更多的创造性和想象力。在本研究中，大语言模型的任务是解析暖通规范，其中可能涉及建筑内交叉行业的内容，如给水、排水、电气等。因此，相应的角色可以是“建筑规范撰写者”、“设计院暖通设计师”或“建筑领域教授”。

指令明确可以提高模型运行效率，因为它消除了歧义。明确的指令减少了系统解释指令的不确定性，从而降低了错误解释的可能性。这种明确性使大语言模型能够更快地处理信息，并提供更精确的结果。

### 2) 领域知识

本研究在第三章将现有的条款进行了总结并提出了一种基于大语言模型解析的语义结构，并基于这个语义结构进行了自动打标模型的训练并带入大语言模型解析。对于大语言模型来说，它并不清楚本研究中所输入的语义标签含义。所以，在提示中对于语义结构的描述和说明是必要的。

在这一部分，本研究添加了：

1. **obj(object)**: 表示条款中出现的语义元素，通常是一个建筑构件，元素之间存在父级和属性关系。例如：“窗户的反射率”中，存在“窗户”和“反射率”两个语义元素，“反射率”是“窗户”的属性。元素之间的连接可能是并列也可能是上下游关系，比如：窗户[obj]的长[obj]和宽[obj]；也可能是存在包含关系，比如：窗户[obj]边框[obj]的厚度[obj]，这里的厚度和边框就是从属关系而不是并列关系。

2. **cmp(compare)**: 是比较谓词，比如：“应当满足”、“不应大于”、“不应小于”，表示比较和判断。一个待执行条款必须存在至少一个比较词。

3. **qtg(quoted trigger)**: 属于触发标签的一种，是一种表示引用的触发标签。比如：“应当满足[cmp]表 1-1 的要求[qtg]；应当满足[cmp]《某个规范》的要求[qtg]。”

4. **ctg: conditinal trigger**, 属于触发标签的一种，是一种表示条件的触发标签。比如：“当 xxxx 时[ctg]”、“在 xxxx 条件下[ctg]”。

5. **value**: 代表具体值，比如：“甲类[fobj]商业建筑[fobj]窗户[obj]的反射率[prop]应当大于[cmp]0.6[value]。”

## 4.2.2 思维链构建

思维链方法通过让大型语言模型生成一系列中间推理步骤来工作，显著提高了它们执行复杂推理任务的能力<sup>[66]</sup>。这种方法之所以有效，是因为它模仿了人类解决问题的过程，将复杂的任务分解成更小、更易管理的步骤。因此，思维链方法的优越性在于它能够分解多步骤问题，使得对更复杂任务分配更多的计算变得更容易。首先，设计一个包含连续推理步骤的提示（如图 3），引导语言模型不仅仅给出答案，而是生成一系列解决问题的中间步骤。这些步骤通常是自然语言形式的，仿照人类解决问题时的思考过程。通过这种方式，模型被引导去明确地表达其推理路径，而不是直接跳至最终答案。这种方法使得模型的输出更加透明和可解释，有助于理解模型是如何得出特定答案的，并且使得推理路径的调试更加容易。

此外，这种方法在包括算术、常识推理和符号操作等多种任务中都适用，使其具有高度的通用性。最后，它可以通过包含思维链序列的少数示例提示，轻松地在大型现成的语言模型中实现。

因此，对于条款解析任务来说，本研究拟采用“零样本提示”思维链和“多样本提示”思维链来分别对应忽略以及包含“解析数据集”制作过程的思维链提示。

### 1) 零样本思维链

“零样本思维链”（Zero-shot Chain of Thought）是指在不提供特定任务训练样本的情况下，利用大型语言模型解决问题的能力<sup>[95]</sup>。具体来说，这种方法不依赖于针对特定任务的训练样本，而是直接利用模型对问题进行推理和解决。通过这种方式，模型能够在没有直接训练的情况下展示出解决问题的能力，显示了模型的推广能力和适应新任务的潜力。

从技术上讲，完整的零样本思维链过程涉及两个单独的提示/补全结果。在图 22 中，左侧的顶部气泡生成一个思维链，而右侧的顶部气泡接收来自第一个提

示（包括第一个提示本身）的输出，并从思维链中提取答案。这个第二个提示是一个自我增强的提示。

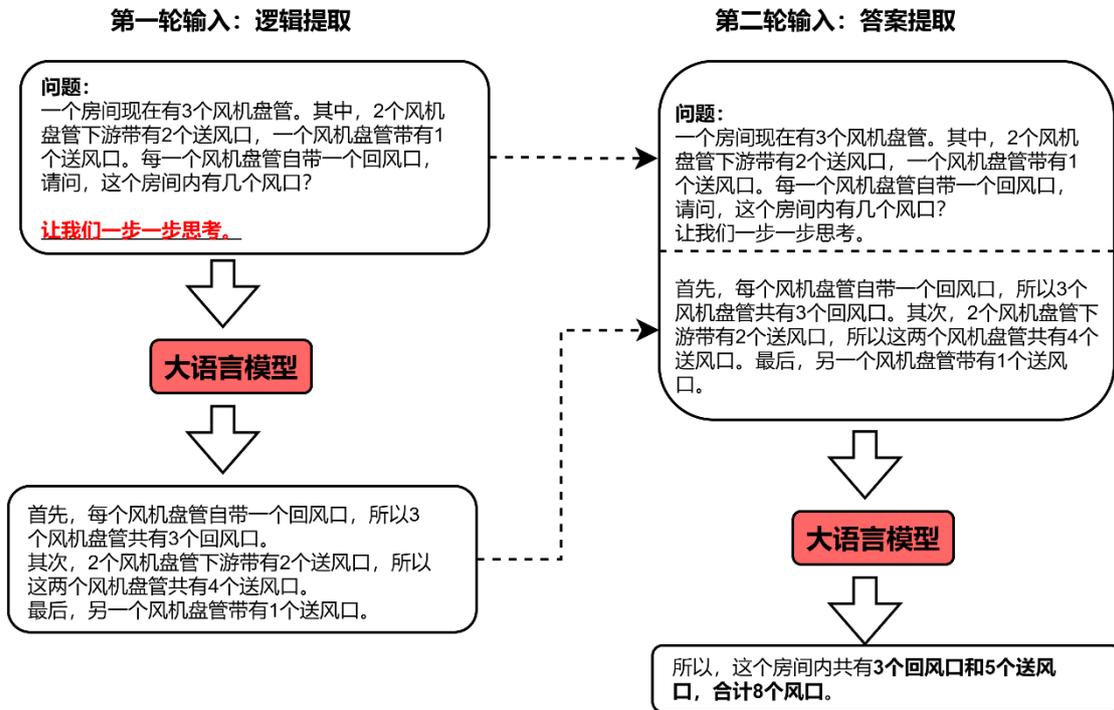


图 22 零样本思维链原理

人们尝试了许多不同的零样本思维链提示（例如“让我们按步骤解决这个问题。”或“让我们逻辑思考一下。”），但普遍发现“让我们一步一步地思考”对于他们选择的任务最有效。因此，在本研究中，零样本思维链的构建为：

“让我们一步一步思考：

- 1.首先，尝试识别条款的不同部分，可能包括条件句、主句（通常只有一个）、引用句（通常只有一个）。
- 2.其次，提取每个句子内的元素，并连接这些元素。
- 3.最后，将答案填入预置架构中！”

## 2) 多样本思维链

“多样本思维链”（Few-shot Chain of Thought）是指在有限的样本训练基础上，使用大型语言模型解决问题的方法。这种方法涉及向模型提供少量特定任务的示例，以指导其理解和解决类似问题。这些示例通常包含了完成任务所需的逐步推理过程。通过这种训练方式，模型学习如何对类似的新问题进行有效的推理，

从而提高在新任务上的性能<sup>[72]</sup>。

与“零样本思维链”不同的是，“多样本思维链”需要对大语言模型的输入案例提供分析过程，帮助大语言模型理解特定任务步骤的解决范式。因此，按照第三章所提出的语义结构。本研究拟采用“有标签”和“无标签”两类数据通过人工撰写提示的方式，抽样形成 2×20 的提示案例数据库，其中：

#### 有标签案例（例）：

- 输入：“[在/ctg]名义工况和规定[条件下/ctg]，[锅炉/obj]的[热效率/prop][不应低于/cmp][表 4.2.5/qtg]的[数值/prop]。”
- 思维链：
  - (1) 为解析这个条款，需要从条件句、引用句和中心句三个角度进行分析
  - (2) 中心句 (Main Sentence)：中心句阐述了该条款中的主要规定，本条款最终的目的是规定锅炉的热效率。因此，中心句在“锅炉的热效率不应低于表 4.2.5 的数值”这句话中。条款在 cmp 之后依然存在 obj，这可能存在引用。根据判断，“表 4.2.5 的数值”确实为引用。因此，中心句为：“锅炉-热效率-不应低于-quote”
  - (3) 条件句 (Conditional Sentence)：表明这个标准是在特定操作条件下适用，在这里为“名义工况&名义规定”。
  - (4) 引用句 (Quoted Sentence)：引用的是具体的表或条款或标准，本条款存在引用，即“表 4.2.5-数值”。
- 输出：“main\_sentence: [锅炉 - 热效率 - 不应低于 -quote], conditional\_sentence: [名义工况&名义规定], quoted\_sentence: [表 4.2.5-数值]”

#### 无标签案例（例）：

- 输入：“在名义工况和规定条件下，锅炉的热效率不应低于表 4.2.的数值。”
- 思维链：
  - (1) 为解析这个条款，需要从条件句、引用句和中心句三个角度进行分析
  - (2) 中心句 (Main Sentence)：中心句阐述了该条款中的主要规定，本条款最终的目的是规定锅炉的热效率。因此，中心句在“锅炉的热效率不应低于表 4.2.5 的数值”这句话中。根据判断，“表 4.2.5 的数值”为引用。因此，中心句为：“锅炉-热效率-不应低于-quote”
  - (3) 条件句 (Conditional Sentence)：表明这个条款是在特定操作条件下适用，在这里为“名义工况&名义规定”。
  - (4) 引用句 (Quoted Sentence)：引用的是具体的表或条款或标准，表示本条

款存在引用内容，在本条款中指代“表 4.2.5-数值”。

- 输出：“main\_sentence: [锅炉 - 热效率 - 不应低于 -quote], conditional\_sentence: [名义工况&名义规定], quoted\_sentence: [表 4.2.5-数值]”

### 4.2.3 动态案例构建

动态案例选择的必要性在于，为了提高语言模型在特定任务上的表现，需要提供相关、数量有限且多样化的示例，这有助于模型更好地理解 and 应对各种问题。通过使用案例选择排序并选取排名靠前的案例作为最终案例提示。

在提示工程中，通过使用案例选择排序并选取排名靠前的案例作为最终案例提示，前面排名的案例往往具有更高的信息丰富度，能够提供更多的细节和背景信息，有助于用户更好地理解 and 应用这些案例<sup>[93]</sup>。排名靠前的案例通常与当前输入或问题更相关，能够更好地满足用户的需求，提供更有针对性的信息和解决方案。良好的案例选择排序算法会考虑到多样性，确保选取的案例在主题、内容、观点等方面具有一定的差异性，从而提供更全面的视角和解决方案。由于提示工程通常用于帮助用户快速获取信息或解决问题，因此选取排名靠前的案例可以提高任务流程效率，减少用户阅读和筛选的时间成本。综上所述，通过案例选择排序并选取前面几个案例作为最终案例提示，可以确保所提供的提示信息具有更高的信息质量、相关性和实用性，从而更好地满足用户的需求。

#### 最大边际相关性 (Maximal Marginal Relevance, MMR):

MMR 排序算法结合了案例与输入的相似度和多样性优化。这通过寻找与输入最具余弦相似度的嵌入案例，然后在添加它们的同时对它们进行接近已选案例的惩罚来实现。MMR 平衡了相关性和多样性，选择既与查询相关又在已选示例中独特的案例。这有助于防止重复且提供全面的示例。

MMR 被用来选择与输入之间相关性最高的案例。这种算法旨在平衡两个关键因素：1) 选择与输入最相关的案例，以提供有效的信息；2) 选择与已选案例之间的差异最大的案例，以提供多样性和全面性。这一过程基于以下两个关键参数：1) 相关性 (Relevance): 衡量输入与候选案例之间的相似性或相关性。通常使用余弦相似度或其他相似性度量来衡量。2) 差异性 (Diversity): 衡量已选案例之间的差异程度，以确保所选案例具有多样性和全面性。这可以通过计算候选案例与已选案例之间的距离或差异性来实现。具体步骤如下：

1. 计算相关性：对于每个候选案例，计算其与输入之间的相关性得分。
2. 计算差异性：对于已选案例，计算它们之间的差异性得分。这可以通过计算

已选案例之间的相互距离或差异性来实现。

3. 综合相关性和差异性：将相关性和差异性得分结合起来，通常使用以下公式：

$$MMR = \arg \max_D \lambda \cdot \text{Relevance}(D, q) - (1 - \lambda) \cdot \max_{d_i \in D} \text{Similarity}(d_i, D) \quad (4.1)$$

其中：

- $MMR$  表示最大边际相关性（Max Marginal Relevance）。
- $D$  是已选案例的集合，即条款案例库。
- $q$  是输入条款句子。
- $\text{Relevance}(D, q)$  表示已选案例与输入之间的相关性。
- $\text{Similarity}(d_i, D)$  表示候选案例  $d_i$  和案例集合  $D$  中最相似的案例之间的相似性。
- $\lambda$  是一个权衡相关性和差异性的参数。

4. 选择案例：根据综合得分选择最佳的候选案例加入到已选案例中，重复该过程直到达到所需数量的案例或其他终止条件。

在 4.2.2 节的“多样本思维链”中，本研究制作了一套提示案例数据库（详见附录 C），每个条款均按照“输入-思维链-输出”的范式被制作成“黄金标准”。而在实际输入一个条款时，并不需要将所有的案例全部打包制作成提示的一部分。在实际的多样本案例提示工程实践中，一般只需 4-8 个案例对于简单的任务已经达到较为明显的提高，为了减少解析的成本，本研究选取排序后前 4 个条款作为最终样本案例。

如图 23，在本课题中， $MMR$  选择器从一系列案例中选择，这些案例通过 OpenAI 嵌入类和 FAISS 向量存储类来测量语义相似度，并基于这些信息选择  $k$  个案例。这种方法有助于保证所选案例不仅相关性高，而且具有多样性，提高了案例选择的有效性和覆盖面。

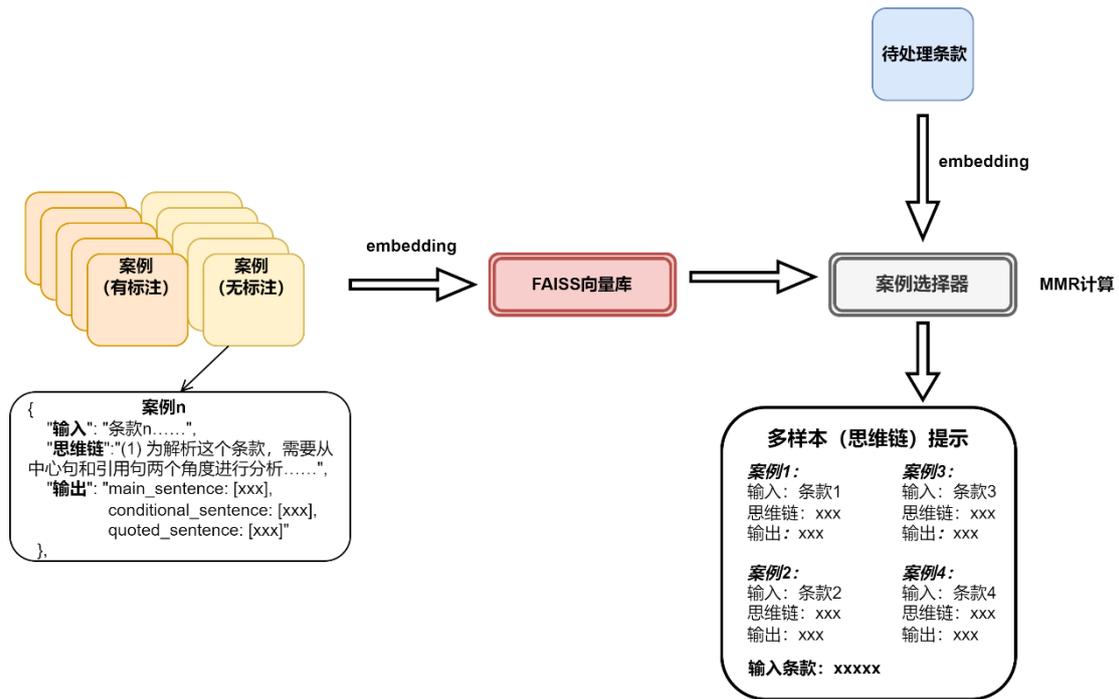


图 23 动态案例制作流程

### 4.3 大语言模型、提示词组合实验

由于当下大模型仍然处于快速增长期，虽然学界、业界在不断地推出各自的大语言模型，但是模型的训练语料绝大部分都是英文语料，因为建筑行业的中文语料（尤其是经过业内审核、被广泛传播和认可的中文语料）更是少之又少。因此，大语言模型在 AEC 行业中的应用依然存在较大的专业壁垒。大语言模型在 AEC 行业中的应用也依然处于摸索阶段，现阶段针对建筑自动合规性检查而言，业界没有一个较为明确的模型偏好，尤其是针对中文条款的解析。

因此，本节将尝试使用当下较为流行的多个大语言模型并利用多种提示策略进行组合实验对建筑条款进行解析，尝试得出更为适合的大语言模型和提示策略组合。本研究考虑到模型在测评中的表现、模型在开源社区中的受欢迎程度、模型所经过的训练语料选择了以下八个当下较为主流的大语言模型进行测试，如表 6:

表 6 实验所选取大模型

模型	是否开源	参数量（亿）	发布时间
----	------	--------	------

GPT-3.5	×	175	2022 年 11 月
GPT-4	×	1750	2023 年 3 月
ERNIE-Bot 4.0	×	2600	2023 年 10 月
Qianfan-Chinese-Llama-2-13B	×	130	2023 年 3 月
Baichuan2-7B-Chat	√	70	2023 年 9 月
通义千问	×	720	2023 年 9 月
ChatGLM2-6B	√	60	2023 年 3 月
Ernie-bot 3.5	×	100	2023 年 3 月

在本研究中，我们聚焦于探索适用于建筑条款解析的高效大语言模型和提示策略组合。基于模型的性能、社区影响力以及训练语料的广泛性，选定了八款领先的大语言模型，即 GPT-3.5、GPT-4、ERNIE-Bot 4.0、Qianfan-Chinese-Llama-2-13B、Baichuan2-7B-Chat、通义千问、ChatGLM2-6B 和 Ernie-bot，涉及开源与非开源、不同参数规模与发布时间跨度。其中，OpenAI 的 GPT-4 与 ERNIE 系列均突破了百亿级别的参数量，显示出对复杂话题与上下文的深度理解能力；而百度文心大模型与百川智能的 Baichuan2-7B-Chat 分别聚焦于知识增强与高质量语料训练，旨在不同的应用需求和性能优化上取得领先。ChatGLM2-6B 则以开源双语对话模型身份，通过混合目标函数和人类偏好对齐训练，在多个数据集上展现出卓越性能。

表 7 单个模型提示策略组合

实验编号	角色	指令	标签	输出编码	标签说明	少样本案例	零样本思维链	少样本思维链
1	√	√	√	√				
2	√	√	√	√	√			
3	√	√	√	√		√		
4	√	√	√	√			√	
5	√	√	√	√				√
6	√	√	√	√	√	√	√	

7	√	√	√	√	√		√	√
8	√	√	√		√	√	√	
9	√	√	√	√			√	√
10	√	√		√				
11	√	√		√		√		
12	√	√		√			√	
13	√	√		√				√
14	√	√		√		√	√	
15	√	√		√			√	√
16	√	√				√	√	

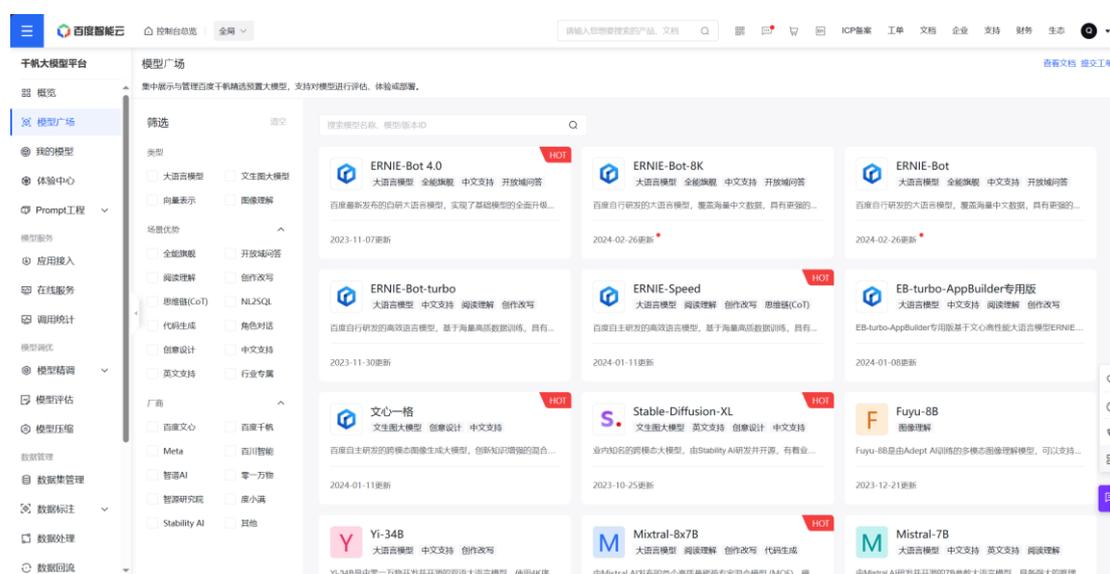
基于4.2节所提出的提示策略，本研究设计了一系列实验以探索大模型在特定角色、指令、标签、输出编码、标签说明以及少样本或零样本情境下的应用（表7，内容详见附录A），从而根据指令遵循率、格式遵循率和解析正确率三个指标来评判不同模型再不同提示策略下对建筑条款解析中的效果，并计算解析所需要的时间成本（s/条款）以及计算花费成本（元/条款）。这些实验组通过不同组合的配置揭示了多种提示策略对模型性能的影响：

实验1-5提供了基础提示策略的效果评估，从最基本的角色、指令、标签及输出编码构成（实验1），到加入标签说明（实验3）及探究零样本（实验4）与少样本（实验5）的思维链变量对模型性能的影响；实验6与7包含了所有变量，因此，通过实验6和实验7的对比，预计可以得到表现最佳的策略提示组合。两组的不同之处在于实验6包含的是少样本案例而实验7包含的是少样本思维链，区别在于所提供的案例中是否包含所提供案例的解析思维链过程，从而探究同样提供部分案例的情况下，思维链对于解析过程的重要性；实验8可以和实验6进行对比，探究输出编码对于存在样本输出案例提示的重要性，即：大语言模型是否能够不依靠输出格式知识，从建筑条款解析案例中学会正确输出格式。实验9和实验7可以试求探究大语言模型是否需要通过具体的标签说明才能完整地运用条款标签这一辅助工具，亦或是可以通过案例说明可以直接对标签的含义进行参悟。

实验10-16进一步探索了在移除标签制约的情形下模型性能的变化。这种配置与实验1、3、4、5、6、7、8进行对应，可以对比得出跳过前序词性标注信息

时，大预言模型对于原始条款的解析能力。

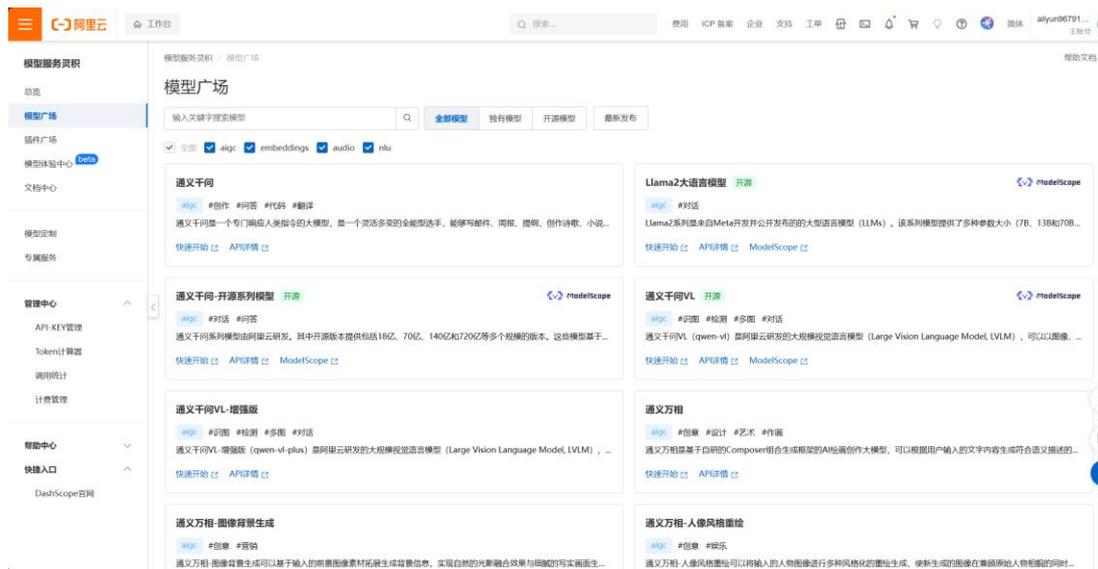
本课题将通过云服务大模型平台所提供的 API 在 Python 环境下进行调用。由于实验数据组数较大，本课题将从《公共建筑节能设计标准》(GB50189-2015) 中随机筛选 20 条规范条款，作为每一组策略下的待解析对象。其中在阿里云灵积模型平台 (<https://dashscope.console.aliyun.com/model>) 使用通义千问大模型；在 opena (<https://platform.openai.com/>) 使用 GPT-4、GPT-3.5 大模型；在百度智能云千帆平台 (<https://console.bce.baidu.com/qianfan/modelcenter/model/buildIn/list>) 使用文心 4 大模型、Ernie-bot-3.5 大模型、Qianfan-Chinese-Llama-2-13B 大模型和 ChatGLM2-6B 大模型



a 百度智能云千帆大模型平台模型界面



b Openai API 界面



c 阿里云灵积大模型平台

图 24 所使用大模型平台

## 4.4 条款解析结果和实验分析

### 4.4.1 解析结果评价指标

#### A. 指令遵循-->True or False:

- a) 判定大预言模型是否按照要求基于主句，条件句，引用句的角度对条款进行解析并输出结果；
- b) 如图 25, 模型回复内容必须出现“main\_sentence”、“conditional\_sentence”、“quoted\_sentence”字段或是这三种类型的句子字段的中文版本如：“主句”、“条件句”、“引用句”；
- c) 如图 25 解析结果示例，输出的隔断线或是空格数量可以不一致，因为此类情况的输出可以通过后续统一去除句子之间空格进行弥补；
- d) 若输出不完整但是省略的内容为 null，则判定遵循指令；
- e) 若模型生成结果后不断自主生成和条款无关的条款并继续解析，则判定模型没有理解解析任务，为 False。

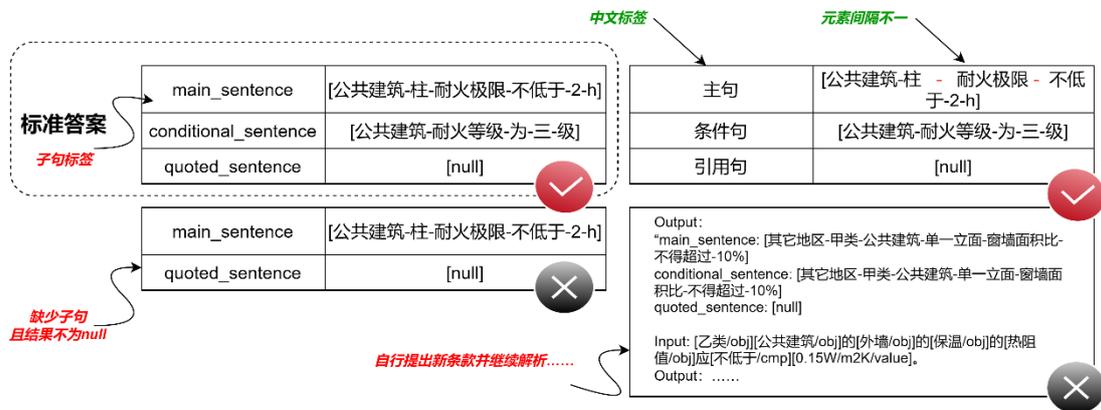


图 25 解析结果示例

**B. 格式遵循-->True or False:**

该指标关注解析结果是否按照规定格式输出，以确保在后续的逻辑提取中正确获得语义元素。即：词语之间需要用“-”进行连接，如“屋顶-透光部分-面积”，若出现没有按照这种形式呈现，判定为 False；解析结果可以出现解析结果之外的内容，如说明性内容、冗余内容，此情况整体不影响三种类型子句的提取。

**C. 分类正确-->True or False:**

该指标和指标 A 有关，指标 A 目的是判断结果生成的是否正确理解指令，指标 C 则关注三个子句结果是否分类正确；同时，正确答案不唯一，例如：“夏热冬暖、夏热冬冷、温和地区的建筑各朝向外窗（包括透光幕墙）均应采取遮阳措施”中，“夏热冬暖、夏热冬冷、温和地区”可以是作为条件局中的元素，也可以是主句的元素。

**D. 连接正确-->True or False:**

该指标最为严格，首先必须满足指标 A、指标 B、指标 C，且在单个子句内

满足句子元素之间的识别和连接均正确。其中基本连接方式为：obj- com- (obj/value)；

#### E. 主观评分-->分数

该指标为本课题研究人员主管评价，以加强不同模型之间的区分度。本课题列出以下分数区间以供标注人员参考：

- a) 0：模型没有理解指令，没有进行解析行为，输出结果完全无法使用；
- b) 60：模型对条款进行了解析，解析的结果可以让人看懂，但是有明显错误，无法使用；
- c) 80：模型对条款进行了解析，解析的结果基本正确，格式基本正确但是有明显错误，需要进行进一步后处理后才可使用；
- d) 90：模型对条款进行了解析，解析的结果正确，格式正确。存在不明显的错误或者可以通过少量后处理使用；
- e) 100：和正确答案匹配，完美解析。

### 4.4.2 实验评价结果分析

根据图 27 可以看出，在所测试的八种大语言模型中，GPT-4、文心 4、GPT-3.5 和通义千问在整体的条款解析任务中表现较为出色且稳定。表 8-9 展现了多个模型在评价环节中的总实验组平均分以及对应时间、API 花费数据。文心 4 大模型 (ERNIE-Bot-4.0) 在“指令遵循”、“格式遵循”、“分类正确”和“主观评分”四个指标中获得了最高分，仅在最终解析正确上略逊于 GPT-4 的 93.13%，且相比于 GPT-4 单条款解析时间，文心 4 仅需 8.7 秒就能完成一个规范条款的解析。这说明文心 4 在中文语境下的建筑条款解析任务中能够媲美甚至超越“当下最佳”的 GPT-4 模型，并且文心 4 更高的主观评分往往说明该模型解析能力较为稳定，很少输出较低分数的解析结果，这对于条款解析来说是更为利好的；除此以外，通义千问的表现令人惊喜，作为阿里旗下的大语言模型，通义千问在“指令遵循”、“分类正确”、“解析正确”和“主观评分”四类指标中均超过了 GPT-3.5，且和评分最高的文心 4 和 GPT-4 在中文条款解析中的表现差距甚微，并且解析成本仅为文心 4 的 18.16%。

表 8 模型解析结果评价

模型	指令遵循 (%)	格式遵循 (%)	分类正确 (%)	解析正确 (%)	主观评分 (分)
----	----------	----------	----------	----------	----------

GPT-3.5	93.13	96.25	90.94	85.00	90.44
GPT-4	93.75	94.06	87.19	93.13	95.47
ERNIE-Bot-4.0	98.75	98.13	97.50	91.25	96.38
Qianfan- Llama-2-13B	69.06	48.06	67.19	43.13	61.10
Baichuan2-7B-Chat	9.38	7.19	8.13	5.31	7.72
ERNIE-Bot-3.5	90.00	92.19	88.44	80.31	85.69
ChatGLM2-6B	64.69	72.50	62.81	58.75	65.00
通义千问	94.75	95.31	93.31	87.19	93.66

表 9 模型解析消耗成本

模型	时间 (s/条款)	API 花费 (元/百条款)
GPT-3.5	8.26	1.61
GPT-4	13.31	14.68
ERNIE-Bot-4.0	8.70	14.59
Qianfan- Llama-2-13B	9.23	0.56
Baichuan2-7B-Chat	3.31	2.67
ERNIE-Bot-3.5	9.53	1.65
ChatGLM2-6B	11.37	0.54
通义千问	8.70	2.65

相反，百川、千帆和 ChatGLM 实验中的表现均不如人意，其根本原因可能是模型本体参数体量较小导致模型在逻辑推理和对话生成的基本能力不足导致。其中百川大模型在所有的实验组表现最不佳，模型的回复错误主要集中在模型不能准确理解指令意图，导致解析内容以及格式形式发散较为严重。在图 26 中标注了该模型在 4 个案例中的典型错误，在第一个案例中，模型虽然以不满足格式要求的方式列出了三类子句，但是这个回答仅仅只是对前半句进行了解析，并且该子句被错误地分到了主句中，最后模型在引用句部分甚至擅自杜撰了其中的内容；在第二个案例中，可以看到模型在单词的拼写和文字输出环节中出现了很大

的问题，说明模型在基础的对话能力上依然有较大的改进空间；第三个案例同样也说明了这个问题，主要体现在模型的输出中出现了大量不成文或不相干的内容；第四个案例种的错误在百川大模型以及千帆、ChatGLM 中也经常出现，语言模型的输出通常是以预测特定的停止词来决定是否继续进行文字的生成，因此在某些存在思维链的实验组中，模型会因为思维链中貌似和条款本身无关的内容导致提前停止，这个问题的本质也是模型本身没有理解提示中的任务，从而导致解析表现低下。

图 28 和表 10 展现了不同实验组在多个模型下的平均表现分数。其中，绝大部分的主观评价分数在 80 分以下，表现较好的实验组有：3、6、8、11、14、16 组。这几组的最终提示中均包含了少样本提示词，且实验组 2-5 与实验组 1 的主观评价分相减得到 4 组实验组在多个评价指标中的影响如表 12 所示。实验组 3 在 5 个指标中的均有较大的提升，在五个指标中提升比例分别为 8.40%、30.43%、6.16%、25.72%和 16.52%。因此，本课题初步可以得出“少样本案例提示”对于大语言模型解析建筑规范条款这一任务来说是最为必要的。

表 10 不同实验组在多个模型下的平均表现分数

实验组	指令遵循%	格式遵循%	分类正确%	解析正确%	主观评分
1	72.78	63.89	72.22	58.33	68.28
2	83.33	73.33	80.00	60.00	75.83
3	78.89	83.33	76.67	73.33	79.56
4	66.67	57.22	63.89	53.33	60.39
5	70.0	71.11	69.44	67.78	68.11
6	86.67	86.11	82.22	76.11	83.89
7	64.44	67.22	63.89	63.33	63.28
8	84.44	85.00	82.78	75.00	81.83
9	66.11	74.44	65.00	63.33	64.39
10	75.00	66.67	71.67	60.00	68.67
11	88.89	88.89	85.00	78.33	86.44

12	83.89	72.78	83.89	66.11	77.72
13	72.22	77.78	70.00	65.00	71.44
14	86.11	87.78	85.56	77.22	85.89
15	73.89	86.11	75.00	71.67	75.89
16	85.00	85.00	85.00	76.67	82.94

表 11 不同实验组在多个模型下的平均成本

实验组	时间 (s/条款)	API 花费 (元/百条款)
1	5.02	2.45
2	4.85	3.74
3	8.53	4.18
4	5.40	2.69
5	12.32	7.17
6	9.45	5.75
7	13.10	8.64
8	10.32	4.61
9	13.10	7.36
10	4.68	2.45
11	8.47	3.66
12	3.98	2.53
13	12.65	6.73
14	9.05	3.92
15	12.73	7.0
16	8.54	2.8

根据所有模型的平均评分来看，看似具有最完备提示的实验组 6 和实验组 7

并不是表现最好的实验组。表现最好的实验组是实验组 11，即在实验组 6 的提示中删除了标签、标签说明、零样本思维链三个部分；除此之外，在所有模型的平均打分中，实验组 7 和实验组 9 之间存在唯一的“标签知识”提示作为唯一变量，显示似乎在提示中对于标签知识的提供会对模型表现起到负作用。但是如果本研究将模型分成“表现一般”、“表现良好”和“表现优秀”的三组，我们可以观测到在“表现较差”的模型实验中，均是没有词性标签的第 11 组或第 14 组获得最高评分，但是在“表现良好”以及“表现优秀”的模型中，带有标签的实验组 7 和实验组 5 取得了更高的评分。本研究推测是由于所提供的标签为大语言模型的理解能力提出了更高的要求，导致逻辑推理能力较弱的语言模型无法正确意识到标签的作用。这一点最直观的体现可以从“表现良好”的模型中看出：实验组 7 相较于实验组 9 在“指令遵循”、“分类正确”、“解析正确”和“主观打分”中分别有着 47.5%、45%、32.5%和 41.5 分的提升。同样，在“表现一般”模型中，第 2 组和第 1 组之间通常伴有较为明显提升。因为逻辑推理有限的模型无法像人一样自主地理解标签的含义，添加了显式的标签说明提示之后，模型对于带有标签条款的解析能力便有了较大幅度的改观。然而，对于逻辑推理能力较强的模型来说，词语标签实际是通过第 3 章所训练语言模型预测向量的显示近似，对于解析任务来说，通过严格按照预设标签说明进行条款的结构化似乎不是一个最完美的方式。因此，基于固定的标签来解析建筑规范条款可能对于当下表现最为优异的大模型来说是一种枷锁，可以在最终的提示中被舍弃。

在实验组 6 和实验组 8 的对比中，表现良好和表现优异的模型会因为输出编码格式提示均在五类评价中存在少量但是稳定的提升，而在指令遵循能力较差的模型中起到了相反且稳定的干扰作用。但是因为模型从所提供的少样本案例中学到绝大部分的输出范式，所以专门的编码说明给予模型输出效果的提升比例较为有限，所以在模型输入存在限制的情况下（如，出于模型本身 embedding 限制或是 token 计费经济性考虑存在尽量减少提示字数要求），可以被适当舍去。

综上，本研究认为：“标签说明”作为带有标签输入条款的补充知识可以起到一定的作用，对于当下非头部大语言模型来说，是一个提升整体解析效果较为有效的提示方式，而对于顶尖的大模型来说，此提示存在限制其解析能力的隐患，但是对于格式遵循有一定的作用，可选择性使用；“输出编码”对于当下大语言模型的“格式遵循”能力是极为显著的，推荐后续研究继续使用，从而进一步提升最终“解析正确”的比例。“少样本案例”通过部分案例的展示，对于模型解析条款能力的提升是普遍且明显的。而同样给予模型案例展示的“少样本思维链”因为案例所含逻辑不够直接，会给基础推理能力较差的模型产生明显的负面影响，通常表现为“指令遵循”指标快速下降，从而导致后续的解析无法进行。而对于

能力较强的大语言模型来说，此类少样本思维链的展示会在解析方式上给大语言模型提出更多的规则导致其解析效果不及直接提供少量输入和输出样本，因此本研究更加推荐后续研究使用“少样本案例”作为提示的最终组成之一，舍弃“少样本思维链”。但并非所有思维链都会对模型产生负面影响，“零样本思维链”通过“让我们一步一步来……”的方式减少了对推理能力较差模型的干扰，在此类模型中获得了正向的帮助。

因此如表 13-16，从所调用模型角度来看，最适合的提示策略分别为：

对于基础性能较差的大语言模型（百川大模型在解析条款的结果中表现尤其不佳，无法作为比较和推荐的依据，故被舍去），应当以千帆大模型的实验组 11 进行配置并适当添加零样本思维链，即：角色指派+指令明确+输出编码+少样本案例+（零样本思维链）。

对于基础性能良好的大语言模型，应当以实验组 7 进行配置并根据模型逻辑推理能力适当用少样本案例替换少样本思维链，即：角色指派+指令明确+标签+标签说明+输出编码+少样本思维链（案例）+零样本思维链

对于基础性能优异的大语言模型，应当以实验组 11 进行配置并根据情况适当删除标签说明，即：角色指派+指令明确+输出编码+少样本案例+零样本思维链

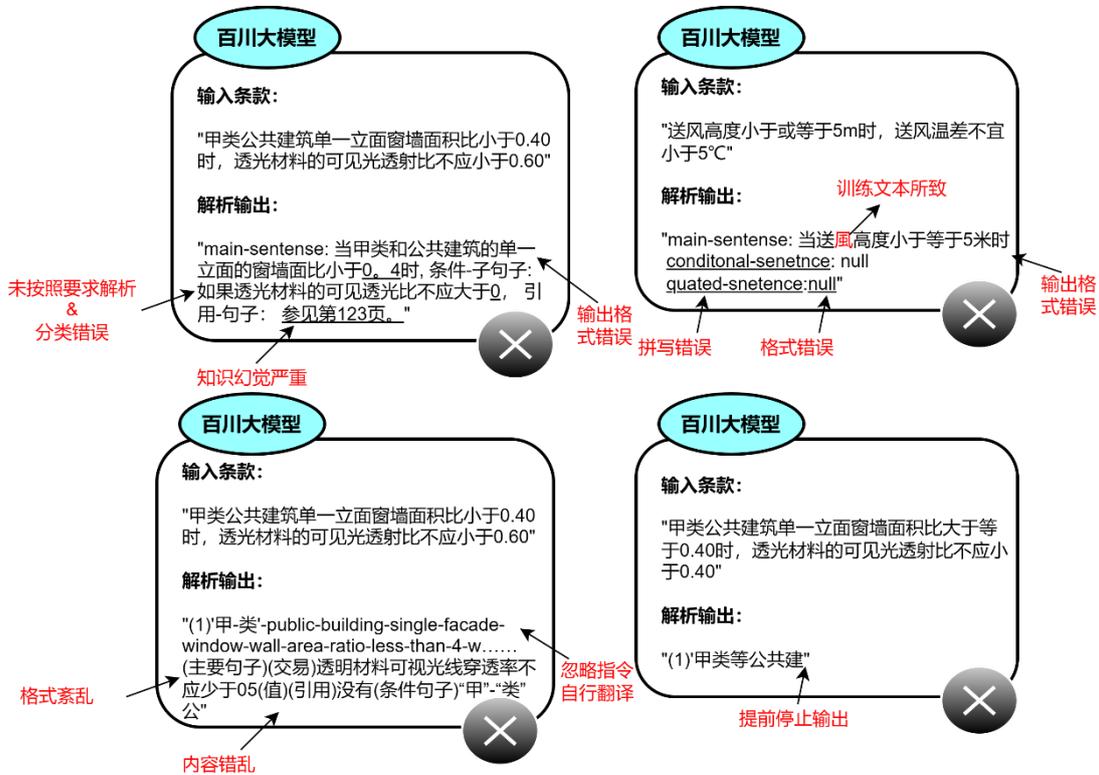


图 26 百川大模型主要错误案例

表 12 实验组 1-5 指标变化率

实验组	指令遵循%	格式遵循%	分类正确%	解析正确%	主观评分%
1	0.00	0.00	0.00	0.00	0.00
2	14.50	14.78	10.77	2.86	11.06
3	8.40	30.43	6.16	25.72	16.52
4	-8.40	-10.44	-11.53	-8.57	-11.56
5	-3.82	11.30	-3.85	16.20	-0.25

表 13 表现较差模型各实验组评价指标（百川、千帆）

实验组	指令遵循%	格式遵循%	分类正确%	解析正确%	主观评分%
1	25.0	2.5	25.0	2.5	16.0
2	62.5	20.0	52.5	7.5	40.25
3	42.5	50.0	40.0	37.5	46.25
4	37.5	2.5	37.5	2.5	23.5
5	5.0	7.5	5.0	5.0	6.5
6	47.5	45.0	42.5	40.0	45.0
7	0.0	0.0	0.0	0.0	0.0
8	52.5	50.0	52.5	45.0	52.5
9	10.0	10.0	10.0	10.0	9.0
10	35.0	7.5	35.0	7.5	24.0
11	57.5	57.5	57.5	52.5	57.0
12	50.0	0.0	50.0	0.0	40.0
13	52.5	42.5	50.0	40.0	43.5
14	50.0	50.0	47.5	47.5	49.0

15	45.0	50.0	42.5	42.5	44.0
16	55.0	55.0	55.0	47.5	54.0

表 14 表现良好模型各实验组评价指标（ChatGLM、ERNIE-Bot-3.5）

实验组	指令遵循%	格式遵循%	分类正确%	解析正确%	主观评分%
1	75.0	57.5	75.0	57.5	68.0
2	75.0	72.5	75.0	62.5	72.5
3	82.5	92.5	80.0	80.0	81.75
4	55.0	55.0	52.5	50.0	53.5
5	65.0	67.5	65.0	65.0	60.75
6	92.5	92.5	92.5	77.5	89.5
7	92.5	100.0	90.0	87.5	96.0
8	87.5	90.0	87.5	80.0	83.5
9	55.0	90.0	55.0	55.0	54.5
10	62.5	47.5	60.0	45.0	54.5
11	100.0	100.0	90.0	87.5	95.5
12	87.5	87.5	87.5	85.0	87.0
13	62.5	82.5	55.0	50.0	69.0
14	90.0	100.0	90.0	85.0	95.5
15	72.5	100.0	72.5	67.5	72.75
16	82.5	82.5	82.5	77.5	77.25

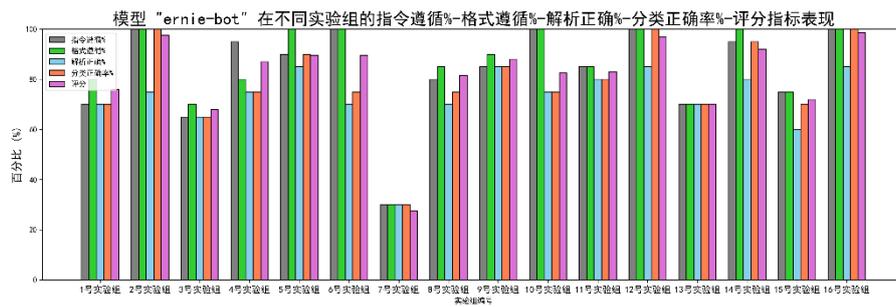
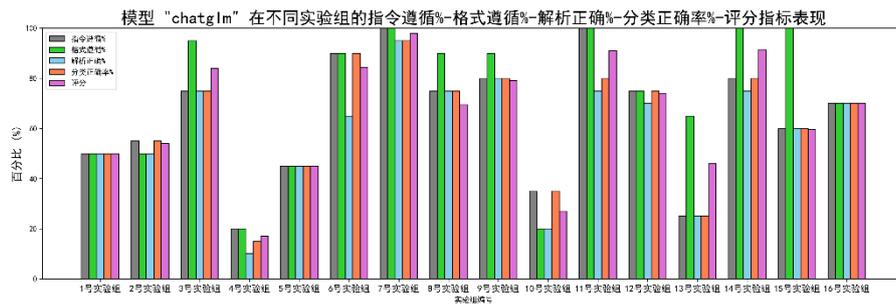
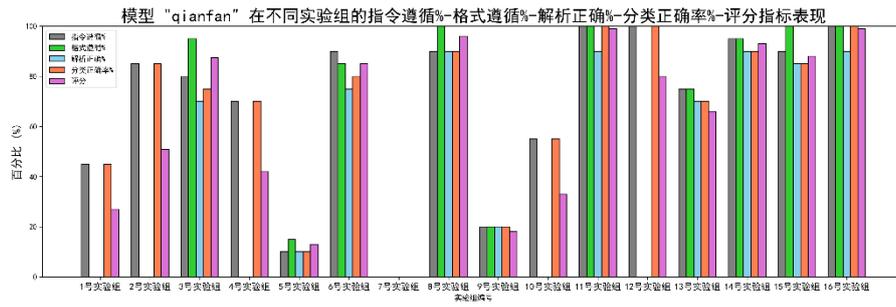
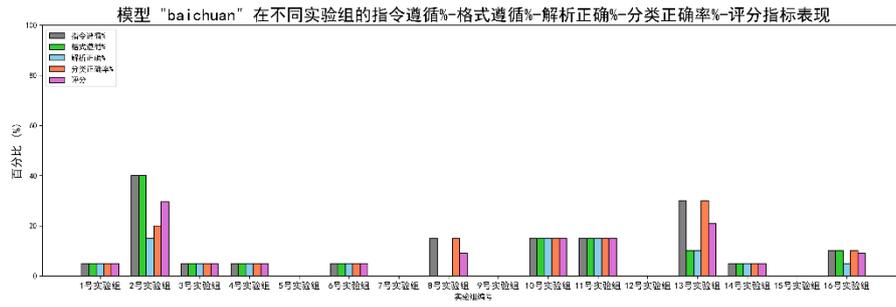
表 15 表现优异模型各实验组评价指标（GPT-3.5、GPT-4、文心 4、通义千问）

实验组	指令遵循%	格式遵循%	分类正确%	解析正确%	主观评分%
1	96.25	93.75	95.0	83.75	92.62

2	93.75	93.75	91.25	81.25	89.88
3	98.75	98.75	96.25	90.0	98.0
4	80.0	80.0	80.0	75.0	75.62
5	100.0	97.5	98.75	96.25	97.25
6	100.0	100.0	98.75	95.0	99.12
7	91.25	93.75	91.25	91.25	90.5
8	100.0	100.0	97.5	88.75	95.75
9	95.0	95.0	92.5	88.75	91.12
10	95.0	97.5	95.0	90.0	94.62
11	100.0	100.0	97.5	86.25	97.5
12	95.0	95.0	95.0	85.0	93.88
13	87.5	95.0	87.5	83.75	87.0
14	100.0	97.5	100.0	87.5	98.0
15	88.75	100.0	93.75	91.25	94.38
16	97.5	97.5	97.5	88.75	96.38

表 16 最佳提示策略组合

模型表现	代表模型	角色	指令	标签	输出编码	标签说明	少样本案例	零样本思维链	少样本思维链
一般	千帆	√	√		√		√		
良好	Ernie-Bot-3.5、 ChatGLM	√	√	√	√	√	√	√	
优秀	GPT-3.5、通义千问、GPT-4、文心4	√	√	√	√		√	√	



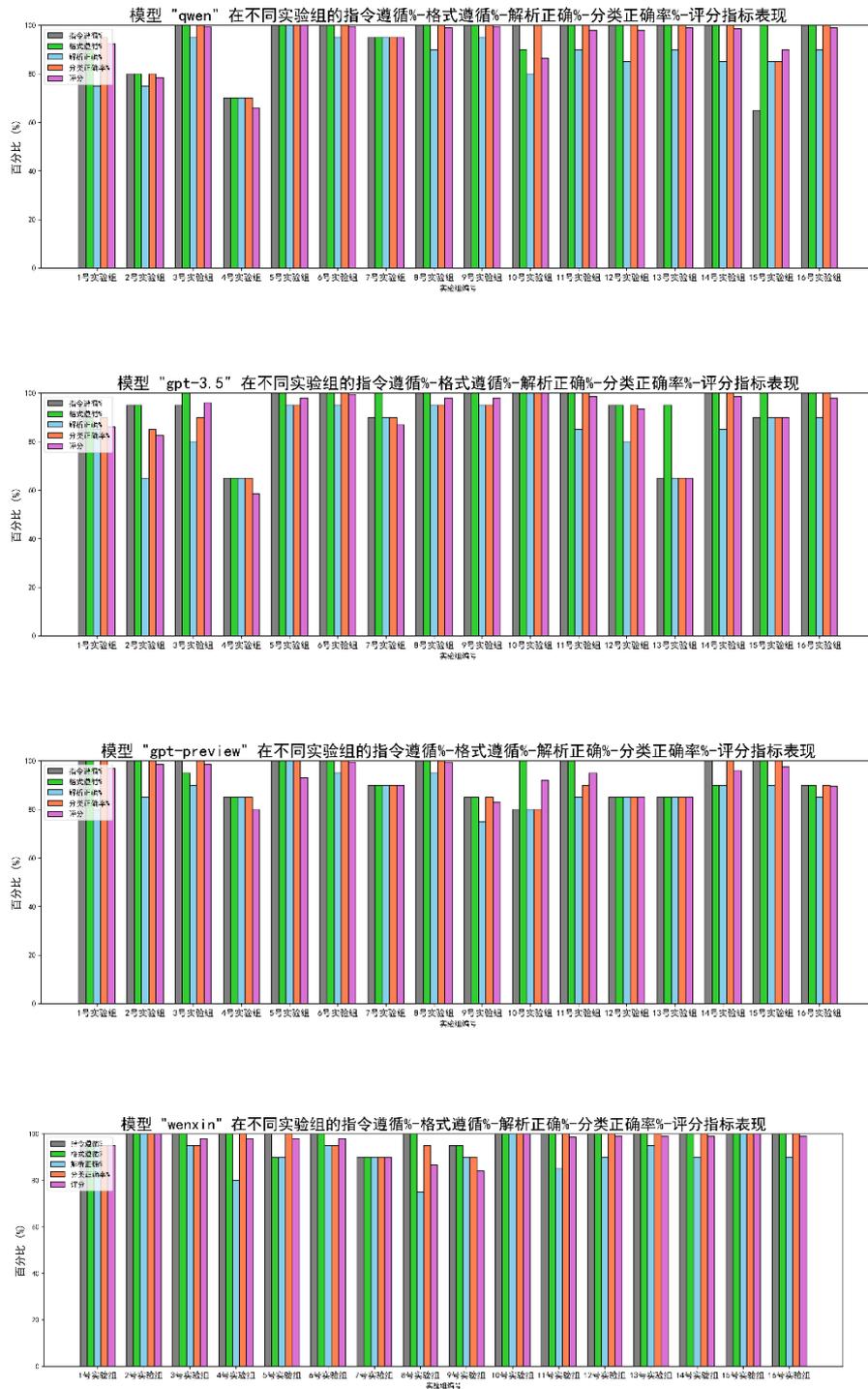


图 27 模型多实验组解析结果评价分数

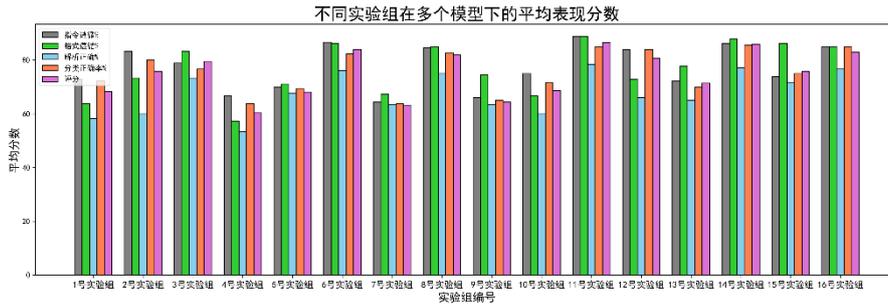


图 28 实验组多模型平均分数

## 4.5 本章小结

相对于以往基于具体任务所使用的预训练+微调方式，大语言模型因为“涌现”现象而出现的强大的上下文学习和逻辑推理能力使得模型不需要进行微调也能成功适配各类自然语言处理任务。然而，因为大语言模型是基于通用文本进行的训练，使得模型往往需要一定清晰的“提示”才能被激发其强大能力。除此以外，随着越来越多的大模型出现，社区对于不同模型在各类任务的应用开发和测评正在如雨后春笋一般涌现，本课题所涉及的中文建筑条款解析任务，也需要参与进去，得到表现优良的模型+提示策略组合。因此，本章首先介绍了一些常见的提示方法，拟定了一个提示词架构；之后，基于各类提示词部件功能，结合本课题的“条款解析”任务，撰写了此任务的提示词并为后续模型调用；最后，本研究使用当下主流且支持中文输出的大模型进行了模型、提示词组合实验，得到了多种提示词在此类任务中对模型表现的普适影响，同时本文在结论分析中给出了三类表现基准模型下不同的最佳提示词组合策略。其中表现最佳的文心4大模型在各实验组中的表现均十分优秀，在“指令遵循”、“格式遵循”、“分类正确”、“解析正确”和“主观评分”中分别获得了98.75%、98.13%、97.50%、91.25和96.38分的平均成绩，并在4类指标中获得最佳。因此，本文选择了表现最好的文心4+实验组11的提示策略作为条文解析的解决方案。同时该实验也证明了大语言模型在建筑自动合规性检查应用中的优越性。

## 第 5 章 基于设计说明的模型参数强化方法

### 5.1 概述

在暖通系统设计工作中，除了 BIM 模型以外，建筑设计说明的重要性不容小觑。因为它们为审图人员、施工队伍、运维团队和其他利益相关者提供了关于系统设计意图、性能要求和操作细节的关键信息。审图人员依赖设计说明来理解设计意图并确保设计方案遵循了相关的法规和标准。设计说明通常会详细说明系统的功能要求、性能标准和特殊的设计考虑。这些说明帮助审图人员识别可能在图纸上不明显的细节，如系统的能效目标、空间条件要求或者特定设备的性能参数。但是，由于 BIM 模型通常用于可视化和协调，而不是性能模拟，因此设计师在 BIM 模型中不会总是更新或修改所有性能参数，例如冷热源的 COP 或材料的热阻。这可能导致直接从模型中提取的参数与实际设计或施工说明中的参数不一致。所以对于自动合规性检查来说，从设计说明中补充设计参数是尤为重要的。

本章将从 3 个部分进行阐述：5.2 节介绍了基于 cnocr 开源库进行的图片转文字提取以及预处理过程，5.3 节阐述了根据第四章的实验结果形成最终提示结合预处理后的设计说明并使用当前表现最好的 GPT4 模型将设计说明中的参数提取成 JSON 格式，完成对模型设计参数的补全。

### 5.2 设计说明预处理

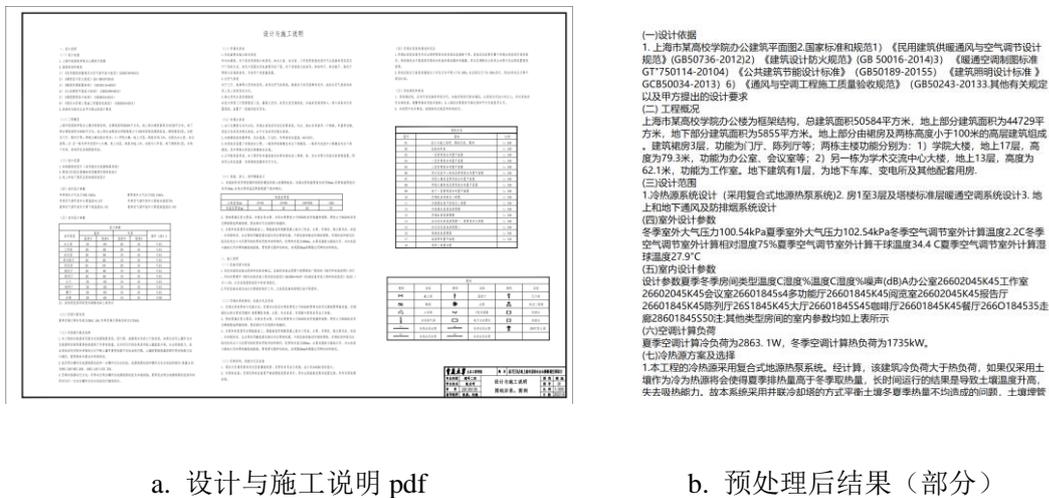
建筑施工设计说明通常由设计师在 Autodesk CAD 中完成编辑，输出 pdf 格式文件。因此相较于文字软件产出的 pdf 文件。设计说明中的文字不能够直接通过 pdf 文件中的 text 根目录进行直接读取，因为内部的文字是以图片的形式进行存储。所以，要对设计说明进行处理首先需要对图片内部的文字进行 OCR(Optical Character Recognition，简称 OCR) 识别。

本研究采用当下开源的 CnOCR 库来完成。CnOCR 是 Python 3 下的文字识别工具包，支持简体中文、繁体中文（部分模型）、英文和数字的常见字符识别，支持竖排文字的识别。自带了 20+ 个训练好的模型，适用于不同应用场景，如图 29。同时，CnOCR 也提供简单的训练命令供使用者训练自己的模型。



图 29 cnocr 库识别案例

针对已经提取的设计说明文字，该预处理步骤会做一些基于正则表达式的标题提取以和空格删除等标准化处理，包括删除多余空格以及根据每一条设计说明中的“中文序号”+“、”的文字标识进行查询，并在该条末尾添加“\n\n”双换行符。注意：文中的分隔符和多余空格对于能力较强的大语言模型来说影响甚微，但这样的预处理模式可以帮助产生中间文件并保存，方便设计者以及检查者进行检查和调试。图 30 展示了施工说明预处理前后效果。



a. 设计与施工说明 pdf

b. 预处理后结果（部分）

图 30 设计与施工说明预处理预览

### 5.3 提示词构建

因为从设计说明中提取设计参数本质上也是一个提取信息并结构化的过程，因此，本研究将采用 4.4 节中得出的文心 4 和提示策略 6 作为设计说明参数提取的模型策略组合参考，由于设计说明无法进行长段词性标注，故本研究最后选取

了“角色指派”、“指令明确”、“输出编码”、“少样本案例”和“零样本思维链”作为最终提示的组成。之后，将设计说明文本添加在最终提示文本中，打包输入文心 4 大模型，如图 30。

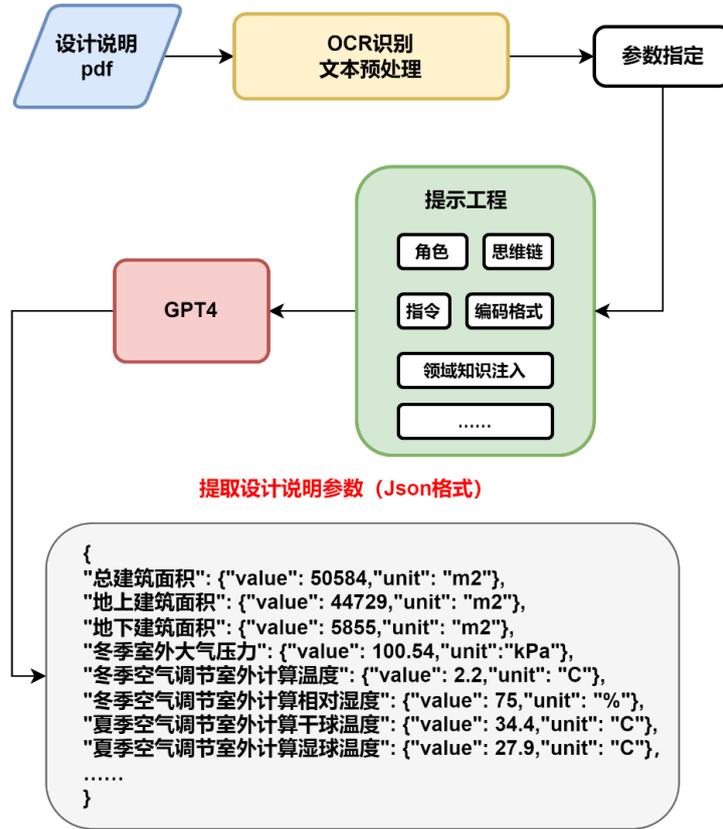


图 31 设计说明提取流程

以下是基于第 6 组提示策略形成的最终提示，完整提示详见附录 x：

\*\*\*\*角色\*\*\*\*

你是一位经验丰富的建筑领域暖通方向的审图专家。擅长从已经完成的设计文档和图纸中获取建筑信息和暖通系统设计相关信息。

\*\*\*\*指令说明\*\*\*\*

你的任务是从一份中文暖通系统'设计说明'中获取设计过程中的重要参数信息。并按照要求输出 JSON 文本。

注意!!! 该设计文档是从 png 图片通过 OCR 识别转化成的文本文档，其中的一些单位在图像识别过程中会出现错误，请你自行修正：

1. 'C 或 C 通常是指摄氏度，例如：12C，通常是指 12 摄氏度；
2. 6O 中的 O，通常是指数字 0 而不是字母 O；
3. ≤45 中的 ≤ 符号常常会被识别成 S 或者 K 等大写字母或者其他和数字

不相关的符号;

4. 有些参数来自于表格 ocr 的直接输出, 有些数字和列名可能会重叠在一起。请你自行分开;

\*\*\*\*输出编码\*\*\*\*

!!!! 请以 JSON 格式提取出建筑的信息参数!!!! 其中涉及到的应当参数包括: 1. 总建筑面积 2. 地上建筑面积 3. 地下建筑面积 4. 建筑层数 5. 建筑高度 6. 冬季室外大气压力 7. 冬季空气调节室外计算温度 8. 冬季空气调节室外计算相对湿度 9. 冬季空调计算热负荷 10. 夏季空气调节室外计算干球温度 11. 夏季空气调节室外计算湿球温度 12. 夏季空调计算冷负荷 13. 冷热源 a. 地源热泵 i. 台数 ii. 单机负荷 b. 螺杆式冷水机组 i. 台数 ii. 单机负荷 14. 风系统 a. 风机盘管+新风 i. 房间类型 ii. 送风方式 b. 全空气系统 i. 房间类型 ii. 送风方式 c. 分层空调系统 i. 房间类型 ii. 送风方式 15. 水系统 a. 闭式双管式系统 b. 夏季供水温度 c. 夏季回水温度 d. 冬季供水温度 e. 冬季回水温度 f. 控制方式: 主机定流量、负荷侧变流量 16. 保温 a. 保温材料 b. 水管保温层厚度 c. 风管保温层厚度

\*\*\*\*输入和输出案例\*\*\*\*

Input:(二) 工程概况

上海市某高校学院办公楼为框架结构, 总建筑面积 50584 平方米, 地上部分建筑面积为 44729 平方米, 地下部分建筑面积为 5855 平方米。地上部分由裙房及两栋高度小于 100 米的高层建筑组成。建筑裙房 3 层, 功能为门厅、陈列厅等; 两栋主楼功能分别为: 1) 学院大楼, 地上 17 层, 高度为 79.3 米, 功能为办公室、会议室等; 2) 另一栋为学术交流中心大楼, 地上 13 层, 高度为 62.1 米, 功能为工作室。

Output:[{'para':'总建筑面积','value':50584,'unit':'m²'},  
{'para':'地上部分建筑面积','value':44729,'unit':'m²'},  
{'para':'地下部分建筑面积','value':5855,'unit':'m²'},  
{'para':'学院大楼建筑高度','value':79.3,'unit':'m'},  
{'para':'学院大楼建筑层数','value':17,'unit':'层'},  
{'para':'学术交流中心大楼建筑高度','value':62.1,'unit':'m'},  
{'para':'学术交流中心大楼建筑层数','value':13,'unit':'层'}]

\*\*\*\*零样本思维链\*\*\*\*

让我们一步一步思考……

\*\*\*\*输入\*\*\*\*

一、设计说明

### (一)设计依据

1. 上海市某高校学院办公建筑平面图
2. 国家标准和规范
  - 1) 《民用建筑供暖通风与空气调节设计规范》(GB50736-2012)
  - 2) 《建筑设计防火规范》(GB 50016-2014)
  - 3) 《暖通空调制图标准》(GT"750114-2010)
  - 4) 《公共建筑节能设计标准》(GB50189-2015)
  - 5) 《建筑照明设计标准》(GCB50034-2013)
  - 6) 《通风与空调工程施工质量验收规范》(GB50243-2013)
3. 其他有关规定以及甲方提出的设计要求
2. ……

## 5.4 设计说明解析结果

输出的设计说明参数以 JSON 文件的形式储存在数据文件中，如图 32。本研究将 JSON 文件转化为 Dataframe 形式后如表 17 所示：

表 17 设计说明参数提取结果

参数名称	值	单位
总建筑面积	50584	m <sup>2</sup>
地上建筑面积	44729	m <sup>2</sup>
地下建筑面积	5855	m <sup>2</sup>
学院大楼层数	17	层
学术交流中心大楼层数	13	层
裙房层数	3	层
地下建筑层数	1	层
学院大楼建筑高度	79.3	m
学术交流中心大楼建筑高度	62.1	m
冬季室外大气压力	100.54	kPa
冬季空气调节室外计算温度	2.2	℃
冬季空气调节室外计算相对湿度	75	%
冬季空调计算热负荷	1735	kW

夏季空气调节室外计算干球温度	34.4	℃
夏季空气调节室外计算湿球温度	27.9	℃
夏季空调计算冷负荷	2863.1	kW
地源热泵台数	2	台
地源热泵单机负荷(制冷)	893.2	kW
地源热泵单机负荷(制热)	803.2	kW
螺杆式冷水机组台数	1	台
螺杆式冷水机组单机负荷(制冷)	1605.1	kW
螺杆式冷水机组单机负荷(制热)	1455.7	kW
风机盘管+新风房间类型	办公室、会议室、工作室	/
风机盘管+新风送风方式	双层百叶下送	/
全空气系统房间类型	门厅、画廊	/
全空气系统送风方式	上送上回	/
分层空调系统房间类型	学院门厅	/
分层空调系统送风方式	球形喷口	/
水系统类型	闭式双管式系统	/
水系统夏季供水温度	7	℃
水系统夏季回水温度	12	℃
水系统冬季供水温度	40	℃
水系统冬季回水温度	55	℃
水系统控制方式	主机定流量、负荷侧变流量	/
保温材料	带玻璃布铝箔防潮层的离心玻璃棉板材	/
水管保温层厚度	20	mm
风管保温层厚度	30	mm

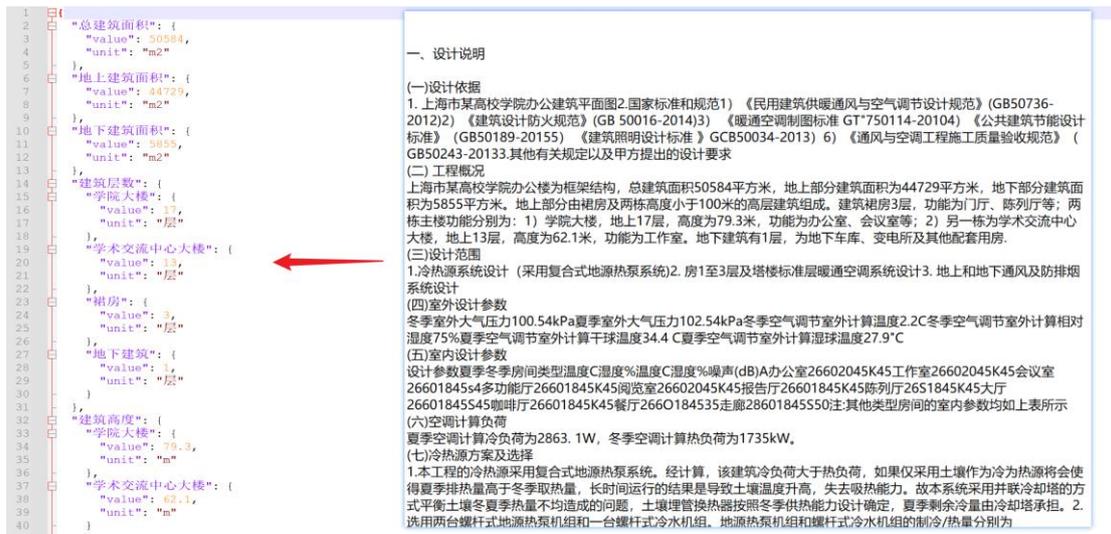


图 32 设计文本参数提取结果 JSON

## 5.5 本章小结

当下设计师在建 BIM 模型时, 往往采用设计院内部预置族库并做部分修改后直接加入现有的工程文件中。建筑负荷计算、暖通设备选型以及管路水力计算等设计步骤往往是和模型创建分开的。这会导致最后模型文件的参数和真实工程中所使用的设计参数出现偏差; 除此之外, 在规范中常被提及的“窗墙比”、“体形系数”等参数也一般不会被记录在模型文件中, 而是被统一记录在《设计和施工说明》中。为解决这个问题, 本章首先使用了 OCR 库将 CAD 输出的《设计与施工说明》pdf 文件转化成.txt 文件; 再构建基于参数提取任务的模型提示, 最后和设计说明文本一起输入到第四章表现最好的文心 4 大模型中完成模型参数的补充和强化。该任务在流程上较为简单, 但对于建筑自动合规性检查来说也是不可或缺的一环。同时也预示着在应用建筑合规性检查系统时, 设计说明在整个工程项目中的重要性将被大大提升。



## 第 6 章 自动合规性检查代码生成方法

### 6.1 概述

模型检查本质是一个交互核对的工作，通过生成条款中涉及到的建筑构件、值、单位等语义元素和实际建筑模型实例、属性之间的映射关系，从而实现条款逻辑到自动检查逻辑的转换。

本研究将利用现阶段设计院广泛应用的 BIM 软件 Autodesk Revit 作为模型检查平台，并利用 Revit 自带的 Model Checker 作为执行条款规定的实现工具，从而实现第四章条款解析后的模型自动检查工作。

本章首先通过编码格式、交互原理介绍 Revit Model Checker 建筑模型自动检查工具的应用原理。其次，介绍如何利用大语言模型和提示工程对现有的建筑参数进行强化，从而扩大系统自动检查的范围。之后，本章介绍了系统规则生成方法，将结构化的检查条款转化为可运行的 xml 格式自动检查代码。最后，基于 Revit Model Checker 的检查报表结果作为母版，生成考虑设计说明参数的建筑模型最终检查报表。

### 6.2 软件：Revit Model Checker

Revit Model Checker 是 Autodesk 提供的一款免费工具，旨在根据一系列建筑信息模型（BIM）要求自动检查 Revit 模型，此工具有助于验证模型是否符合 BIM 要求。软件提供了数十个预先构建和测试的检查集，可使用预定义检查数据集或自定义检查数据集，并允许用户在 Revit 模型中直接运行报告。同时，用户随时可以直接从报告结果中放大和选择不合规元素并执行相应操作。软件配有 Revit Model Checker Configurator 作为辅助工具，允许为 Autodesk Model Checker 创建自定义检查集。

Revit Model Checker 的输入执行代码通过 XML 格式进行定义。基本的 XML（可扩展标记语言）命名规则和一个 XML 树的结构。在图的上方列出了节点的定义，节点是 XML 结构中的任意项。如图 33，在基本层面上，XML 代码是通过以下节点定义的：

- 文档（Document）：描述代码信息的声明，如 XML 版本和字符类型。
- 根（Root）：XML 树中的最顶层元素。它是其余元素节点的“父节点”。

- 元素 (Element): 是根的“子”组件，以开始标签开始并以匹配的结束标签结束。标签之间的字符是元素的内容，并可能包含其他的“子”元素。元素也可以包含一个空的元素标签。
- 属性 (Attribute): 包含一个名称-值对的标记构造，存在于开始标签或空元素标签中。
- 文本 (Text): 在元素内部引号中定义的值。

在 Revit Model Checker 的输入检查集中，以上节点元素的呈现为：

Document		<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
Root	Start-Tag	<MCSettings>
Element	Start-Tag	<Heading>
		<Section>
		< Section>
		<Check>
	Start-Tag / End Tag	<Filter.../>
		<Filter.../>
		<Filter.../>
	End-Tag	</Check>
		</Section>
		</Section>
</Heading>		
Root	End-Tag	</MCSettings>

图 33 Revit Model Checker 输入 XML 文件样式

在 XML 文件中，可以看出该软件是由多个检查构成的集合，一个检查是由多个在 Revit 中被广泛使用的过滤器 (Filter) 构成的集合，允许用户按照特定的参数和属性来进行数据筛选和检查。对于一个过滤器来说，其内部结构为：

- 操作符 (Operator): 包括“与”、“或”和“排除”，它们分别用于组合其他过滤器验证条件、独立于其他过滤器验证条件以及从其他过滤器中排除某些条件。
- 类别 (Category): 涵盖了 Revit API 参数、类型、类别、设计选项、族等多个方面，如指定 Revit API 中定义的属性、类别、参数名和值等。
- 属性 (Property): 部分列出了一些特定的属性，如“标高”、“全类名”、“是否定义”、“是否是元素类型”等，它们用于描述元素的特定信息。
- 条件 (Condition): 定义了如何比较属性值，例如“包含”、“不等于”、“大于”、“小于”、“无值”、“重复”、“匹配正则表达式”等，用于进一步细化搜索或检查条件。

- **值 (Value):** 可以是“真”或“假”，或者是特定的字段值。
- 还包括了对大小写敏感性、单位（如英寸、英尺、厘米、米）和单位类别的定义，以及用户自定义值的设置。

如图 34，为 Revit Model Checker Configurator 中的检查集视角和单个检查的过滤器视角：

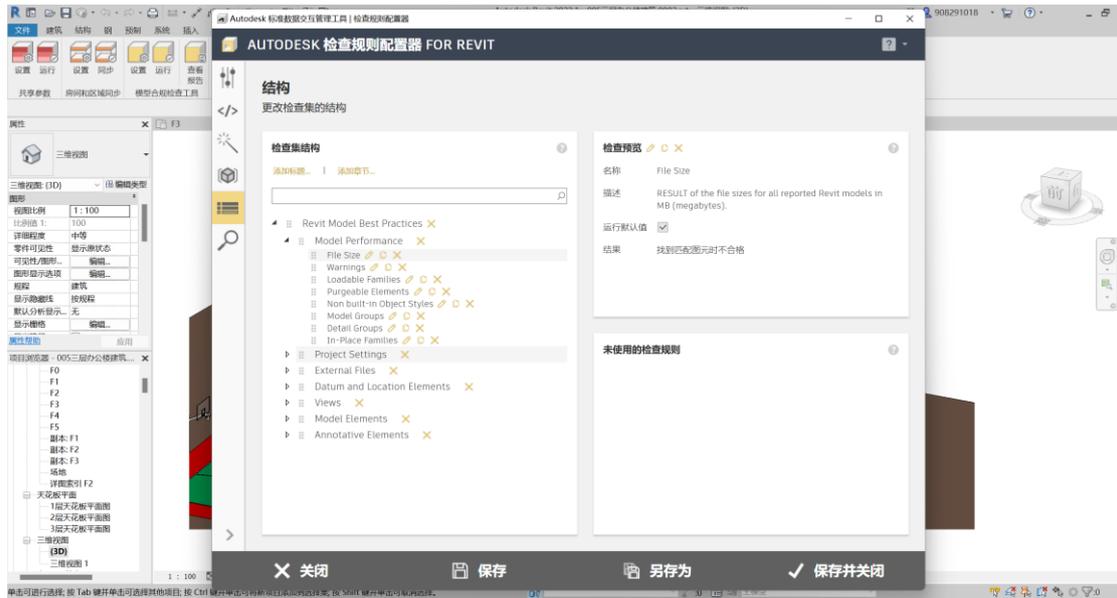


图 34 检查集视角

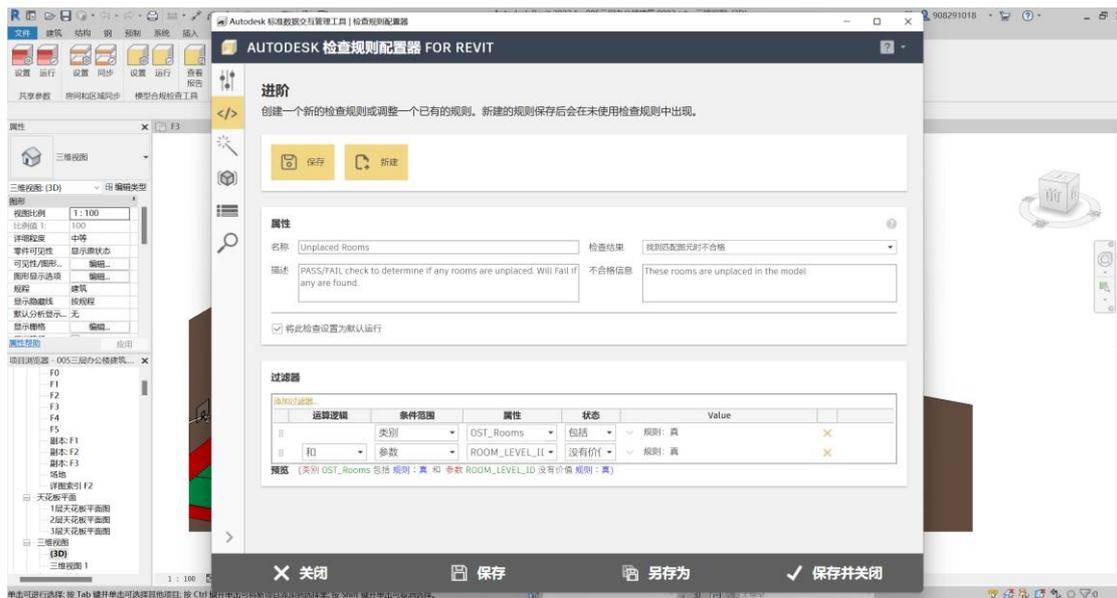


图 35 过滤器视角

### 6.3 自动检查代码生成

检查代码生成的目的是将条款的结构性表达转化为软件特定的 xml 所输入的可执行代码。如图 36，本流程将分为两个步骤：

- 语义元素映射：将条款子句中的语义元素转化为软件可识别的 XML Checkeset 对象属性名称。如“窗”、“门”、“小于”等建筑构建元素或是比较词可以转化为“OST\_WINDOW”、“OST\_DOORS”、“<”等对象。
- 子句结构转换：将条款中的多个子句按照“条件-主句-引用”的形式进行排列执行过滤操作。

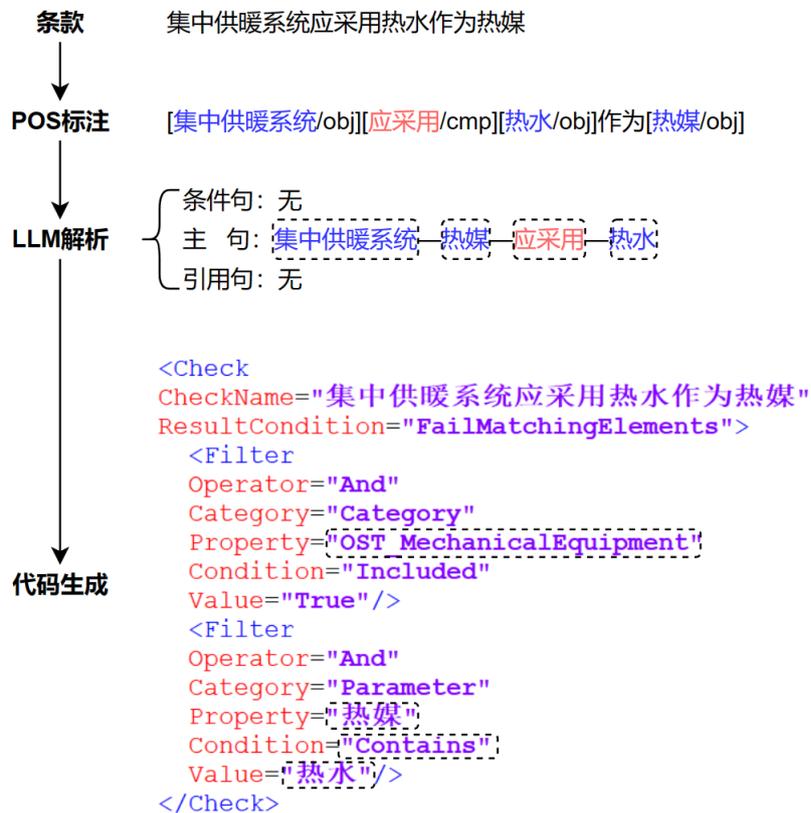


图 36 检查代码生成说明

本研究中所涉及到的语义元素则较为集中。因为建筑规范对于建筑构建种类的划分较为明确，如“门”、“窗”、“柱”、“梁”等。Revit Doc 曾对于用于 Revit 中的建筑类别和参数进行了说明，对应“BuiltInCategory”以及“BuiltInParameter”。

详情可参考 Revit Doc 网站 (<https://www.revitapidocs.com/>)。这两个文档的本质是将 Revit 原生内嵌的建筑构建进行详尽的枚举，以使用户进行二次开发调用。同样，用户可以通过自创类别或是参数赋予模型，软件同样可以通过用户定义的关键字段筛选到具体的建筑构建。本研究将采用部分模型对象和正则表达式 (Regular Expression, 简称 Regex 或 RegExp) 进行建筑规范自动检查的元素映射和演示。正则表达式是一种用来匹配字符串的模式。它是由字符和特殊字符组成的字符串, 用于描述字符的组合、位置和数量, 从而实现对文本的搜索、替换、提取等操作。

在表 18 中展示了本研究所采用的部分正则表达规则所生成的元素映射关系。其中, 第一列表达的是用于匹配解析结果中语义元素的正则表达式, 第二列是该元素对应 XML Checkset 中对象名称。例如: 对于“窗”在条款中的表达, 可能会以“窗”、“窗户”、“透光结构”……等形式出现, 这些语义元素都对应了 Model Checker 中的“OST\_WINDOWS”这一对象。所以可以通过正则表达式“窗\$|^窗|^透光”对条款元素进行筛选。其中: “\$”表示匹配字符串的结尾, 它用于指定模式的结束位置, 涵盖了“某某窗”这类语义元素; “|”表示逻辑上的“或”关系, 用于匹配多个模式中的任意一个, 在这个例子中, 存在三个匹配模式; “^”, 表示匹配字符串的开头, 它指示模式应该从字符串的开头开始匹配, 涵盖了“窗户”、“窗面”、“透光结构”、“透光材料”等语义元素。

表 18 Revit Model Checker 模型对象映射表

类别/参数名	正则表达式	说明
OST_Doors	门\$	门
OST_Stairs	楼梯\$	楼梯
OST_Windows	窗\$ ^窗 ^透光材料	窗、幕墙
OST_Walls	墙\$ ^墙	墙
OST_Floors	楼板\$	楼板
OST_MassShade	遮阳\$	遮阳
OST_Sheets	(甲 乙)?类 严寒 夏热冬(冷 暖)?温和	气候区

OST_Rooms	(房 空)?间	房间
OST_DuctTerminal	(送 回)?风	风管、风口
OST_MechanicalEquipment	散热器 锅炉 供暖系统	设备、系统

对于子句结构转化部分，对应 XML 文件中的 Filter 顺序。如图 35，在条款“集中供暖系统应采用热水作为热媒”中，因条款较为简洁，经解析得到了单一的主句元素连接。经正则表达映射得到其中“集中供暖系统”作为 obj 映射到软件自带的“OST\_MechanicalEquipment”类作为被检查对象，通过第一个过滤器过滤得到相关设备，并通过第二个过滤器将存在“热媒”属性的设备筛选出来，注意这里的“热媒”是研究人员在 Revit 模型中自定义的模型参数，这是因为该参数在 Revit 模型所提供的基础设备模型中没有被预先定义。为了更加贴切本研究所依赖的中文属性，本研究在遇到这种情况时一般均使用自定义参数。“应采用”作为比较词根据预设字典映射成为“Contains”，表 19 记录了所有比较词的映射关系。

当条款所定义构件无法在 Revit API 中找到时，如：窗墙比、模型没有定义的保温层厚度、进出水温度等参数。代码将自动检查从设计说明提取出的设计参数，且因为设计师在使用 Revit 时通常会使用固定的模板而忽略其中参数的修改，所以本研究将保持设计说明优先级高于 BIM 模型中的参数，以确保尽可能保证设计参数的准确性。如果条款元素依然没有被找到，该条款会被标记为人工介入，最后记录下该条款的语义标签、比较词高亮（属性型、间接型）和检查结果生成最终报表。

表 19 比较词映射字典

符号	正则表达式	说明
≤	小于或?等于 不(大于 高于 多于 超过)	小于等于
≥	大于或?等于 不(小于 低于 少于)	大于等于
>	大于 超过 高于	大于
<	小于 低于	小于

≠	不等于 避免 (不 无法 严禁)采用 非 除	不等于
=	等于 为 采?用 按照? 符合 执行	等于
has no	(不 严禁)(有 设置? 具备) 无	不含
has	(具 含 留)?有 设(置 有)? (增 辅 搭)设 安装 具备 采取	包含
not in	(不 除 严禁)在	不在
in	在	在

## 6.4 本章小结

自动合规性检查的自动体现在自动生成条款逻辑,并根据逻辑对模型进行循环检查。但即使生成了结构化的逻辑文本,其本质也是自然语言,并不能让计算机运行。因此,本章的目的就是在执行检查前,将结构化的条款表达转化为机器可运行的代码。首先,本章首先对比了执行模型检查的常见方法,选择并介绍了 Autodesk 开发的 Revit Model Checker 商用软件;然后对软件所需输入的 XML 文件格式进行了描述;接着,依照第四章所输出的条款结构化表达、输入文件格式和 Revit API 所提供的元素映射关系,使用正则表达式的方式完成条款元素和模型构件对齐,生成执行代码文件;最后本研究考虑到检查自动化检查过程的可解释性,介绍了在原插件所产出的基础上对条款分析过程进行覆写并导出了最终报表的方法。



## 第 7 章 BIM 模型规范自动审查验证与分析

### 7.1 概述

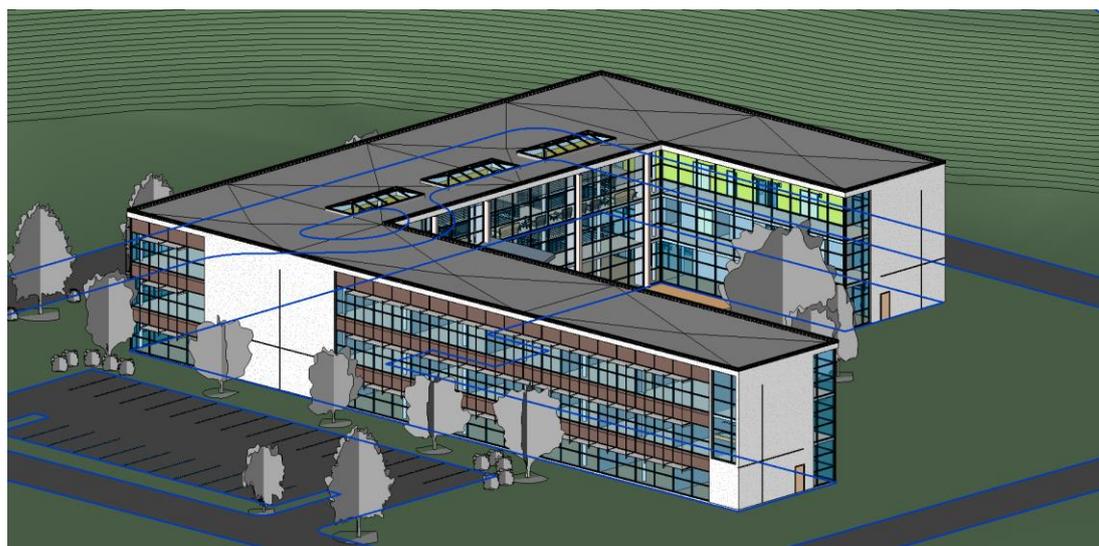
本章将选取 3 栋不同体量的 Revit 建筑模型及设计说明、验证数据集进行语义解析后的自动合规性检查应用的验证。如图 37，验证的流程包括：首先通过条款分类模型将输入条款基于是否能够进行自动检查进行分类，接着对能够自动检查的条款完成自动语义元素标注，然后根据第四章实验结果所得到的最佳提示和模型组合（文心 4 大模型搭配第 11 组提示策略）接入大语言模型完成条款解析，同时输入对应建筑设计说明完成建筑模型参数的并行提取。最后根据结构化解析结果完成 Revit Model Checker 自动检查代码生成，带入 BIM 模型运行 Revit Model Checker 生成最终检查报表。

其中验证建筑模型包括：

案例 1：假定该建筑位于上海市，此模型为 Revit 2022 自带三层办公建筑工程案例。如图 37，建筑面积约 5121.26 m<sup>2</sup>，层高 3.8m。建筑共三层，主要功能房间为办公室、会议室、休息区等。

案例 2：假定该商业建筑位于上海市，建筑面积 8872.17 m<sup>2</sup>，体形系数为 0.23，主体建筑为框架结构。建筑共六层，主要功能为办公室、会议室、宿舍、食堂等。

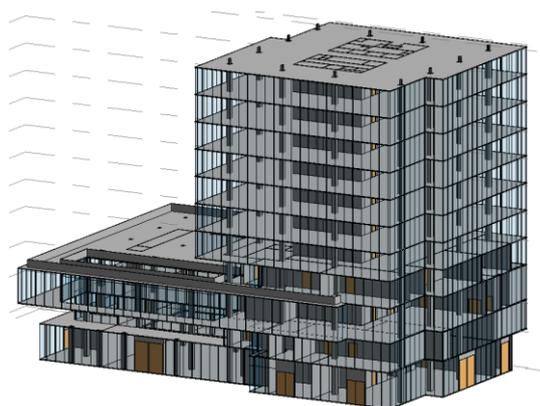
案例 3：假定该商业建筑位于上海市，建筑面积 16805.43 m<sup>2</sup>，体形系数为 0.15，主体建筑为框架结构。建筑共十一层，主要功能为办公室、会议室、礼堂、控制室、等待区等。



a 案例 1



b 案例 2



c 案例 3

图 37 验证案例总览

## 7.2 规范解析结果

着眼于普遍适用在绝大多数 BIM 模型中涉及到的建筑暖通模型参数，如：围护结构传热系数、窗户透射系数、设备选型信息等。并考虑到条款所涉及语义元素和 Revit Model Checker 插件工具中 Categories 和 Parameters 在 Revit API 文档中所涉及到的建筑构件参数对齐。表 20 展示了所选取的规范验证数据集中十

五个条款。

在条款分类后，条款 1-5 被正确归为了 2 条专家型条款和 3 条分类型条款，其中专家型条款一条出自总则，一条出自于第三章规范条款；定义型条款两条出自术语章，一条出自第三章规范条款。这两种类型并不存在具体的数值比较，需要专家最后的人工介入，所以不需要进行后续的条款解析处理。条款 6-15 中有 9 条属性型条款，1 条间接型条款，均被正确分类。在后续的处理中，被输入词性标注模型进行语义打标，并在第 13 条款中出现了一个标签错误。该条款原文是“210 层以下建筑外窗的气密性不应低于 6 级”，标注结果显示其中的“以下”是作为“ctg”即“条件触发标签”，但却理解成了“cmp”即“比较”谓词的属性，这可能和模型训练过程中此类词语的训练标签较为集中导致模型在这种场景下欠拟合导致；除此之外，该模型把“[建筑外窗/obj]”错误地分成了“[建筑/obj][外窗/obj]”进而导致了该条款后续的解析错误。

在条款解析环节，10 个条款中有 9 个解析正确，同样错误出现在了第 13 条款，原因在于模型对条款原意进行了曲解。原文表达的意思是在建筑 210 层以下部分的建筑外窗需要满足一定气密性要求，但是解析后该条款被分割成了条件句和主句，其内涵被曲解成了当满足建筑层数低于 210 层时，建筑外窗的气密性应当满足一定要求。出现这个问题的原因可能时由于“ctg”的标签给大模型带来了一定的误导，因而导致大语言模型认为此类条款存在条件句，进而认为 210 层是用来形容建筑而非具体构建（在这里是建筑外窗），最终导致了解析错误。

表 20 测试条款分类、标注结果

序号	说明	内容	分类	评价
1	原文	为贯彻国家有关法律法规和方针政策，改善公共建筑的室内环境，提高能源利用效率，促进可再生能源的建筑应用，降低建筑能耗，制定本标准。	专家型	正确
2	原文	建筑体形系数 <b>shape factor</b> 建筑物与室外空气直接接触的外表面积与其所包围的体积的比值，外表面积不包括地面和不供暖楼梯间内墙的面积。	定义型	正确
3	原文	参照建筑 <b>reference building</b> 进行围护结构热工性能权衡判断时，作为计算满足标准要求的全年供暖和空气调节能耗用的基准建筑。	定义型	正确
4	原文	独栋建筑面积小于或等于 300 m <sup>2</sup> 的建筑，应为乙类公共建筑。	定义型	正确

5	原文	建筑总平面设计及平面布置应合理确定能源设备机房的位置，缩短能源供应输送距离。同一公共建筑的冷热源机房宜位于或靠近冷热负荷中心位置集中设置。	专家型	正确
6	原文	严寒地区甲类公共建筑各单一立面窗墙面积比(包括透光幕墙)均不宜大于 0.60	属性型	正确
	标注	[严寒地区/obj][甲类/obj][公共建筑/obj]各[单一立面/obj][窗墙面积比/obj](包括透光幕墙)均[不宜大于/cmp][0.60/value]		
	解析	<b>主句:</b> [严寒地区-甲类-公共建筑-各单一立面-窗墙面积比-均不宜大于-0.60], <b>条件句:</b> [null], <b>引用句:</b> [null]		
7	原文	甲类公共建筑单一立面窗墙面积比小于 0.40 时，透光材料的可见光透射比不应小于 0.60	属性型	正确
	标注	[甲类/obj][公共建筑/obj][单一立面/obj][窗墙面积比/obj][小于/cmp][0.40/value][时/ctg], [透光材料/obj]的[可见光透射比/obj][不应小于/cmp][0.60/value]		
	解析	<b>主句:</b> [透光材料-可见光透射比-不应小于-0.60], <b>条件句:</b> [甲类-公共建筑-单一立面-窗墙面积比-小于-0.40], <b>引用句:</b> [null]		
8	原文	甲类公共建筑单一立面窗墙面积比大于等于 0.40 时，透光材料的可见光透射比不应小于 0.40	属性型	正确
	标注	[甲类/obj][公共建筑/obj][单一立面/obj][窗墙面积比/obj][大于等于/cmp][0.40/value][时/ctg], [透光材料/obj]的[可见光透射比/obj][不应小于/cmp][0.40/value]		
	解析	<b>主句:</b> [透光材料-可见光透射比-不应小于-0.40], <b>条件句:</b> [甲类-公共建筑-单一立面-窗墙面积比-大于等于-0.40], <b>引用句:</b> [null]		
9	原文	乙类公共建筑外窗有效通风换气面积不宜小于窗面积的 30%	间接型	正确
	标注	[乙类/obj][公共建筑/obj][外窗/obj][有效通风换气面积/obj][不宜小于/cmp][窗面积/obj]的[30%/value]		

	解析	<b>主句:</b> [乙类-公共建筑-外窗-有效通风换气面积-不宜小于- quote], <b>条件句:</b> [null], <b>引用句:</b> [窗面积-30%]		
10	原文	散热器宜明装;	属性型	正确
	标注	[散热器/obj][宜/cmp][明装/obj];		
	解析	<b>主句:</b> [散热器-宜-明装], <b>条件句:</b> [null], <b>引用句:</b> [null]		
11	原文	送风高度小于或等于 5m 时, 送风温差不宜小于 5°C	属性型	正确
	标注	[送风高度/obj][小于或等于/cmp][5m/value][时/ctg], [送风 温差/obj][不宜小于/cmp][5°C/ value]		
	解析	<b>主句:</b> [送风温差-不宜小于-5°C], <b>条件句:</b> [送风高度-小于或等于-5m], <b>引用句:</b> [null]		
12	原文	在人员长期停留的房间, 其顶棚内表面可见光反射比应当 小于等于 0.9 且大于等于 0.7	属性型	正确
	标注	[在/ctg]人员长期停留的[房间/obj], [其/obj][顶棚/obj][内表 面可见光反射比/obj][应当小于等于/cmp][0.9/value]且[大于 等于/cmp][0.7/value]		
	解析	<b>主句:</b> [房间-顶棚-内表面-可见光反射比-(应当小于等于- 0.9)&(应当大于等于-0.7)], <b>条件句:</b> [null], <b>引用句:</b> [null]		
13	原文	210 层以下建筑外窗的气密性不应低于 6 级;	属性型	错误
	标注	[210 层/value][以下/ctg][建筑/obj][外窗/obj]的[气密性 /obj][不应低于/cmp][6 级/value]		
	解析	<b>主句:</b> [外窗-气密性-不应低于-6-级], <b>条件句:</b> [建筑-低于-210-层],		

		<b>引用句:</b> [null]		
	原文	耐火等级为三级的公共建筑，其柱的耐火极限不低于 2h		
	标注	[耐火等级/obj][为/cmp][三级/value]的[公共建筑/obj]，[其/obj][柱/obj]的[耐火极限/obj][不低于/cmp][2h/value]		
14	解析	<b>主句:</b> [公共建筑-柱-耐火极限-不低于-2-h], <b>条件句:</b> [公共建筑-耐火等级-为-三-级], <b>引用句:</b> [null]	属性型	正确
	原文	集中供暖系统应采用热水作为热媒		
	标注	[集中供暖系统/obj][应采用/cmp] [热水/obj]作为[热媒/obj]		
15	解析	<b>主句:</b> [集中供暖系统-热媒-应采用-热水], <b>条件句:</b> [null], <b>引用句:</b> [null]	属性型	正确

### 7.3 自动审查结果

基于上节解析完成的条款，生成 Revit Model Checker 插件所需 XML 文件。导入到插件后在 Revit 标准交互数据管理工具中运行建筑 BIM 模型案例。导出软件自带检查报表后，再次通过 Python 环境读取基础报表，从而得以写入条款分类、词性打标以及突出比较谓词，最后通过 Pandas 库生成最终报表，如图 38-41。

```

1 <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2 <CheckSettings Allowed="True" Name="CheckFinal" Author="MrKey" Description="This is the final demo of the project" Image="" LastModified="
3 "133538558739839281">
4 <Heading ID="bd242dcd7-4c55-4bcf-b13a-9ef21152aa89" HeadingText=""&lt;!--Unused--&gt;" Description="Checks that have been created but not placed in
5 the structure yet." IsChecked="True">
6 <Section ID="9a189d47-6ee7-44e9-9413-70c9ba599e4a" SectionName=""&lt;!--Unused--&gt;" Title="Checks that have not been placed" IsChecked="True"
7 Description="" />
8 </Heading>
9 <Heading ID="95505ec1-9564-4625-b875-bb2ded2311bf" HeadingText="test" Description="" IsChecked="True">
10 <Section ID="07fad44f-23da-4dd0-9fe2-95df2e250b1" SectionName="validation" Title="" IsChecked="True" Description="">
11 <Check ID="4731f9b5-41c1-4ba6-bce2-1d7b6624791d" CheckName=""&lt;!--Unused--&gt;" Description="严寒地区甲类公共建筑各单一立面窗墙面积比(包括透光幕墙)均不宜大于0.60" Description="
12 "" FailureMessage="" ResultCondition="FailMatchingElements" CheckType="Custom" IsRequired="False" IsChecked="True">
13 <Filter ID="1f581b9e-5247-4d53-bb50-921bb1349ca" Operator="And" Category="Category" Property="OST_Walls" Condition="Included" Value="True"
14 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
15 <Filter ID="2ebc70a-d593-4e2a-92a4-1f7ee6a87821" Operator="And" Category="Parameter" Property="窗墙比" Condition="GreaterThan" Value="0.6"
16 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
17 </Check>
18 <Check ID="62511484-43dd-40cb-91e8-e92076981d46" CheckName="甲类公共建筑单一立面窗墙面积小于0.40时,透光材料的可见光透射比不应小于0.60"
19 Description="" FailureMessage="" ResultCondition="FailMatchingElements" CheckType="Custom" IsRequired="False" IsChecked="True">
20 <Filter ID="7df445df-345b-4bea-a578-965f227a5ac0" Operator="And" Category="Category" Property="OST_Walls" Condition="Included" Value="True"
21 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
22 <Filter ID="49006877-e4d8-4344-84f7-3df1c5115bbd" Operator="And" Category="Parameter" Property="窗墙比" Condition="LessThan" Value="0.4"
23 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
24 <Filter ID="5922427d-1191-402d-950c-8069039524b6" Operator="And" Category="Category" Property="OST_Windows" Condition="Included" Value="True"
25 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
26 <Filter ID="2be06705-30b4-4e1b-bd86-03d628f09eef" Operator="And" Category="Parameter" Property="ANALYTICAL_VISUAL_LIGHT_TRANSMITTANCE"
27 Condition="LessThan" Value="0.6" CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
28 </Check>
29 <Check ID="3f4d0ba0-b76e-4db9-abd7-9f4356aad427" CheckName="甲类公共建筑单一立面窗墙面积大于等于0.40时,透光材料的可见光透射比不应小于0.40"
30 Description="" FailureMessage="" ResultCondition="FailMatchingElements" CheckType="Custom" IsRequired="False" IsChecked="True">
31 <Filter ID="a2990758-453a-4981-bef4-5538e37a118e" Operator="And" Category="Category" Property="OST_Walls" Condition="Included" Value="True"
32 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
33 <Filter ID="b33d0fa23-9b98-4ef2-8c5d-ee08ac55e8f5" Operator="And" Category="Parameter" Property="窗墙比" Condition="GreaterOrEqual" Value="0.4"
34 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
35 <Filter ID="7df5baa7-0bbe-40e4-898a-d1eeb99c7c5" Operator="And" Category="Category" Property="OST_Windows" Condition="Included" Value="True"
36 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
37 <Filter ID="1cec66e1-2392-4213-b7f7-5bfac5ed553a" Operator="And" Category="Parameter" Property="ANALYTICAL_VISUAL_LIGHT_TRANSMITTANCE"
38 Condition="LessThan" Value="0.4" CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
39 </Check>
40 <Check ID="f5205a27-1e4c-4775-81e5-642ea474b1b7" CheckName="乙类公共建筑外窗有效通风换气面积不宜小于窗面积的30%" Description="" FailureMessage=""

```

图 38 自动检查 XML 文件生成

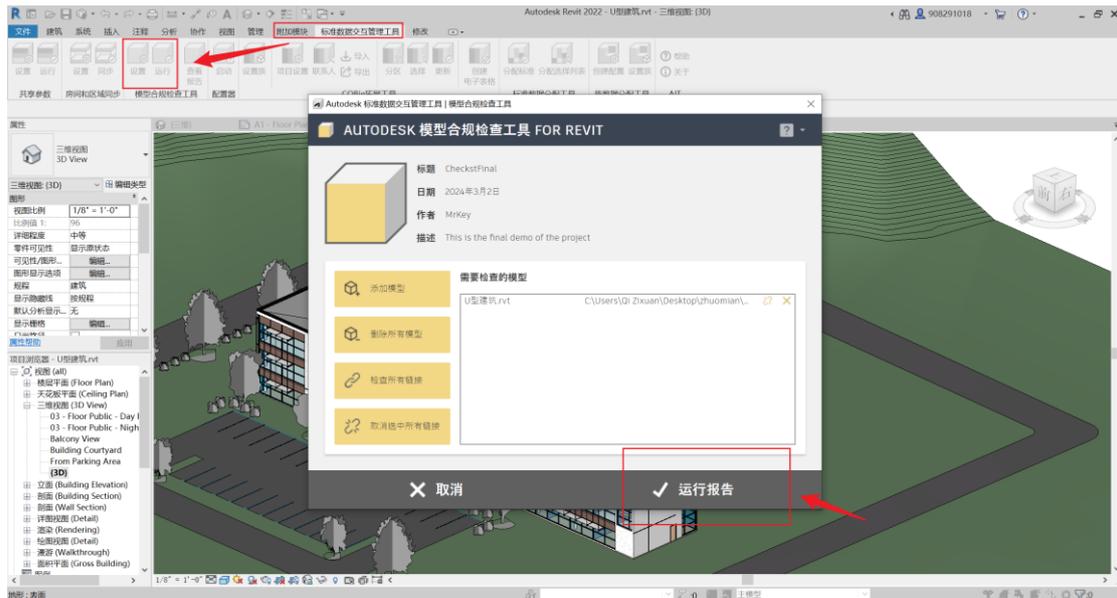


图 39 Revit Model Checker 运行界面

Check ID	Element ID	Ca	Family	Type	Name	Value	Cour	Revit File	Object Handle
55784074-019014	窗	固定	0406 x 0610mm	窗: 类型: 固定: 0406 x 0610mm		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-019016	窗	固定	0406 x 1220mm	窗: 类型: 固定: 0406 x 1220mm		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-019026	窗	固定	0406 x 1830mm	窗: 类型: 固定: 0406 x 1830mm		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-019018	窗	固定	0610 x 0610mm	窗: 类型: 固定: 0610 x 0610mm		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-019020	窗	固定	0610 x 1220mm	窗: 类型: 固定: 0610 x 1220mm		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-019028	窗	固定	0610 x 1830mm	窗: 类型: 固定: 0610 x 1830mm		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-019022	窗	固定	0915 x 0610mm	窗: 类型: 固定: 0915 x 0610mm		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-019024	窗	固定	0915 x 1220mm	窗: 类型: 固定: 0915 x 1220mm		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-019030	窗	固定	0915 x 1830mm	窗: 类型: 固定: 0915 x 1830mm		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0423760	窗	窗户_单开	100 x 1500mm	窗: 类型: 窗户_单开: 100 x 1500mm		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0226778	窗	窗户_单开	200 x 1500mm	窗: 类型: 窗户_单开: 200 x 1500mm		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0296275	窗	窗户_单开	300 x 1500mm	窗: 类型: 窗户_单开: 300 x 1500mm		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0423766	窗	窗户_单开	500 x 1500mm	窗: 类型: 窗户_单开: 500 x 1500mm		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0226730	窗	窗户_单开	600 x 1200 mm	窗: 类型: 窗户_单开: 600 x 1200 mm		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0226792	窗	窗户_单开	600 x 1500mm	窗: 类型: 窗户_单开: 600 x 1500mm		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0383490	窗	窗户_单开	BY0302_300 x 200n	窗: 类型: 窗户_单开: BY0302_300 x		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0226816	窗	窗户_单开	BY0332_300 x 320n	窗: 类型: 窗户_单开: BY0332_300 x		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0226772	窗	窗户_单开	BY0403_400 x 300n	窗: 类型: 窗户_单开: BY0403_400 x		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0226804	窗	窗户_单开	BY0506_500 x 600n	窗: 类型: 窗户_单开: BY0506_500 x		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0423770	窗	窗户_单开	BY0602_600 x 200n	窗: 类型: 窗户_单开: BY0602_600 x		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0226802	窗	窗户_单开	BY0612_600 x 1200	窗: 类型: 窗户_单开: BY0612_600 x		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0383492	窗	窗户_单开	BY0622_600 x 2200	窗: 类型: 窗户_单开: BY0622_600 x		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0226818	窗	窗户_单开	BY0632_600 x 3200	窗: 类型: 窗户_单开: BY0632_600 x		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0226808	窗	窗户_单开	BY0647_600 x 4700	窗: 类型: 窗户_单开: BY0647_600 x		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0363845	窗	窗户_单开	BY0803_800 x 300n	窗: 类型: 窗户_单开: BY0803_800 x		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0364427	窗	窗户_单开	BY1003_1000 x 300	窗: 类型: 窗户_单开: BY1003_1000 x		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0364140	窗	窗户_单开	BY1005_1000 x 500	窗: 类型: 窗户_单开: BY1005_1000 x		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0364257	窗	窗户_单开	BY1105_1100 x 500	窗: 类型: 窗户_单开: BY1105_1100 x		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0226796	窗	窗户_单开	C0412_400 x 1200r	窗: 类型: 窗户_单开: C0412_400 x 1		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0226632	窗	窗户_单开	C0412_450 x 1200r	窗: 类型: 窗户_单开: C0412_450 x 1		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0226806	窗	窗户_单开	C0615_600 x 1500r	窗: 类型: 窗户_单开: C0615_600 x 1		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	
55784074-0226810	窗	窗户_单开	C0630_600 x 3000r	窗: 类型: 窗户_单开: C0630_600 x 3		1	1	C:\Users\Qi Zixuan\Desktop\zhuor	

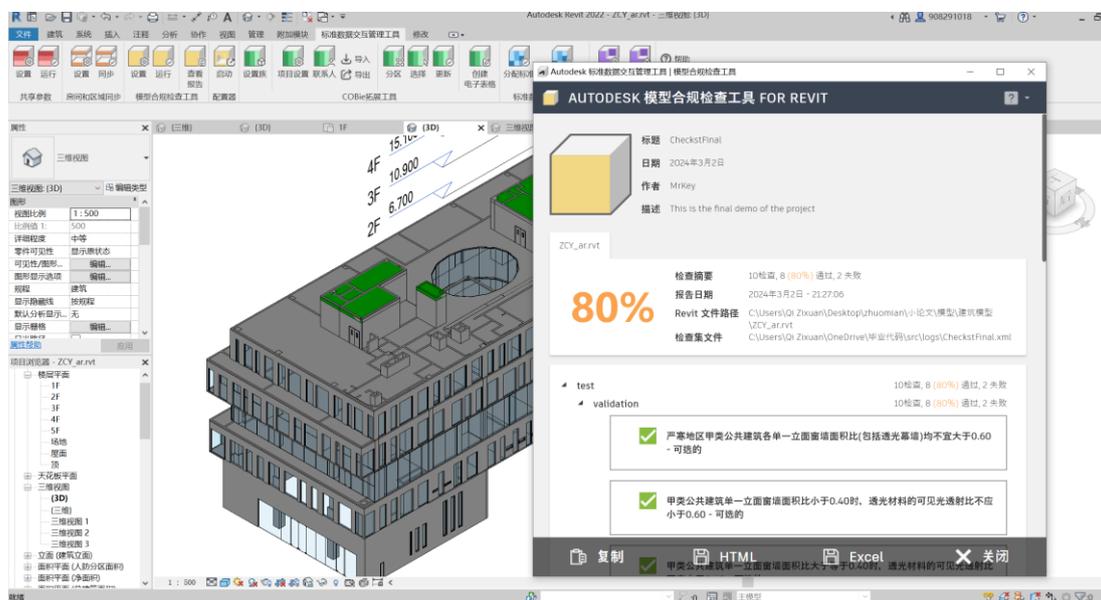
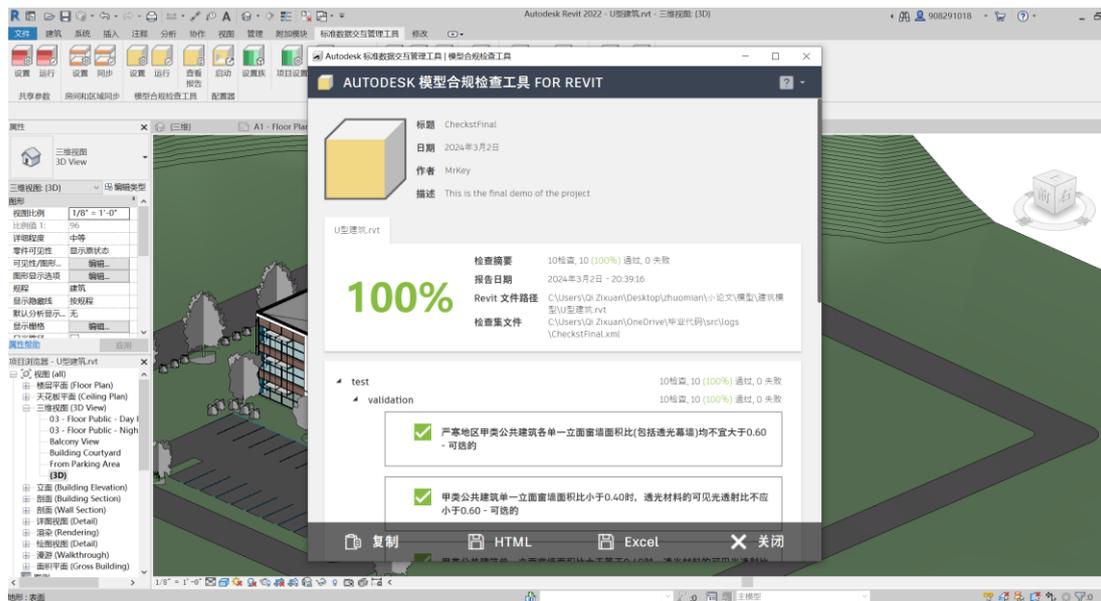
图 40 违规构件列表 (基础报告)

序号	内容	结果	错误	计数	分类	标注	解析	程度词
1	为贯彻国家有关法律法规和方针政策,改善公共建筑的室内环境,提高能源利用效率,促进可再生能源的建筑应用,降低建筑能耗,制定本标准。	人工介入			专家型			
2	建筑体形系数shape factor建筑物与室外空气直接接触的外表面积与其所包围的体积的比值。外表面积不包括地面和不采暖楼梯间内墙的面积				定义型			
3	参照建筑reference building进行围护结构热工性能参数判断时,作为计算满足标准要求的全年供暖和空气调节能耗用的基准建筑。				定义型			
4	单栋建筑面积小于或等于300m2的建筑,应为乙类公共建筑。				定义型			
5	建筑总平面设计及平面布置应合理确定能源设备机房的位置,缩短能源供应输送距离,同一公共建筑的冷热源机房宜位于或靠近冷热负荷中心位置集中设置。	人工介入			专家型			应
6	严寒地区甲类公共建筑各单一立面窗墙面积比(包括透光幕墙)均不宜大于0.60	通过		0	属性型	[严寒地区/obj[甲类/obj[公共建筑/obj[各单一立面窗墙面积比/obj[包括透光幕墙]均不宜大于/cmp[0.60/value]	main_sentence: [严寒地区-甲类-公共建筑-各单一立面-窗墙面积比-均不宜大于-0.60]. conditional_sentence: [包括-透光幕墙]. quoted_sentence: [null]	不宜
7	甲类公共建筑单一立面窗墙面积比小于0.40时,透光材料的可见光透射比不应小于0.60	通过		0	属性型	[甲类/obj[公共建筑/obj[单一立面窗墙面积比/obj[小于/cmp[0.40/value]时/ctg]. [透光材料/obj[的可见光透射比/obj[不应小于/cmp[0.60/value]	main_sentence: [透光材料-可见光透射比-不应小于-0.60]. conditional_sentence: [甲类-公共建筑-单一立面-窗墙面积比-小于-0.40]. main_sentence: [透光材料-可见光透射比-不应小于-0.40]. conditional_sentence: [甲类-公共建筑-单一立面-窗墙面积比-大于等于-0.40]. quoted_sentence: [null]	不应
8	甲类公共建筑单一立面窗墙面积比大于0.40时,透光材料的可见光透射比不应小于0.40	通过		0	属性型	[甲类/obj[公共建筑/obj[单一立面窗墙面积比/obj[大于等于/cmp[0.40/value]时/ctg]. [透光材料/obj[的可见光透射比/obj[不应小于/cmp[0.40/value]	main_sentence: [乙类-公共建筑-外窗-有效通风换气面积-不宜小于-quoted]. conditional_sentence: [null]. quoted_sentence: [面积-30%]	不应
9	乙类公共建筑外窗有效通风换气面积不宜小于窗面积的30%	通过		0	间接型	[乙类/obj[公共建筑/obj[外窗/obj[有效通风换气面积/obj[不宜小于/cmp[窗面积/obj[的[30%/value]	main_sentence: [乙类-公共建筑-外窗-有效通风换气面积-不宜小于-quoted]. conditional_sentence: [null]. quoted_sentence: [面积-30%]	不宜
10	散热器宜明装	通过		0	属性型	[散热器/obj[宜/cmp[明装/obj]	main_sentence: [散热器-宜-明装]. conditional_sentence: [null]. quoted_sentence: [null]	宜
11	送风高度小于或等于5m时,送风温差宜小于5℃	通过		0	属性型	[送风高度/obj[小于或等于/cmp[5m/value]时/ctg]. [送风温差/obj[不宜小于/cmp[5/℃/value]	main_sentence: [送风高度-不宜小于-5℃]. conditional_sentence: [送风高度-小于或等于-5m]. quoted_sentence: [null]	不宜
12	在人员长期停留的房间,其顶棚内表面可见光反射比应当小于等于0.9且大于等于0.7。	通过		0	属性型	[在/ctg]人员长期停留的房间/obj]. [其/obj[顶棚内表面可见光反射比/obj[应当小于等于/cmp[0.9/value]且大于等于/cmp[0.7/value]	main_sentence: [房间-顶棚-内表面-可见光反射比-应当小于等于-0.9]且[应当大于等于-0.7].	应

图 41 最终生成报表

如图 42, 在三个案例中, 通过率分别为: 100%、80%和 90%。案例 1 得到了较高的通过分数, 主要原因是案例 1 中并没有布置任何暖通系统设备并且模型中的围护结构构件参数经过课题组自动设计项目精调, 使之满足现国标标准; 案例 2 和案例 3 中出现的违规情况主要集中在房间顶棚的可见光反射比, 链接条文“在人员长期停留的房间, 其顶棚内表面可见光反射比应当小于等于 0.9 且大于等于 0.7。”以及对于供暖系统热媒的规定上, 即“集中供暖系统应采用热水

作为热媒。”在运行案例 2 检查前，本研究将 1F 的休息室中其中一个散热器热媒手动修改为蒸汽。因此，在图 43 中可以看到，经过修改的散热器在检查中被检查出来并高亮，在它旁边的另一个使用热水作为热媒的散热器并没有被插件记录在最终错误报表中。本系统从载入待检查条款到最终生成 3 栋建筑、10 个条款的审查报表输出仅仅用时 293 秒，这大大得提升了审查模型的效率。



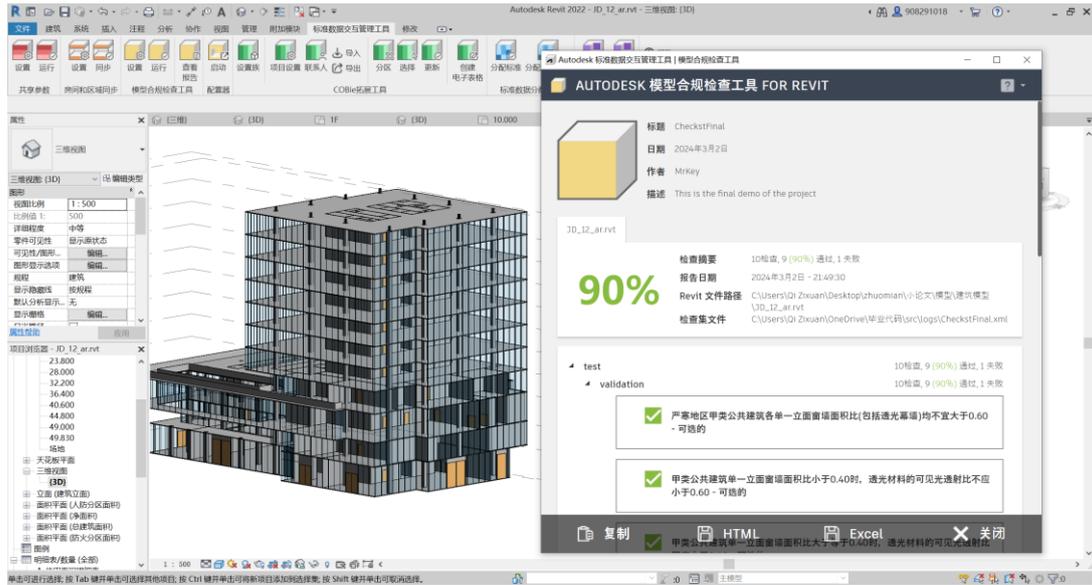


图 42 案例检查结果

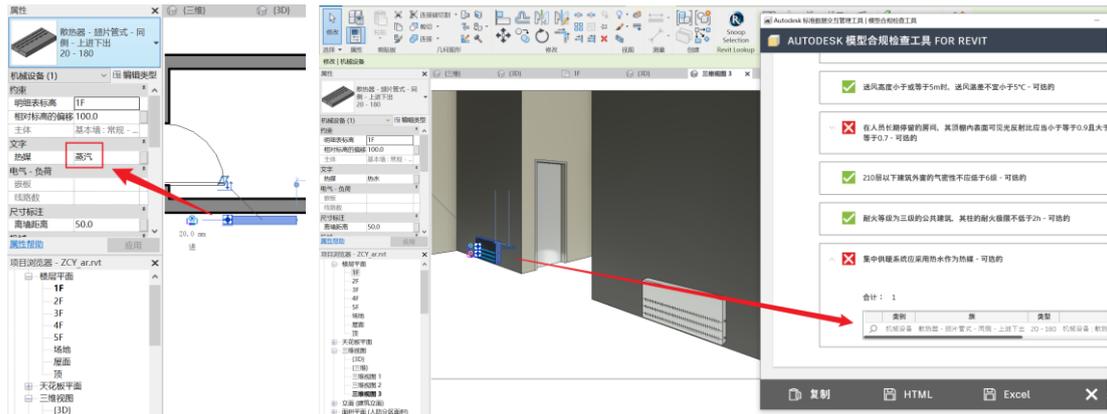


图 43 手动修改散热器热媒检查结果

## 7.4 本章小结

本章基于前几章的技术内容,选取了三栋形状、体量均各异的测试建筑模型和 15 条来自词性标注模型训练测试集的建筑规范条款,对条款的分类模型、词性标注模型、大语言模型解析结果以及模型检查报表进行了测试验证并对其中的错误原因进行了分析。测试结果表明该任务流程运行顺畅,从载入待检查条款到最终生成 3 栋建筑、10 个条款的审查报表输出仅仅用时 293 秒,提升了审查模型的效率。

## 第 8 章 结论与展望

### 8.1 研究总结

BIM 模型审查依赖人工方法，效率低下且易出错，进而产生了建筑自动合规性检查的需求。但即使是当下最先进的全自动检查方法也仍存在瓶颈，如标注数据集不统一，过分依赖规则化的条款解析方式所造成的解析复杂、覆盖范围窄，经常需要大量人力维护等问题，限制了其在实际用途中的推广。因此，本文采用大语言模型、深度学习、NLP 等技术，提出一种全新的规则自动解析方法并应用到建筑自动合规性检查系统。本研究主要的研究内容和贡献包括：

- (1) 基于此研究方向中词性标注数据集难以互通问题，提出了一种更为普适的条款语义简化结构：大语言模型以其强大的理解能力和推理能力在信息提取任务中表现出色。因此，对于大语言模型来说，并不需要为其提供多余的规则用于条款句子的解析，仅仅只需要简单的物件标签和子句关系标签帮助大语言模型规避专有名词和连接子句之间逻辑即可，而这几类标签在以往的研究者是普遍存在的，从而解决了各研究数据集难以互通的问题。
- (2) 将最先进的大语言模型应用于暖通领域的规范解析，建立并验证基于大语言模型的建筑自动合规性检查流程框架：以往的研究过于依赖与基于标签规则的解析方法，因此不仅仅要引入复杂多样的词性标签，还要基于所提出的标签提出一系列条款解析规则。而大语言模型的信息提取能力可以很好地取代这样的信息提取方法，因此，考虑到现有规范并非完全适合建筑自动合规性检查，本课题首先根据《标准化工作导则 第 1 部分：标准化文件的结构和起草规则》（GB/T 1.1-2020）和建筑规范特点，定义了四种条款类型；然后从《建筑设计防火规范》和《公共建筑节能设计标准》规范中经过文本预处理获得了 1097 例条款，并在标注后使用 BERT 模型微调训练得到条款分类模型；接着，本研究提出了一种普适的条款语义结构并微调训练得到词性自动标注模型；同时基于提示工程对不同条款自动生成对应提示词，输入大语言模型进行条款解析。最后，本研究通过 Revit Model Checker 插件开发了一套全自动 BIM 模型检查工具，并选取了三栋体量各异的 BIM 建筑模型对该系统进行了测试，验证了该方法的可行性。
- (3) 随着越来越多的大模型出现，社区对于不同模型在各类任务的应用开发和测评正在如雨后春笋一般涌现，本课题所涉及的中文建筑条款解析任务，也需

要参与进去，得到表现优良的模型+提示策略组合。因此，本研究首先介绍了一些常见的提示方法，拟定了一个提示词架构；之后，基于各类提示词部件功能，结合本课题的“条款解析”任务，撰写了此任务的提示词并为后续模型调用；最后，本研究使用当下主流且支持中文输出的大模型进行了模型、提示词组合实验，得到了多种提示词在此类任务中对模型表现的普适影响，同时本文在结论分析中给出了三类表现基准模型下不同的最佳提示词组合策略。其中表现最佳的文心 4 大模型在各实验组中的表现均十分优秀，在“指令遵循”、“格式遵循”、“分类正确”、“解析正确”和“主观评分”五类指标中分别获得了 98.75%、98.13%、97.50%、91.25 和 96.38 分的平均成绩，并在 4 类指标中获得最佳。因此，本文选择了表现最好的文心 4+实验组 11 的提示策略作为后续条文解析的解决方案。同时该实验也证明了大语言模型在建筑自动合规性检查应用中的优越性。

- (4) 基于 Revit Model Checker 软件开发出一套少人力工作的全自动建筑合规性检查系统。该系统基于插件导出的检查报表生成，除了将不符合条款逻辑的构件以列表的方式进行罗列以外。最终报表还记录了每个条款在条款分类、条款标注以及条款解析的中间结果，为审查人员提供更为全面、可解释性更强的模型检查报表，从而帮助审图人员进行后续判别，提高该系统在实际工程中的实用性。

## 8.2 未来展望

本研究相较于其他建筑自动合规性检查研究，最重要的出发点就是尝试通过大语言模型推进简化当下过于复杂的“条款解析”环节，并且在一定程度上取得了成功。但是因为本人的时间和精力有限，从条款解析到模型检查中间依然需要大量的元素对齐工作，并且因为部分条款本身表达并不包含“数值对比”，导致许多条款的语义逻辑在插件检查中无法对应到具体的运算逻辑中，这是在近几年此方向研究中最需要被解决的问题。除此之外，本研究还更为具体地罗列出了未来的研究可以关注的问题：

- (1) 建筑自动合规性检查流程的全黑箱化：本研究在尝试用黑盒式的大语言模型取代以往基于规则和语法计算器的方式进行条款解析，仅仅只是将黑盒模型的输出推进到结构化的语义表达。然而模型检查最终需要的是检查系统可运行的代码文件（在本研究中就是 Revit Model Checker 插件的 XML 输入文件）。随着大语言模型逻辑推理能力不断增强，这个流程将会被很快黑箱化。从而规避前序工作中的语义结构定义不够普适以及部分条款语料和建筑构

件模型之间参数对齐工作所带来的不稳定因素。

- (2) 建筑模型数据增强方法：本研究使用了设计说明作为外挂的项目设计参数来源。即使如此，当下的模型质量依然不支持全面的自动合规性检查。现实工程所产出的设计模型往往存在很多的数据缺失，这些数据往往藏在各类设计文档中或是被行业内视为工程经验从而被忽略甚至是标注成错误的数据。除此之外，建筑规范条款中所涉及的许多参数往往需要进行额外的计算（例如：最短逃生距离、窗墙比、体型系数等）。数据的缺失会造成检查系统对于事实情况的忽略，从而导致漏查、错查的情况。因此，如何从现有的工程文件中获取更加准确且全面的建筑模型参数依然是一个值得探究的问题。
- (3) 基于开源协议和平台的模型审查工具：本研究所使用的 Revit Model Checker 是 Autodesk 公司所推出的商业软件。该软件首先在建筑构件的定义结构上极为复杂，因此在条款元素映射的过程中经常出现无法和某类构件代码进行对应的情况，从而降低了系统通用性；其次，该软件运算逻辑较为简单，无法对间接型引用条款进行处理。当下开源的模型审查库极少，其中使用最多的 MVDXML-master 库需要用户熟练掌握 IFC 协议以及对应的代码开发水平，这对于一线设计院工作人员来说是无法进行普及的；同时，巨量 IFC 数据所带来的计算量激增也会对模型检查效率带来挑战。因此，未来需要更多针对开源模型审查工具的研究。
- (4) 非自然语言文本建筑规范研究：自动合规性检查一直在尝试跨越自然语言规范文本所带来的模糊性、多义性鸿沟。这会带来一系列因“解读”环节的存在而出现的各类复杂问题，是否存在一个更为高效且对计算机友好的设计规则标准表达方式，让文本条款跨越至抽象逻辑规则的规范制定方法，从而使自动审查变得更为轻量、高效。
- (5) 更加多样的检查交互方式：除了用自然语言文本以及模型数据作为此类任务的输入，设计师需要更加直观的实时交互方式对模型进行审查，从而提高模型审查的可解释性以及准确性。例如，在《建筑设计防火规范》（GB50016-2014）中，有对于最短逃生距离的规定。此类参数如果仅通过基于几何数据来获取则显得极不直观且难以计算，而通过设计师屏幕所展现的实时图像可以更加直观地对整个建筑的布局、走管以及相对高度等因素进行同时考虑，从而减少系统所需的中间流程，规避各流程所带来的错误隐患。同时，发展图像、语音所驱动的审查系统也可以为下一代智能终端所带来的设计方式剧变提前做好技术储备。



## 附录 A：条款解析提示组件（JSON）

```

{
  "role": {
    "chinese": "****角色****\n 你是一位在建筑领域非常有经验的教授，擅长解析建筑规范条款。",
    "english": "****Role****You are a highly experienced professor in the field of architecture, skilled at analyzing building regulatory clauses."
  },
  "instruct": {
    "chinese": "****指令****\n 你的任务是解析现有的中文建筑条款，从提取出条款的主句、条件子句、引用子句。并按照以下格式输出：\n'main_sentence:[xxx],conditional_sentence:[xxx],quoted_sentence:[xxx]'\n 其中：\n1. 主句(main_sentence): 条款中心句，主要句子不能是[null]或为空!!! 而且可能会有多个连接，请用'&'表示'和'，或'/'表示'或'来连接这些连接。\n2. 条件句(conditional_sentence):在子句开始调节主要句子之前表示条件限制。条件句可以是[null]! 而且可能会有多个连接，请用'&'表示'和'，或'/'表示'或'来连接这些连接。\n3. 引用句(quoted_sentence):一些主要子句可能不会调节特定元素的值，它们可能会提及其他元素与之比较。引用的句子可以是[null]",
    "english": "****Instruction**** Your task is to analyze existing Chinese architectural clauses, extracting the main clause, conditional clause, and quoted clause from them. The format for the output should be as follows:\n'main_sentence:[xxx],conditional_sentence:[xxx],quoted_sentence:[xxx]'.\nWhere:\n1. Main sentence (main_sentence): The central sentence in the clause. The main sentence cannot be [null] or empty!!! There may be multiple connections, please use '&' to represent 'and' or '/' to represent 'or' to connect these connections.\n2. Conditional sentence (conditional_sentence): Represents conditional restrictions before modifying the main sentence. The conditional sentence can be [null]! There may be multiple connections, please use '&' to represent 'and' or '/' to represent 'or' to connect these connections.\n3. Quoted sentence (quoted_sentence): Some main clauses may not modify the values of specific elements. They may refer to other elements for comparison. Quoted sentences can be [null]."
```

```

    "z_cot": {
        "chinese": "零样本思维链\n 让我们一步一步思考：\n1.首先，尝试识别条款的不同部分，可能包括条件句、主句（通常只有一个）、引用句（通常只有一个）。\n2.其次，提取每个句子内的元素，并连接这些元素。\n3.最后，将答案填入架构中！",
        "english": "zero-shot chain of thought Remember: let's think step by step: \n1. Firstly, try to identify different part of a clause, there could be conditional sentences, main sentence(usually only one), quoted sentence(usually only one).\n2. Secondly, extract elements inside of each sentence and connect those elements.\n3. Finally you fill the answer into the schema!"
    },
    "f_example": {
        "chinese": "多样本思维链\n 让我们一步一步思考：\n1.首先，尝试识别条款的不同部分，可能包括条件句、主句（通常只有一个）、引用句（通常只有一个）。\n2.其次，提取每个句子内的元素，并连接这些元素。\n3.最后，将答案填入架构中！",
        "english": "few-shot chain of thought Remember: let's think step by step: \n1. Firstly, try to identify different part of a clause, there could be conditional sentences, main sentence(usually only one), quoted sentence(usually only one).\n2. Secondly, extract elements inside of each sentence and connect those elements.\n3. Finally you fill the answer into the schema!"
    },
    "f_cot": {
        "chinese": "多样本思维链\n 让我们一步一步思考：\n1.首先，尝试识别条款的不同部分，可能包括条件句、主句（通常只有一个）、引用句（通常只有一个）。\n2.其次，提取每个句子内的元素，并连接这些元素。\n3.最后，将答案填入架构中！",
        "english": "few-shot chain of thought Remember: let's think step by step: \n1. Firstly, try to identify different part of a clause, there could be conditional sentences, main sentence(usually only one), quoted sentence(usually only one).\n2. Secondly, extract elements inside of each sentence and connect those elements.\n3. Finally you fill the answer into the schema!"
    },
    "tot": {

```

"chinese": "零样本思维树" \n 让我们一步一步思考： \n1.首先，尝试识别条款的不同部分，可能包括条件句、主句（通常只有一个）、引用句（通常只有一个）。 \n2.其次，提取每个句子内的元素，并连接这些元素。 \n3.最后，将答案填入架构中！ "

"english": "zero-shot tree of thought" Remember: let's think step by step: \n1. Firstly, try to identify different part of a clause, there could be conditional sentences, main sentence(usually only one), quoted sentence(usually only one).\n2. Secondly, extract elements inside of each sentence and connect those elements.\n3. Finally you fill the answer into the schema!"

},

"schema": {

"chinese": "输出编码和注意事项" \n 例如，在句子：'高层厂房和甲、乙、丙类多层厂房的疏散楼梯应采用封闭楼梯间或室外楼梯' 中，'高层厂房' 和 '甲、乙、丙类多层厂房'是平行关系，用'OR'连接；'封闭楼梯间'和'室外楼梯'也是平行关系，同样用'OR'连接（有时也可以是'AND'，您应该能够判断）。因此，将这些元素组合起来将产生四个具有不同含义的子句。这个句子应该在完整的输出上下文中表示如下： \nOutput:main\_sentence:[(高层-厂房-疏散楼梯-应采用-封闭楼梯间)/(甲、乙、丙类多层-厂房-疏散楼梯-应采用-封闭楼梯间)/(高层-厂房-疏散楼梯-应采用-室外楼梯)/(甲、乙、丙类多层-厂房-疏散楼梯-应采用-室外楼梯)],conditional\_sentence:[null],quoted\_sentence:[null]有几点需要注意： \n1.不要忘记使用 '-' 来分隔可能出现在每种类型的句子内部的元素！ \n2.当文本中出现字符'&'和'/'（其中这些符号分别表示'and'和'or'）时，请使用成对的括号来澄清左侧和右侧元素之间的关系。此外，将两种情况都呈现为完整的子句，而不仅仅是列出平行的元素。 "

"english": "输出编码" \nFor example, in the sentence: 'High-rise factories and Class A, B, and C multi-story factories should use enclosed stairwells or outdoor staircases,' 'High-rise factories' and 'Class A, B, and C multi-story factories' are parallel elements connected by 'OR'; 'enclosed stairwells' and 'outdoor staircases' are also parallel elements, and they are connected by 'OR' as well (sometimes it could be 'AND,' but you should be able to determine it). Therefore, combining these elements will produce four subclauses with different meanings. This sentence should be represented in the complete output context as follows:\nOutput:main\_sentence:[(高层-厂房-疏散楼梯-应采用-封闭楼梯间)/(甲、乙、丙类多层-厂房-疏散楼梯-应采用-封闭楼梯间)/(高层-厂房-疏散楼梯-应采用-室外楼梯)/(甲、乙、丙类多层-厂房-疏

散楼梯-应采用-室外楼梯)],conditional\_sentence:[null],quoted\_sentence:[null]\nhere are a few points to keep in mind:\nDon't forget to use '-' to separate elements that may appear within each type of sentence!When the characters '&' and '/' (where these symbols represent 'and' and 'or,' respectively) appear in the text, use paired parentheses to clarify the relationship between the left and right elements. Additionally, present both cases as complete clauses, not just listing parallel elements."

},

"klg": {

"chinese":"\*\*\*标签知识\*\*\*\n 在我展示给你输出结构之前,我会给你一些已经设定好的学习材料,希望你在进行解析工作时能记住这些元素的作用!\n 标签解释:\n1. Obj(object): 条款中出现的语义元素,通常是一个建筑构件,元素之间存在父级和属性关系。例如:“窗户的反射率”中,存在“窗户”和“反射率”两个语义元素,“反射率”是“窗户”的属性。元素之间的连接可能是并列也可能是上下游关系,比如:窗户[obj]的长[obj]和宽[obj];也可能是有包含关系的,比如:窗户[obj]边框[obj]的厚度[obj],这里的厚度和边框就是从属关系而不是并列关系.\n2. cmp(compare): 是比较谓词,比如:应当满足,不应大于,不应小于,表示一个比较的判断.\n3. qtg(quoted trigger): 属于 trigger 的一种,是一种表示引用的触发标签。比如:应当满足[cmp]表 1-1 的要求[qtg];应当满足[cmp]《某个规范》的要求[qtg]\n4. ctg: conditinal trigger, 属于 trigger 的一种,是一种表示条件的出发标签。比如:当 xxxx 时[ctg],在 xxxx 条件下[ctg]\n5. value: 代表具体值,比如:甲类[fobj]商业建筑[fobj]窗户[obj]的反射率[prop]应当大于[cmp]0.6[value]\n6. unit: 表示单位,比如:窗户[obj]边框[prop]的厚度[prop]应当小于[cmp]100[value]mm[unit]。在条款中往往在[value]后面,可能和 value 混在一起,你需要把它挑出来。",

"english":"\*\*\*Label Reference\*\*\*\nBefore i show you the schema, I will give you something(in Chinese) that has been set up to learn and hope you can keep these elements in mind when you are doing your parsing job!\n 标签解释:\n1. Obj(object): Semantic elements appearing in the clause, usually a building component. Elements have hierarchical and attribute relationships. For example, in 'window reflectivity,' 'window' and 'reflectivity' are semantic elements, with 'reflectivity' being an attribute of 'window'. Connections between elements can be parallel or hierarchical, e.g., the length[obj] and width[obj] of a window[obj]; or contain relationships, e.g., the thickness[obj] of a window[obj] frame[obj], where thickness and frame have a subordinate relationship.\n2. cmp(compare): A comparative predicate, such as 'should

meet,' 'should not exceed,' 'should not be less than,' indicating a comparative judgment.\n3. qtg(quoted trigger): A type of trigger, indicating a quotation. For example: should meet[cmp] the requirements[qtg] of Table 1-1; should meet[cmp] the requirements[qtg] of a certain standard.\n4. ctg: conditinal trigger, conditional trigger, a type of trigger, indicating a condition. For example: when xxxx[ctg], under the condition of xxxx[ctg].\n5. value: Represents a specific value, e.g., for a Class A[fobj] commercial building[fobj], the reflectivity[prop] of the windows[obj] should be greater than[cmp] 0.6[value].\n6. unit : Indicates a unit, e.g., the thickness[prop] of a window[obj] frame[prop] should be less than[cmp] 100[value] mm[unit]. In clauses, it often follows [value] and may be mixed with it; you need to distinguish it."

```
    },  
    "input": {  
        "chinese": "*****输入*****",  
        "english": "*****Input*****"  
    }  
}
```



## 附录 B：案例提示库（JSON）

```
[
  {
    "input": "[乙类/obj][公共建筑/obj][外窗/obj][有效通风换气面积/obj][不宜小于/cmp][窗面积/obj]的[30%/value]。",
    "cot": "(1) 为解析这个条款，需要从中心句和引用句两个角度进行分析。
\n(2) 中心句（Main Sentence）：中心句阐述了该条款中的主要规定，即乙类公共建筑外窗的有效通风换气面积不应小于窗面积的一定比例。因此，中心句为：‘乙类-公共建筑-外窗-有效通风换气面积-不宜小于-’。
\n(3) 条件句（Conditional Sentence）：这个条款没有明确的条件句，因此可以视为无条件或条件不适用。
\n(4) 引用句（Quoted Sentence）：条款中提到了‘窗面积的 30%’，这是一个具体的比例引用，因此引用句为：‘外窗-窗面积-30%’。",
    "output": "main_sentence: [乙类-公共建筑-外窗-有效通风换气面积-不宜小于-], conditional_sentence: [null], quoted_sentence: [外窗-窗面积-30%]"
  },
  {
    "input": "[当/ctg][供暖系统/obj]的[设计回水水温/obj][小于或等于/cmp][50°C/value][时/ctg], [宜采用/cmp][冷凝式锅炉/value]。",
    "cot": "(1) 为解析这个条款，需要从条件句和中心句两个角度进行分析。
\n(2) 中心句（Main Sentence）：中心句阐述了该条款中的主要规定，即在特定条件下建议使用的设备类型。本条款最终的目的是推荐在某些条件下使用冷凝式锅炉。因此，中心句为：‘供暖系统-宜采用-冷凝式锅炉’。
\n(3) 条件句（Conditional Sentence）：这部分表明了这个建议在特定操作条件下适用，即当供暖系统的设计回水水温小于或等于 50°C 时。因此，条件句为：‘供暖系统-设计回水水温-小于或等于-50-°C’。
\n(4) 引用句（Quoted Sentence）：这个条款中没有明确的引用句。",
    "output": "main_sentence: [供暖系统 - 宜采用 - 冷凝式锅炉], conditional_sentence: [供暖系统 - 设计回水水温 - 小于或等于 -50- °C], quoted_sentence: [null]"
  },
  {
```

"input": "[耐火等级/obj][为/cmp][四级/value]的[厂房/obj], [其/obj][防火墙/obj]的[耐火极限/obj][不低于/cmp][3h/value]。",

"cot": "(1) 为解析这个条款, 需要从条件句和中心句两个角度进行分析。  
 \n(2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 即四级耐火等级的厂房的防火墙耐火极限。条款中的主要内容是关于防火墙的耐火极限, 且要求不低于 3 小时。因此, 中心句为: ‘厂房-防火墙-耐火极限-不低于-3h’。  
 \n(3) 条件句 (Conditional Sentence): 这个条款的适用条件是厂房的耐火等级为四级。因此, 条件句为: ‘厂房-耐火等级-为-四级’。  
 \n(4) 引用句 (Quoted Sentence): 这个条款中没有明确的引用句。",

"output": "main\_sentence: [厂房-防火墙-耐火极限-不低于-0.5-h],  
 conditional\_sentence: [厂房-耐火等级-为-四-级], quoted\_sentence: [null]"

},

{

"input": "[严寒地区/obj][甲类/obj][公共建筑/obj]各[单一立面/obj][窗墙面积比/obj](包括透光幕墙)均[不宜大于/cmp][0.60/value]。",

"cot": "(1) 为解析这个条款, 需要从条件句和中心句两个角度进行分析。  
 \n(2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 即在特定条件下公共建筑立面窗墙比不应过大。本条款最终的目的是限制建筑的窗墙比。因此, 中心句为: ‘供暖系统-宜采用-冷凝式锅炉’。  
 \n(3) 条件句 (Conditional Sentence): 这个条款中没有明确的条件句。  
 \n(4) 引用句 (Quoted Sentence): 这个条款中没有明确的引用句。",

"output": "main\_sentence: [严寒地区-甲类-公共建筑单一立面-窗墙面积比-不宜大于-0.60], conditional\_sentence: [null], quoted\_sentence: [null]"

},

{

"input": "[水冷变频螺杆式冷水机组/obj]的[综合部分负荷性能系数(IPLV)/obj][不应低于/cmp][表 2.11 中/qtg][水冷螺杆式冷水机组/obj][限值/obj]的[1.15 倍/value]。",

"cot": "(1) 为解析这个条款, 需要从条件句、引用句和中心句三个角度进行分析。  
 \n(2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 即水冷变频螺杆式冷水机组的综合部分负荷性能系数(IPLV)不应低于表 2.11 中水冷螺杆式冷水机组的限值的 1.15 倍。因此, 中心句为: “水冷变频螺杆式冷水机组-综合部分负荷性能系数(IPLV)-不应低于-”  
 \n(3) 条件句 (Conditional Sentence): 这个条款中没有明确的条件句。  
 \n(4) 引用句 (Quoted Sentence): 条

款中引用了“1.15 倍”，这是一个具体的数值引用，因此引用句为：“表 2.11 中-水冷螺杆式冷水机组-综合部分负荷性能系数(IPLV)-限值-1.15-倍 “。”,

```
"output": "main_sentence: [水冷变频螺杆式冷水机组-综合部分负荷性能系数(IPLV)-不应低于-quote], conditional_sentence: [null], quoted_sentence: [表 2.11 中-水冷螺杆式冷水机组-综合部分负荷性能系数(IPLV)-限值-1.15-倍]"
```

```
},
```

```
{
```

```
"input": "[高层/obj][厂房/obj]和[甲、乙、丙类/obj][多层/obj][厂房/obj]的[疏散楼梯/obj][应采用/cmp][封闭楼梯间/value]或[室外楼梯/value]。",
```

"cot": "(1) 为解析这个条款，需要从条件句、引用句和中心句三个角度进行分析。  
(2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定，即高层厂房和甲、乙、丙类多层厂房的疏散楼梯应采用封闭楼梯间或室外楼梯。因此，中心句为：“(高层/甲、乙、丙类-多层)-厂房-疏散楼梯-应采用-(封闭楼梯间/室外楼梯)”。  
(3) 条件句 (Conditional Sentence): 这个条款中没有明确的条件句。  
(4) 引用句 (Quoted Sentence): 这个条款没有明确的引用句，因此可以视为无条件或条件不适用。”

```
"output": "main_sentence: [(高层/甲、乙、丙类-多层)-厂房-疏散楼梯-应采用-(封闭楼梯间/室外楼梯)], conditional_sentence: [null], quoted_sentence: [null]"
```

```
},
```

```
{
```

```
"input": "单栋[建筑面积/obj][小于或等于/cmp][300m2/value]的[建筑/obj], [应为/cmp][乙类公共建筑/value]。",
```

"cot": "(1) 为解析这个条款，需要从条件句、引用句和中心句三个角度进行分析。  
(2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定，即建筑面积小于或等于 300m<sup>2</sup> 的建筑应为乙类公共建筑。因此，中心句为：“建筑-应为-乙类公共建筑”。  
(3) 条件句 (Conditional Sentence): 条件句表明了这个标准是在建筑面积小于或等于 300m<sup>2</sup> 的情况下适用，因此条件句为：“建筑-建筑面积-小于或等于-300-m2”。  
(4) 在这个条款中没有引用具体的表或条款或标准，因此引用句为：“null”。”

```
"output": "main_sentence: [建筑 - 应为 - 乙类公共建筑], conditional_sentence: [建筑-建筑面积-小于或等于-300-m2], quoted_sentence: [null]"
```

```
},
```

```
{
```

"input": "[甲类/obj][公共建筑/obj][单一立面/obj][窗墙面积比/obj][小于/cmp][0.40/value][时/ctg], [透光材料/obj]的[可见光透射比/obj][不应小于/cmp][0.60/value]。",

"cot": "(1) 为解析这个条款, 需要从条件句、引用句和中心句三个角度进行分析。  
 (2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 即当甲类公共建筑单一立面的窗墙面积比小于 0.40 时, 透光材料的可见光透射比不应小于 0.60。因此, 中心句为: “透光材料-可见光透射比-不应小于-0.60”。  
 (3) 条件句 (Conditional Sentence): 条件句表明了这个标准是在窗墙面积比小于 0.40 的情况下适用, 因此条件句为: “甲类-公共建筑-单一立面-窗墙面积比-小于-0.40”。  
 (4) 在这个条款中没有引用具体的表或条款或标准, 因此引用句为: “null”。”

"output": "main\_sentence: [透光材料-可见光透射比-不应小于-0.60],  
 conditional\_sentence: [甲类-公共建筑-单一立面-窗墙面积比-小于-0.40],  
 quoted\_sentence: [null]"  
 },  
 {

"input": "[当/ctg][透光幕墙/obj]受条件限制[无法设置/cmp][可开启窗扇/value][时/ctg], [应设置/cmp][通风换气装置/value]。",

"cot": "(1) 为解析这个条款, 需要从条件句、引用句和中心句三个角度进行分析。  
 (2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 即当透光幕墙受条件限制无法设置可开启窗扇时, 应设置通风换气装置。因此, 中心句为: “透光幕墙-应设置-通风换气装置”。  
 (3) 条件句 (Conditional Sentence): 条件句表明了这个标准是在透光幕墙受条件限制无法设置可开启窗扇的情况下适用, 因此条件句为: “透光幕墙-无法设置-可开启窗扇”。  
 (4) 在这个条款中没有引用具体的表或条款或标准, 因此引用句为: “null”。”

"output": "main\_sentence: [透光幕墙-应设置-通风换气装置],  
 conditional\_sentence: [透光幕墙-无法设置-可开启窗扇], quoted\_sentence: [null]"  
 },  
 {

"input": "[乙类/obj][公共建筑/obj][外窗/obj][有效通风换气面积/obj][不宜小于/cmp][窗面积/obj]的[30%/value]。",

"cot": "(1) 为解析这个条款, 需要从条件句、引用句和中心句三个角度进行分析。  
 (2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 即乙类公共建筑外窗的有效通风换气面积不宜小于窗面积的 30%。因此, 中心句为: “乙类-公共建筑-外窗-有效通风换气面积-不宜小于-quote”。  
 (3) 条件句

(Conditional Sentence): 这个条款没有明确的条件句, 因此可以视为无条件或条件不适用。  
 \n(4) 引用句 (Quoted Sentence): 条款中引用了“30%”, 这是一个具体的比例引用, 因此引用句为: “外窗-窗面积-30%”。”,

```
"output": "main_sentence: [乙类-公共建筑-外窗-有效通风换气面积-不宜小于-quote], conditional_sentence: [null], quoted_sentence: [外窗-窗面积-30%]"
```

```
},
```

```
{
```

```
"input": "[屋面/obj]、[外墙/obj]和[地下室/obj]的[热桥部位/obj]的[内表面温度/obj][不应低于/cmp][室内空气/obj][露点温度/obj]。",
```

"cot": "(1) 为解析这个条款, 需要从条件句、引用句和中心句三个角度进行分析。  
 \n(2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 即屋面、外墙和地下室的热桥部位的内表面温度不应低于室内空气的露点温度。因此, 中心句为: “(屋面&外墙&地下室)-热桥部位-内表面温度-不应低于-quote”。

\n(3) 条件句 (Conditional Sentence): 这个条款没有明确的条件句, 因此可以视为无条件或条件不适用。  
 \n(4) 引用句 (Quoted Sentence): 条款中引用了“室内空气的露点温度”, 因此引用句为: “室内空气-露点温度”。”,

```
"output": "main_sentence: [(屋面&外墙&地下室)-热桥部位-内表面温度-不应低于-quote], conditional_sentence: [null], quoted_sentence: [室内空气-露点温度]"
```

```
},
```

```
{
```

```
"input": "[210 层/value][以下/ctg]的[建筑/obj], [外窗/obj]的[气密性/obj][不应低于/cmp][6 级/value]。",
```

"cot": "(1) 为解析这个条款, 需要从条件句、引用句和中心句三个角度进行分析。  
 \n(2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 即 210 层以下的建筑, 外窗的气密性不应低于 6 级。因此, 中心句为: “外窗-气密性-不应低于-6-级”。  
 \n(3) 条件句 (Conditional Sentence): 条件句表明了这个标准是在建筑高度低于 210 层的情况下适用, 因此条件句为: “建筑-低于-210-层”。

\n(4) 引用句 (Quoted Sentence): 在这个条款中没有引用具体的表或条款或标准, 因此引用句为: “null”。”,

```
"output": "main_sentence: [外窗 - 气密性 - 不应低于 -6- 级 ], conditional_sentence: [建筑-低于-210-层], quoted_sentence: [null]"
```

```
},
```

```
{
```

"input": "[严寒/obj]和[寒冷地区/obj][外门/obj]的[气密性/obj][不应低于/cmp][4级/value]",

"cot": "(1) 为解析这个条款, 需要从条件句、引用句和中心句三个角度进行分析。  
 (2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 即在严寒和寒冷地区, 外门的气密性不应低于 4 级。因此, 中心句为: “(严寒&寒冷地区)-外门-气密性-不应低于-4-级”。  
 (3) 条件句 (Conditional Sentence): 这个条款没有明确的条件句, 因此可以视为无条件或条件不适用。  
 (4) 引用句 (Quoted Sentence): 在这个条款中没有引用具体的表或条款或标准, 因此引用句为: “null”。”

"output": "main\_sentence: [(严寒&寒冷地区)-外门-气密性-不应低于-4-级], conditional\_sentence: [null], quoted\_sentence: [null]"

},

{

"input": "[单台锅炉/obj]的[实际运行负荷率/obj][不宜低于/cmp][50%/value]。”

"cot": "(1) 为解析这个条款, 需要从条件句、引用句和中心句三个角度进行分析。  
 (2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 即单台锅炉的实际运行负荷率不宜低于 50%。因此, 中心句为: “单台锅炉-实际运行负荷率-不宜低于-50%”。  
 (3) 条件句 (Conditional Sentence): 这个条款没有明确的条件句, 因此可以视为无条件或条件不适用。  
 (4) 引用句 (Quoted Sentence): 在这个条款中没有引用具体的表或条款或标准, 因此引用句为: “null”。”

"output": "main\_sentence: [单台锅炉-实际运行负荷率-不宜低于-50%], conditional\_sentence: [null], quoted\_sentence: [null]"

},

{

"input": "[在/ctg]名义工况和规定[条件下/ctg], [锅炉/obj]的[热效率/obj][不应低于/cmp][表 4.2.5/qtg]的[数值/obj]。”

"cot": "(1) 为解析这个条款, 需要从条件句、引用句和中心句三个角度进行分析。  
 (2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 即在名义工况和规定条件下, 锅炉的热效率不应低于表 4.2.5 的数值。因此, 中心句为: “锅炉-热效率-不应低于-”。  
 (3) 条件句 (Conditional Sentence): 条件句表明了这个标准是在名义工况和规定条件下适用, 因此条件句为: “(名义工况&名义规定)”。  
 (4) 引用句 (Quoted Sentence): 在这个条款中引用了“表 4.2.5”

的数值，因此引用句为：“表 4.2.5-数值”。”，

```
"output": "main_sentence: [ 锅炉 - 热效率 - 不应低于 -quote],
conditional_sentence: (名义工况&名义规定), quoted_sentence: [表 4.2.5-数值]"
```

```
},
```

```
{
```

```
"input": "[地面/obj][辐射供暖/obj][面层材料/obj]的[热阻/obj][不宜大于
/cmp][0.05m2 • K/W/value]。”，
```

"cot": "(1) 为解析这个条款，需要从条件句、引用句和中心句三个角度进行分析。  
 (2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定，即地面辐射供暖面层材料的热阻不宜大于  $0.05\text{m}^2 \cdot \text{K/W}$ 。因此，中心句为：“地面-辐射供暖-面层材料-热阻-不宜大于- $0.05\text{-m}^2 \cdot \text{K/W}$ ”。  
 (3) 条件句 (Conditional Sentence): 这个条款没有明确的条件句，因此可以视为无条件或条件不适用。  
 (4) 引用句 (Quoted Sentence): 在这个条款中没有引用具体的表或条款或标准，因此引用句为：“null”。”，

```
"output": "main_sentence: [地面-辐射供暖-面层材料-热阻-不宜大于-
0.05-m2 • K/W], conditional_sentence: [null], quoted_sentence: [null]"
```

```
},
```

```
{
```

```
"input": "[建筑/obj][空间/obj][高度/obj][大于等于/cmp][10m/value]、且
[体积/obj][大于/cmp][10000m3/value][时/ctg], [宜采用/cmp][辐射供暖供冷/obj]或
[分层空气调节系统/obj]。”，
```

"cot": "(1) 为解析这个条款，需要从条件句、引用句和中心句三个角度进行分析。  
 (2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定，即当建筑空间的高度大于等于 10m 且体积大于  $10000\text{m}^3$  时，宜采用辐射供暖供冷或分层空气调节系统。因此，中心句为：“空间-宜采用-辐射供暖供冷/分层空气调节系统”。  
 (3) 条件句 (Conditional Sentence): 条件句表明了这个标准是在建筑空间的高度大于等于 10m 且体积大于  $10000\text{m}^3$  的情况下适用，因此条件句为：“建筑-空间-(高度-大于等于-10-m)&(体积-大于- $10000\text{-m}^3$ )”。  
 (4) 引用句 (Quoted Sentence): 在这个条款中没有引用具体的表或条款或标准，因此引用句为：“null”。”，

```
"output": "main_sentence: [空间-宜采用-辐射供暖供冷/分层空气调节系
统], conditional_sentence: [建筑-空间-(高度-大于等于-10-m)&(体积-大于-10000-
m3)], quoted_sentence: [null]"
```

```
},
```

```

    {
      "input": "[当采用/ctg][空气源热泵热水机组/obj][制备/cmp][生活热水/obj][时/ctg], [制热量/obj][大于/cmp][10kW/value]的[热泵热水机/obj]在名义制热工况和规定条件下, [性能系数（COP）/obj][不宜低于/cmp][表 5.3.3/qtg]的[规定/obj], 并[应有/cmp]保证水质的[有效措施/obj]。",

```

"cot": "(1) 为解析这个条款, 需要从条件句、引用句和中心句三个角度进行分析。  
 \n(2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 即当采用空气源热泵热水机组制备生活热水时, 制热量大于 10kW 的热泵热水机在名义制热工况和规定条件下, 性能系数 (COP) 不宜低于表 5.3.3 的规定, 并应有保证水质的有效措施。因此, 中心句为: “热泵热水机-(性能系数-不宜低于- quote)&(应有-保证水质-有效措施)”。  
 \n(3) 条件句 (Conditional Sentence): 条件句表明了这个标准是在采用空气源热泵热水机组制备生活热水, 并且制热量大于 10kW 的热泵热水机的情况下适用, 因此条件句为: “(空气源热泵热水机组-制备-生活热水)&(热泵热水机-制热量-大于-10-kW)”。  
 \n(4) 引用句 (Quoted Sentence): 在这个条款中引用了“表 5.3.3”的规定, 因此引用句为: “表 5.3.3-规定”。”,

```

      "output": "main_sentence: [热泵热水机-(性能系数-不宜低于- quote)&(应有-保证水质-有效措施)], conditional_sentence: [(空气源热泵热水机组-制备-生活热水)&(热泵热水机-制热量-大于-10-kW)], quoted_sentence: [表 5.3.3-规定]"

```

```

    },
    {
      "input": "[在/ctg]人员长期停留的[房间/obj], [其/obj][墙面/obj][内表面/obj][可见光反射比/obj][应当小于等于/cmp][0.8/value]且[大于等于/cmp][0.5/value]。",

```

"cot": "(1) 为解析这个条款, 需要从条件句、引用句和中心句三个角度进行分析。  
 \n(2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 即在人员长期停留的房间中, 其墙面内表面的可见光反射比应当小于等于 0.8 且大于等于 0.5。因此, 中心句为: “房间-墙面-内表面-可见光反射比-(应当小于等于-0.8)&(应当大于等于-0.5)”。  
 \n(3) 条件句 (Conditional Sentence): 条件句表明了这个标准是在人员长期停留的房间中适用, 因此条件句为: “人员长期停留-房间”。  
 \n(4) 引用句 (Quoted Sentence): 在这个条款中没有引用具体的表或条款或标准, 因此引用句为: “null”。”,

```

      "output": "main_sentence: [房间-墙面-内表面-可见光反射比-(应当小于等于-0.8)&(应当大于等于-0.5)], conditional_sentence: [人员长期停留-房间], quoted_sentence: [null]"

```

},

{

"input": "[在/ctg][严寒 A 区/obj]或[B 区/obj], [体形系数/obj][小于等于/cmp][0.50/value]且[大于/cmp][0.30/value]的[甲类/obj][公共建筑/obj]的[外墙/obj]的[传热系数/obj][应当小于等于/cmp][0.35W/m2K/value]",

"cot": "(1) 为解析这个条款, 需要从条件句、引用句和中心句三个角度进行分析。  
 (2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 即在严寒 A 区或 B 区, 体形系数小于等于 0.50 且大于 0.30 的甲类公共建筑的外墙的传热系数应当小于等于 0.35 W/m<sup>2</sup> · K。因此, 中心句为: “公共建筑-外墙-传热系数-应当小于等于-0.35-W/m2K”。  
 (3) 条件句 (Conditional Sentence): 条件句表明了这个标准是在严寒 A 区或 B 区, 并且体形系数小于等于 0.50 且大于 0.30 的甲类公共建筑的情况下适用, 因此条件句为: “(严寒 A 区/严寒 B 区)&(甲类-公共建筑-体形系数-(小于等于-0.50)&(大于-0.30))”。  
 (4) 引用句 (Quoted Sentence): 在这个条款中没有引用具体的表或条款或标准, 因此引用句为: “null”。

"output": "main\_sentence: [公共建筑-外墙-传热系数-应当小于等于-0.35-W/m2K], conditional\_sentence: [(严寒 A 区/严寒 B 区)&(甲类-公共建筑-体形系数-(小于等于-0.50)&(大于-0.30))], quoted\_sentence: [null]"

},

{

"input": "[建筑高度/obj][大于/cmp][32m/value] 且[任一层/obj][人数/obj][超过/cmp][超过/ctg][10 人 /value]的[厂房/obj], [应采用/cmp][防烟楼梯间/obj]或[室外楼梯/obj]。",

"cot": "(1) 为解析这个条款, 需要从条件句、引用句和中心句三个角度进行分析。  
 (2) 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 即建筑高度大于 32m 且任一层人数超过 10 人的厂房, 应采用防烟楼梯间或室外楼梯。因此, 中心句为: “厂房-应采用-防烟楼梯间/室外楼梯”。  
 (3) 条件句 (Conditional Sentence): 条件句表明了这个标准是在建筑高度大于 32m 且任一层人数超过 10 人的厂房的情况下适用, 因此条件句为: “(厂房-建筑高度-大于-32-m)&(厂房-任一层-人数-超过-10-人)”。  
 (4) 引用句 (Quoted Sentence): 在这个条款中没有引用具体的表或条款或标准, 因此引用句为: “null”。

"output": "main\_sentence: [厂房-应采用-防烟楼梯间/室外楼梯], conditional\_sentence: [(厂房-建筑高度-大于-32-m)&(厂房-任一层-人数-超过-10-人)], quoted\_sentence: [null]"

```

    },
    {
        "input": "[首层/obj][外门/obj]的[最小净宽度/obj][不应小于/cmp][1.20m/value]。",
        "cot": "(1) 为解析这个条款，需要从条件句、引用句和中心句三个角度进行分析。
        \n(2) 中心句（Main Sentence）：中心句阐述了该条款中的主要规定，即首层外门的最小净宽度不应小于 1.20m。因此，中心句为：“首层-外门-最小净宽度-不应小于-1.20”。
        \n(3) 条件句（Conditional Sentence）：这个条款没有明确的条件句，因此可以视为无条件或条件不适用。
        \n(4) 引用句（Quoted Sentence）：在这个条款中没有引用具体的表或条款或标准，因此引用句为：“null”。",
        "output": "main_sentence: [首层-外门-最小净宽度-不应小于-1.20],
        conditional_sentence: [null], quoted_sentence: [null]"
    }
]

```

## 附录 C：设计说明参数提取提示组件（JSON）

```

{
  "role": "你是一个经验丰富的建筑领域暖通方向的审图专家。擅长从已经完成的设计文档和图纸\n 中获取建筑信息和暖通系统设计相关信息。",
  "task": "你的任务是从一份中文暖通系统'设计说明'中获取设计过程中的重要参数信息。",
  "z_cot": "",
  "f_cot": "\n#####以下是输入和输出案例#####\nInput:(二) 工程概况\n上海市某高校学院办公楼为框架结构,总建筑面积 50584 平方米,地上部分建筑面积为 44729 平方米,地下部分建筑面积为 5855 平方米。地上部分由裙房及两栋高度小于 100 米的高层建筑组成。建筑裙房 3 层,功能为门厅、陈列厅等;两栋主楼功能分别为: 1) 学院大楼,地上 17 层,高度为 79.3 米,功能为办公室、会议室等; 2) 另一栋为学术交流中心大楼,地上 13 层,高度为 62.1 米,功能为工作室。
\nOutput:[{'para':'总建筑面积','value':50584,'unit':'m2'},\n{'para':'地上部分建筑面积','value':44729,'unit':'m2'},\n{'para':'地下部分建筑面积','value':5855,'unit':'m2'},\n{'para':'学院大楼建筑高度','value':79.3,'unit':'m'},\n{'para':'学院大楼建筑层数','value':17,'unit':'层'},\n{'para':'学术交流中心大楼建筑高度','value':62.1,'unit':'m'},\n{'para':'学术交流中心大楼建筑层数','value':13,'unit':'层'}]",
  "tot": "",
  "schema": "!!!! 请以 json 格式提取出建筑的信息参数!!!! 其中涉及到的应当参数包括:\n1. 总建筑面积\n2. 地上建筑面积\n3. 地下建筑面积\n4. 建筑层数\n5. 建筑高度\n6. 冬季室外大气压力\n7. 冬季空气调节室外计算温度    \n8. 冬季空气调节室外计算相对湿度\n9. 冬季空调计算热负荷\n10. 夏季空气调节室外计算干球温度\n11. 夏季空气调节室外计算湿球温度\n12. 夏季空调计算冷负荷\n13. 冷热源\n\ta. 地源热泵\n\t\ti. 台数\n\t\tt.ii. 单机负荷\n\t\tb. 螺杆式冷水机组\n\t\t\ti. 台数\n\t\t\tt.ii. 单机负荷\n14. 风系统\n\ta. 风机盘管+新风\n\t\ti. 房间类型\n\t\tt.ii. 送风方式\n\t\tb. 全空气系统\n\t\t\ti. 房间类型\n\t\t\tt.ii. 送风方式\n\t\tc. 分层空调系统\n\t\t\t\ti. 房间类型\n\t\t\t\tt.ii. 送风方式\n15. 水系统\n\ta. 闭式双管式系统\n\t\tb. 夏季供水温度\n\t\tc. 夏季回水温度\n\t\t\td. 冬季供水温度\n\t\t\tte. 冬季回水温度\n\t\t\t\tf. 控制方式: 主机定流量、负荷侧变流量\n16. 保温\n\ta. 保温材料\n\t\tb. 水管

```

保温层厚度\n c. 风管保温层厚度",

"klg": "注意!!! 该设计文档是从 png 图片通过 OCR 识别转化成的文本文档, 其中的一些单位在图像识别过程中会出现错误, 请你自行修正, 例如:\n(1)'C 通常是指摄氏度;\n(2)12C, 通常是指 12 摄氏度;\n(3)6O 中的 O, 通常是指数字 0 而不是字母 O;\n(4)≤45 中的 ≤ 符号常常会被识别成 S 或者 K 等大写字母或者其他和数字不相关的符号;\n(5)有些参数来自于表格 ocr 的直接输出, 有些数字和列名可能会重叠在一起。请你自行分开;"

}

## 参考文献

- [1] Zhang J, El-Gohary N M. Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking. *Autom Constr*, 2017, 73: 45~57
- [2] Eastman C, Lee J min, Jeong Y suk, et al. Automatic rule-based checking of building designs. *Autom Constr*, 2009, 18(8): 1011~1033
- [3] Nawari N O. *Building Information Modeling: Automated Code Checking and Compliance Processes*. Boca Raton: CRC Press, 2018
- [4] Xu X, Cai H. Ontology and rule-based natural language processing approach for interpreting textual regulations on underground utility infrastructure. *Adv Eng Inform*, 2021, 48: 101288
- [5] Automated code compliance checking in the construction domain using semantic natural language processing and logic-based reasoning - ProQuest. <https://www.proquest.com/openview/8fceb4008f140d7277a94ae1452dccb4/1?pq-origsite=gscholar&cbl=18750>
- [6] Fenves S J. Tabular Decision Logic for Structural Design. *J Struct Div*, 1966, 92 (6): 473~490
- [7] Delis E A, Delis A. Automatic Fire-Code Checking Using Expert-System Technology. *J Comput Civ Eng*, 1995, 9 (2): 141~156
- [8] Shu-Hsien Liao. Expert system methodologies and applications—a decade review from 1995 to 2004. *Expert Syst Appl*, 2005, 28 (1): 93~103
- [9] Chollet F. *Deep Learning with Python, Second Edition*. Simon and Schuster, 2021
- [10] Dimyadi J, Amor R. Automated Building Code Compliance Checking – Where is it at?
- [11] Sabouni A R, Al-Mourad O M. Quantitative knowledge based approach for preliminary design of tall buildings. *Artif Intell Eng*, 1997, 11 (2): 143~154
- [12] Ren R, Zhang J. Comparison of BIM Interoperability Applications at Different Structural Analysis Stages. 2020: 537~545
- [13] Wong Chong O, Zhang J. Logic representation and reasoning for automated BIM analysis to support automation in offsite construction. *Autom Constr*, 2021, 129: 103756
- [14] Wu J, Sadraddin H L, Ren R, et al. Invariant Signatures of Architecture, Engineering, and Construction Objects to Support BIM Interoperability between Architectural Design and Structural Analysis. *J Constr Eng Manag*, 2021, 147 (1): 04020148
- [15] Wang J, Mu L, Zhang J, et al. On Intelligent Fire Drawings Review Based on Building Information Modeling and Knowledge Graph. 2020: 812~820
- [16] Li H, Zhang J. Interoperability between BIM and BEM Using IFC. 2022: 630~637
- [17] Zhang L, El-Gohary N M. Automated IFC-based building information modelling and extraction for supporting value analysis of buildings. *Int J Constr Manag*, 2020, 20 (4): 269~288
- [18] Li H, Zhang J, Xue X, et al. Issues in Bi-Directional Interoperability between BIM and BEM. 2022: 1355~1364

- [19] Akanbi T, Zhang J, Lee Y C. Data-Driven Reverse Engineering Algorithm Development Method for Developing Interoperable Quantity Takeoff Algorithms Using IFC-Based BIM. *J Comput Civ Eng*, 2020, 34 (5): 04020036
- [20] Zhang J, Laddipeerla S. A Feasibility Study of IFC-Based BIM 4D Simulation Using Commercial Systems to Support Construction Planning in the U.S.
- [21] Gusmao Brissi S, Wong Chong O, Debs L, et al. A review on the interactions of robotic systems and lean principles in offsite construction. *Eng Constr Archit Manag*, 2021, 29 (1): 383~406
- [22] Lacny C, Zhang J. Computer Vision-Based Geometry Mapping and Matching of Building Elements for Construction Robotic Applications. 2022: 541~549
- [23] Nguyen T H, Kim J L. Building code compliance checking using BIM technology. 见: *Proceedings of the 2011 Winter Simulation Conference (WSC)*. Phoenix, AZ, USA: IEEE, 2011: 3395~3400
- [24] Choi J, Kim I. A Methodology of Building Code Checking System for Building Permission based on openBIM. 2017
- [25] Ma L, Sacks R, Kattel U, et al. 3D Object Classification Using Geometric Features and Pairwise Relationships. *Comput-Aided Civ Infrastruct Eng*, 2018, 33(2): 152~164
- [26] Fenves S J, Gaylord E H, Goel S K. Decision table formulation of the 1969 AISC specification. *Civ Eng Stud SRS-347*, 1969
- [27] Tan X, Hammad A, Fazio P. Automated Code Compliance Checking for Building Envelope Design. *J Comput Civ Eng*, 2010, 24 (2): 203~211
- [28] Fenves S J, Wright R N, Stahl F I, et al. Introduction to SASE ::standards analysis, synthesis, and expression. NIST, 1987
- [29] Macit İlal S, Günaydın H M. Computer representation of building codes for automated compliance checking. *Autom Constr*, 2017, 82: 43~58
- [30] 吉久茂, 童华炜, 张家立. 基于 Solibri Model Checker 的 BIM 模型质量检查方法探究. *土木建筑工程信息技术*, 2014, 6 (01): 14~19
- [31] Automated checking of building requirements on circulation over a range of design phases - ProQuest. <https://www.proquest.com/openview/ca28bb63d0441d31bcb4fb7112796251/1?pq-origsite=gscholar&cbl=18750>
- [32] Soliman J J J, Tzortzopoulos P, Baldauf J P, et al. Automated compliance checking in healthcare building design. *Autom Constr*, 2021, 129: 103822
- [33] Soliman-Junior J, Formoso C T, Tzortzopoulos P. A semantic-based framework for automated rule checking in healthcare construction projects. *Can J Civ Eng*, 2020, 47 (2): 202~214
- [34] Solihin W, Eastman C. A knowledge representation approach in BIM rule requirement analysis using the conceptual graph. *J Inf Technol Constr ITcon*, 2016, 21(24): 370~401
- [35] Preidel C, Borrmann A. Automated Code Compliance Checking Based on a Visual Language and Building Information Modeling. *ISARC Proc*, 2015: 1~8
- [36] Zhang J, El-Gohary N M, Asce A M. Semantic NLP-based Information Extraction from Construction Regulatory Documents for Automated Compliance Checking. 2015: 42
- [37] Zhang J, El-Gohary N. Automated Information Transformation for Automated Regulatory Compliance Checking in Construction. *J Comput Civ Eng*, 2015, 29: B4015001

- [38] Zhou P, El-Gohary N. Ontology-based automated information extraction from building energy conservation codes. *Autom Constr*, 2017, 74: 103~117
- [39] Xu X, Cai H. Semantic approach to compliance checking of underground utilities. *Autom Constr*, 2020, 109: 103006
- [40] Gao J, Lin C Y. Introduction to the special issue on statistical language modeling. *ACM Trans Asian Lang Inf Process*, 2004, 3 (2): 87~93
- [41] Rosenfeld R. Two decades of statistical language modeling: where do we go from here?. *Proc IEEE*, 2000, 88 (8): 1270~1278
- [42] Stolcke A. SRILM-an extensible language modeling toolkit, Seventh international conference on spoken language processing. 2002
- [43] Statistical Language Modeling for Information Retrieval. <https://apps.dtic.mil/sti/citations/ADA440321>
- [44] Thede S M, Harper M P. A Second-Order Hidden Markov Model for Part-of-Speech Tagging, Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics. College Park, Maryland, USA: Association for Computational Linguistics, 1999: 175~182
- [45] Bahl L R, Brown P F, De Souza P V, et al. A tree-based statistical language model for natural language speech recognition. *IEEE Trans Acoust Speech Signal Process*, 1989, 37 (7): 1001~1008
- [46] Brants T, Popat A C, Xu P, et al. Large Language Models in Machine Translation, Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). Prague, Czech Republic: Association for Computational Linguistics, 2007: 858~867
- [47] Katz S. Estimation of probabilities from sparse data for the language model component of a speech recognizer. *IEEE Trans Acoust Speech Signal Process*, 1987, 35(3): 400~401
- [48] Gale W A, Sampson G. Good-turing frequency estimation without tears\*. *J Quant Linguist*, 1995, 2 (3): 217~237
- [49] Bengio Y, Ducharme R, Vincent P. A Neural Probabilistic Language Model. *Advances in Neural Information Processing Systems*: 13. MIT Press, 2000
- [50] Mikolov T, Karafiát M, Burget L, et al. Recurrent neural network based language model. In: *Interspeech 2010. ISCA, 2010*: 1045~1048
- [51] Kombrink S, Mikolov T, Karafiát M, et al. Recurrent neural network based language modeling in meeting recognition. In: *Interspeech 2011. ISCA, 2011*: 2877~2880
- [52] Collobert R, Weston J, Bottou L, et al. Natural Language Processing (Almost) from Scratch. *Nat Lang Process*
- [53] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space. *arXiv*, 2013
- [54] Mikolov T, Sutskever I, Chen K, et al. Distributed Representations of Words and Phrases and their Compositionality, *Advances in Neural Information Processing Systems*: 26. Curran Associates, Inc, 2013
- [55] Peters M E, Neumann M, Iyyer M, et al. Deep contextualized word representations. *arXiv*, 2018

- [56] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv, 2019
- [57] Lewis M, Liu Y, Goyal N, et al. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. arXiv, 2019
- [58] Fedus W, Zoph B, Shazeer N. Switch transformers: scaling to trillion parameter models with simple and efficient sparsity. J Mach Learn Res, 2022, 23 (1): 120:5232-120:5270
- [59] Radford A, Wu J, Child R, et al. Language Models are Unsupervised Multitask Learners
- [60] Liu Y, Ott M, Goyal N, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv, 2019
- [61] Sanh V, Webson A, Raffel C, et al. Multitask Prompted Training Enables Zero-Shot Task Generalization. arXiv, 2022
- [62] Wang T, Roberts A, Hesslow D, et al. What Language Model Architecture and Pretraining Objective Works Best for Zero-Shot Generalization? In: Proceedings of the 39th International Conference on Machine Learning. PMLR, 2022: 22964~22984
- [63] Kaplan J, McCandlish S, Henighan T, et al. Scaling Laws for Neural Language Models. arXiv, 2020
- [64] Wei J, Tay Y, Bommasani R, et al. Emergent Abilities of Large Language Models. arXiv, 2022
- [65] Shanahan M. Talking about Large Language Models. Commun ACM, 2024, 67 (2): 68~79
- [66] Wei J, Wang X, Schuurmans D, et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. arXiv, 2023
- [67] Zhao W X, Zhou K, Li J, et al. A Survey of Large Language Models. arXiv, 2023
- [68] Kaddour J, Harris J, Mozes M, et al. Challenges and Applications of Large Language Models. arXiv, 2023
- [69] Lu Y, Bartolo M, Moore A, et al. Fantastically Ordered Prompts and Where to Find Them: Overcoming Few-Shot Prompt Order Sensitivity. arXiv, 2022
- [70] Webson A, Pavlick E. Do Prompt-Based Models Really Understand the Meaning of their Prompts?. arXiv, 2022
- [71] Dong Q, Li L, Dai D, et al. A Survey on In-context Learning. arXiv, 2023
- [72] Brown T B, Mann B, Ryder N, et al. Language Models are Few-Shot Learners. arXiv, 2020
- [73] Min S, Lyu X, Holtzman A, et al. Rethinking the Role of Demonstrations: What Makes In-Context Learning Work?. arXiv, 2022
- [74] Zhao Z, Wallace E, Feng S, et al. Calibrate Before Use: Improving Few-shot Performance of Language Models. In: Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021: 12697~12706
- [75] Liu J, Shen D, Zhang Y, et al. What Makes Good In-Context Examples for GPT-3?. arXiv, 2021
- [76] Lee Y J, Lim C G, Choi H J. Does GPT-3 Generate Empathetic Dialogues? A Novel In-Context Example Selection Method and Automatic Evaluation Metric for Empathetic Dialogue Generation, Calzolari N, Huang C R, Kim H, et al. Proceedings of the 29th International Conference on Computational Linguistics. Gyeongju, Republic of Korea: International

- Committee on Computational Linguistics, 2022: 669~683
- [77] Levy I, Bogin B, Berant J. Diverse Demonstrations Improve In-context Compositional Generalization. arXiv, 2023
- [78] Su H, Kasai J, Wu C H, et al. Selective Annotation Makes Language Models Better Few-Shot Learners. arXiv, 2022
- [79] Ye X, Iyer S, Celikyilmaz A, et al. Complementary Explanations for Effective In-Context Learning. arXiv, 2023
- [80] Li X, Qiu X. Finding Support Examples for In-Context Learning. arXiv, 2023
- [81] Rubin O, Herzig J, Berant J. Learning To Retrieve Prompts for In-Context Learning. arXiv, 2022
- [82] Zhang Y, Feng S, Tan C. Active Example Selection for In-Context Learning. arXiv, 2022
- [83] Lin Y T, Papangelis A, Kim S, et al. Selective In-Context Data Augmentation for Intent Detection using Pointwise V-Information. arXiv, 2023
- [84] Cobbe K, Kosaraju V, Bavarian M, et al. Training Verifiers to Solve Math Word Problems. arXiv, 2021
- [85] Patel A, Bhattamishra S, Goyal N. Are NLP Models really able to Solve Simple Math Word Problems?. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Online: Association for Computational Linguistics, 2021: 2080~2094
- [86] Miao S Y, Liang C C, Su K Y. A Diverse Corpus for Evaluating and Developing English Math Word Problem Solvers. arXiv, 2021
- [87] Talmor A, Herzig J, Lourie N, et al. CommonsenseQA: A Question Answering Challenge Targeting Commonsense Knowledge. arXiv, 2019
- [88] Geva M, Khashabi D, Segal E, et al. Did aristotle use a laptop? a question answering benchmark with implicit reasoning strategies. Trans Assoc Comput Linguist, 2021, 9: 346~361
- [89] 中华人民共和国住房和城乡建设部. GB 50189-2015 公共建筑节能设计标准[A]. 中华人民共和国, 2015.
- [90] 国家标准化管理委员会. GBT 1.1-2020 标准化工作导则 第1部分: 标准化文件的结构和起草规则[A]. 中华人民共和国, 2020.
- [91] Zhou Y C, Zheng Z, Lin J R, et al. Integrating NLP and context-free grammar for complex rule interpretation towards automated compliance checking. Comput Ind, 2022, 142: 103746
- [92] OpenAI Platform.[EB/OL] (2023)[2024-3-8]. <https://platform.openai.com>
- [93] Prompt Engineering Guide. [EB/OL] .(2023)[2024-3-8]. <https://www.promptinguide.ai/zh/introduction>
- [94] Chen B, Zhang Z, Langrené N, et al. Unleashing the potential of prompt engineering in Large Language Models: a comprehensive review. arXiv, 2023
- [95] Wei J, Bosma M, Zhao V Y, et al. Finetuned Language Models Are Zero-Shot Learners. arXiv, 2022



## 致谢

在 A434 课题组的三年使我成长很多，回身俯望三年前自己跌跌撞撞的进组经历，我只深感幸运。

首先要感谢的是我的导师许鹏教授，许老师学识渊博，为人儒雅，我喜欢和许老师交流。许老师响应我的好奇，顺应我的性格，给了我很多自由，同时也常常教导我要勇于承担随之而来的责任，感谢。

我要感谢我的家人，尤其是我的父母。在研究生期间我逐渐真正意识到了他们一直以来无私的爱，并时常感到受之有愧。

感谢清华大学林佳瑞老师和周育丞学长，他们无私地将先前的研究分享给我，给予了我不少启发和更为清晰的方向；感谢肖桐师姐关于 LLM 的分享给予我启发，使我跳出此前课题被困的泥淖；感谢王鸿鑫师兄和顾洁帆师姐从我研一上手到后续的课题中都不断给予我很多有用的知识和经验；感谢课上课下各位老师的指点迷津。

感谢关航师姐和郭明月师姐在申请季和平日工作生活中的鼓励和照顾；感谢夏壮在学习工作中给予我必要的帮助，她是位很棒的搭档！感谢亚珂从本科到同组求学路上一直的陪伴，祝你一路顺遂；感谢吉林 F4 的过命滑雪之旅，是一生难忘的体验；感谢师弟师妹们平日的陪伴和小礼物。

感谢宋睿治、许泽晗和之前电影社的伙伴陪我看电影、聊电影；感谢徐安、韩威和中华陪我打 CS；感谢我的室友和宿管阿姨（尤其是扫烟头的阿姨）平日对我的包容。

感谢兰鑫宇、舒昶、胡艺宁、韩威、符程皓、刘啸林、冯时、何静、汪宇腾、杨静亭等诸多老友在硕士期间依然给予我陪伴和关心。

以及，其他在硕士生涯中和我产生交错的每一位人，他们多少以直接或间接的方式使我成为了当下的样子。我接受，所以我感谢。

2024 年 3 月



## 个人简历、在读期间发表的学术成果

### 个人简历:

齐梓轩, 男, 1998 年 12 月生。

2021 年 6 月毕业于重庆大学 建筑环境与能源应用工程专业 获学士学位。

2021 年 9 月入同济大学读硕士研究生

### 已发表论文:

- [1] **Qi Z**, Jin R, Li J, et al. BIM-Based Automated Multi-Air Distribution Layout Generation for Office Buildings: A Case Study. Buildings, 2023, 13(7): 1819.
- [2] Xia Z, Guan H, **Qi Z**, Xu P. Multi-Zone Infection Risk Assessment Model of Airborne Virus Transmission on a Cruise Ship Using CONTAM. Buildings. 2023; 13(9):2350.
- [3] 周克楠,陈斐然,**齐梓轩**等.多气候区夜间通风相变蓄冷系统优化设计及控制策略研究[J].建筑节能(中英文),2023,51(05):47-61.
- [4] 周克楠,郭子科,何德卫,陈斐然,**齐梓轩**等.基于机房设备表具图像的非侵入式故障诊断算法[J].建筑节能(中英文),2023,51(06):1-11.

### 研究及项目经历:

- [1] 基于人脸识别和空气检测的报警系统, 课程大作业, 组长。
- [2] 暖通空调设计自动化研究: 风系统设备自动布置和选型, 主要负责人。
- [3] 基于虚拟电厂的需求主动响应与资源分析方法研究, 横向课题, 主要负责人。

## 同济大学学位论文原创性声明

本人郑重声明：所呈交的学位论文《基于自然语言处理（NLP）的暖通规范语义解析方法研究》，是本人在导师指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人创作的、已公开发表或者没有公开发表的作品的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。本学位论文原创性声明的法律责任由本人承担。

学位论文作者签名：

日期： 年 月 日

## 同济大学学位论文授权使用授权书

本人完全了解同济大学关于收集、保存、使用学位论文的规定，同意如下各项内容：按照学校要求提交学位论文的印刷本和电子版本；学校有权保留学位论文的印刷本和电子版，并采用影印、缩印、扫描、数字化或其它手段保存论文；学校有权提供目录检索以及提供本学位论文全文或者部分的阅览服务；学校有权按有关规定向国家有关部门或者机构送交论文的复印件和电子版；允许论文被查阅和借阅。学校有权将本学位论文的全部或部分内 容授权编入有关数据库出版传播，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于（在以下方框内打“√”）：

保密，在\_\_\_\_\_年解密后适用本授权书。

不保密。

学位论文作者签名：

指导教师签名：

日期： 年 月 日

日期： 年 月 日

# 同济大学

# 硕士学位论文答辩

题目： 基于自然语言处理（NLP）的暖通规范语义解析方法研究

硕士生： 齐梓轩

导师： 许鹏 教授

2024年03月19日

# 同济大学硕士学位论文答辩

主席： 苏醒      副教授      同济大学  
委员： 许鹏      教授      同济大学  
      李铮伟      副教授      同济大学  
      叶蔚      副教授      同济大学  
秘书： 曾令杰      助理教授      同济大学  
  
答辩人： 齐梓轩

2024年03月19日



同濟大學  
TONGJI UNIVERSITY

College of Mechanical Engineering

# 基于自然语言处理（NLP）的 暖通规范语义解析方法研究



姓名：齐梓轩

学号：2130276

学位：学术型硕士

导师：许鹏教授

# 目录

## CONTENT

1

引言

2

规范条款自动分类

3

条款词性自动标注

4

基于大语言模型&提示工程的条款解析方法

5

自动合规性检查代码生成方法

6

BIM模型规范自动审查验证与分析

7

结论与展望

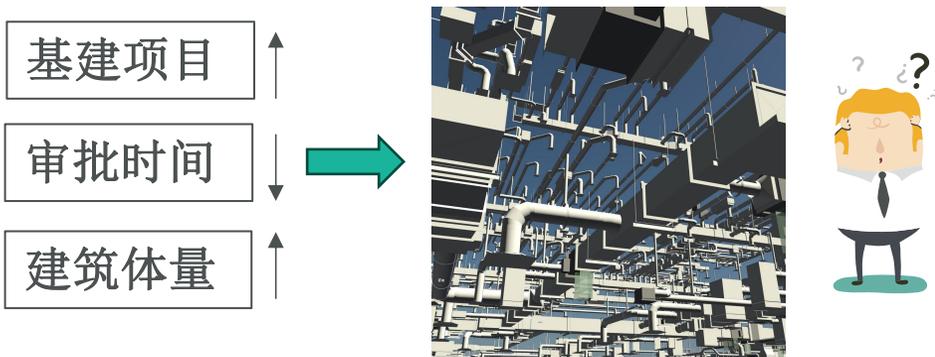
1/7

引言

# 引言-建筑合规性检查

“.....在工程项目设计中，普及应用BIM进行设计方案的性能和功能模拟分析、优化、绘图、审查，以及成果交付和可视化沟通，提高设计质量.....”

——住建部《2016-2020年建筑业信息化发展纲要》



建筑合规性检查 (Automated Compliance Checking) 仍大多依赖于传统的人工方式，比如三维浏览或内部漫游。而传统的手动检查流程不仅效率低下、成本高昂，还易于出错<sup>[1]</sup>。

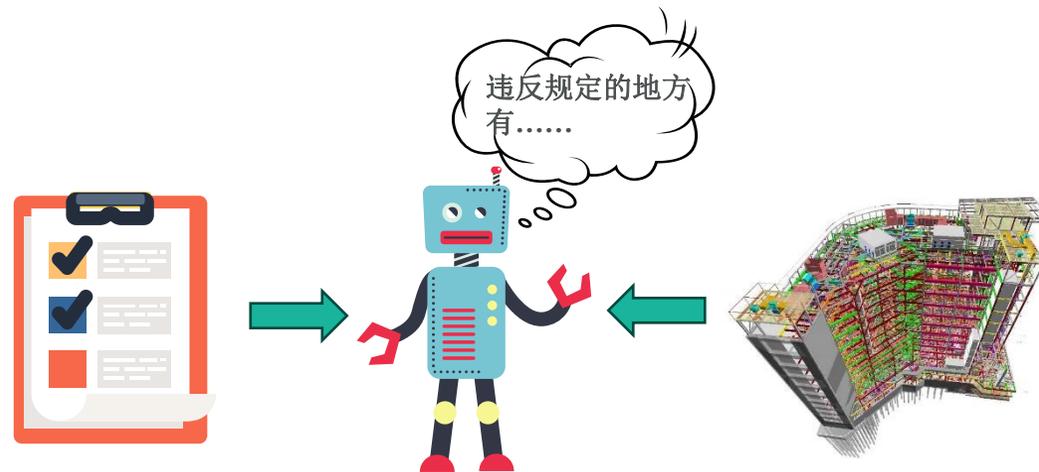
主观理解

熟悉度

注意力

不擅长!

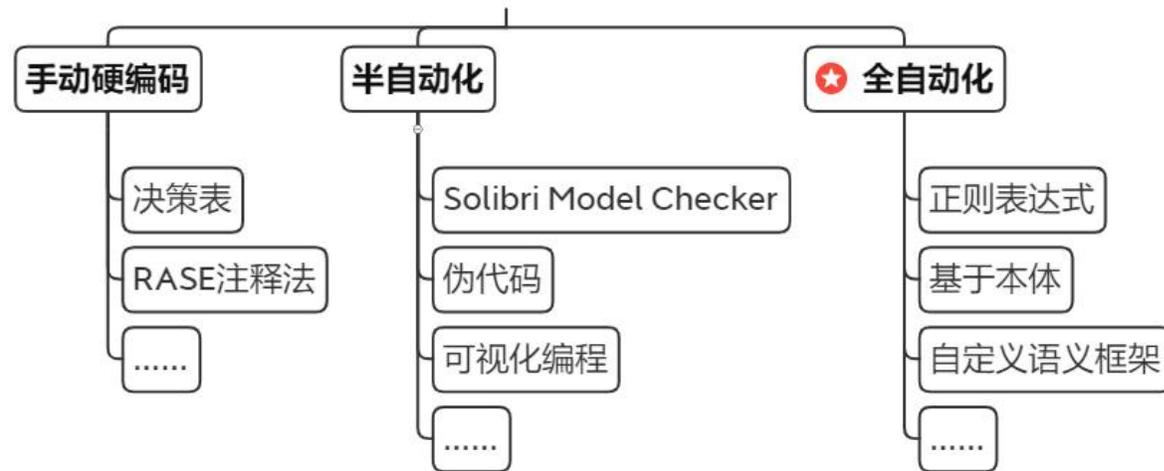
审图人员在审查过程中过分依赖于个人经验和对规范的熟悉度，这不可避免地引入了主观因素，降低了审查的客观性和准确性。



如何告诉机器怎么做?

机器如何执行检查?

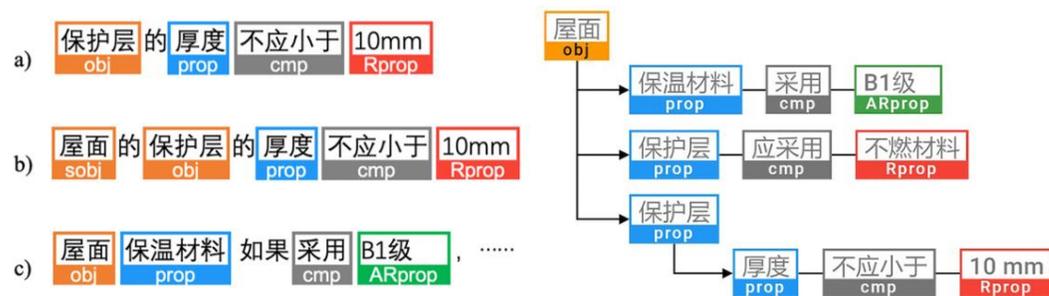
## 规则翻译



[1] Zhang J, El-Gohary N M. Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking. Autom Constr, 2017, 73: 45~57.

# 引言-当下“全自动”方法并不全自动

以往最先进的条款翻译方法[91]:



词性标注：找到句子中的待检查元素和逻辑元素

条款解析：根据语义结构将元素组成逻辑连接

## 缺点：

- 1.即使是先进的条款解析方法，也无法避免手工逻辑编码操作的大量使用。
- 2.依赖于条款语料词性标签训练的语言模型因缺乏统一的数据集和可扩展的规则，导致不同研究之间的标签设置各不相同，进而影响模型训练的效率 and 可迁移性。
- 3.在语义解析部分，尽管简化标签旨在结构化表达，但众多元素组合的解析规则和迭代算法难以全面覆盖所有情况，在遇到复杂句时，条款解析环节往往构成自动合规性审查系统的短板。

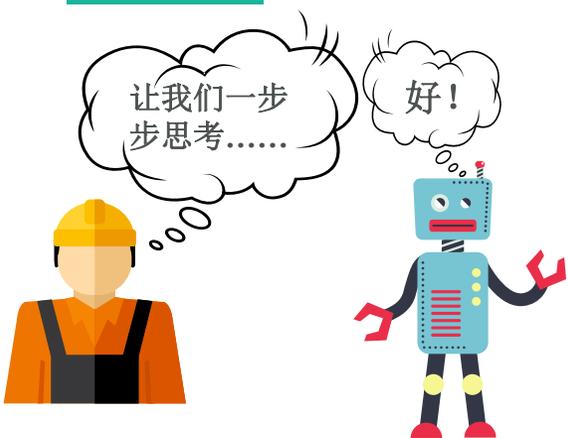
## 当下该方向迫切需要推进的点：

当前建筑自动合规性检查的主要框架在实际应用中存在一系列挑战，这要求进一步研究和开发更加智慧和高效的模型检查自动化方法，以更好地应对复杂多变的审图需求，从而提高建筑设计和审图工作的整体质量及效率。

- 1.当前研究对于复杂条款依然没有实现真正全自动化的规范翻译；
- 2.当前研究很少将最先进的大语言模型应用到建筑自动合规性检查中；
- 3.当前研究很少改善模型参数不可信导致的检查错误；
- 4.研究中标注数据集需要一个普适的条款语义结构；

[91]Zhou Y C, Zheng Z, Lin J R, et al. Integrating NLP and context-free grammar for complex rule interpretation towards automated compliance checking. *Comput Ind*, 2022, 142: 103746





如何让语言模型回答我们希望得到的、有用的答案?  
如何激发大语言模型的知识 and 推理能力?

### 更多提示技巧

### 大语言模型存在的问题:

#### 知识幻觉:

- 不懂装懂, 还会犟嘴;
- 编造规范, 编造数值;
- 泛泛而谈, 说废话;

#### 知识截止:

- 训练时间往后的知识无法知道

#### 训练及其消耗资源:

例如: GPT-3有**1750亿参数**, 在训练的过程中, 批数量最大可达到320万token, 预训练规模使用**3000万亿token**的数据。这些资源对于个人用户来说通常是**不可获得或难以承担的**[67]。

### 直接提问

**模型输入**

**问题:**  
一个房间现在有3个风机盘管。其中, 2个风机盘管下游带有2个送风口, 一个风机盘管带有1个送风口。每一个风机盘管自带一个回风口, 请问, 这个房间内有几个风口?

**模型输出**

这个房间内共有**5个风口**。

### 思维链提示 (零样本)

**模型输入**

**问题:**  
一个房间现在有3个风机盘管。其中, 2个风机盘管下游带有2个送风口, 一个风机盘管带有1个送风口。每一个风机盘管自带一个回风口, 请问, 这个房间内有几个风口?

**让我们一步一步思考。**

**模型输出**

首先, 每个风机盘管自带一个回风口, 所以3个风机盘管共有3个回风口。其次, 2个风机盘管下游带有2个送风口, 所以这两个风机盘管共有4个送风口。最后, 另一个风机盘管带有1个送风口。所以, 这个房间内共有**3个回风口**和**5个送风口**, 合计**8个风口**。

在实际应用中, **提示**是模型的输入, 其不同设计可能导致显著的输出差异[68]。**不同格式 (例如, 改变长度、实例排列方式) 或不同内容 (例如, 措辞、插图选择、指示)**可能会对模型生成的输出产生显著影响[69]。

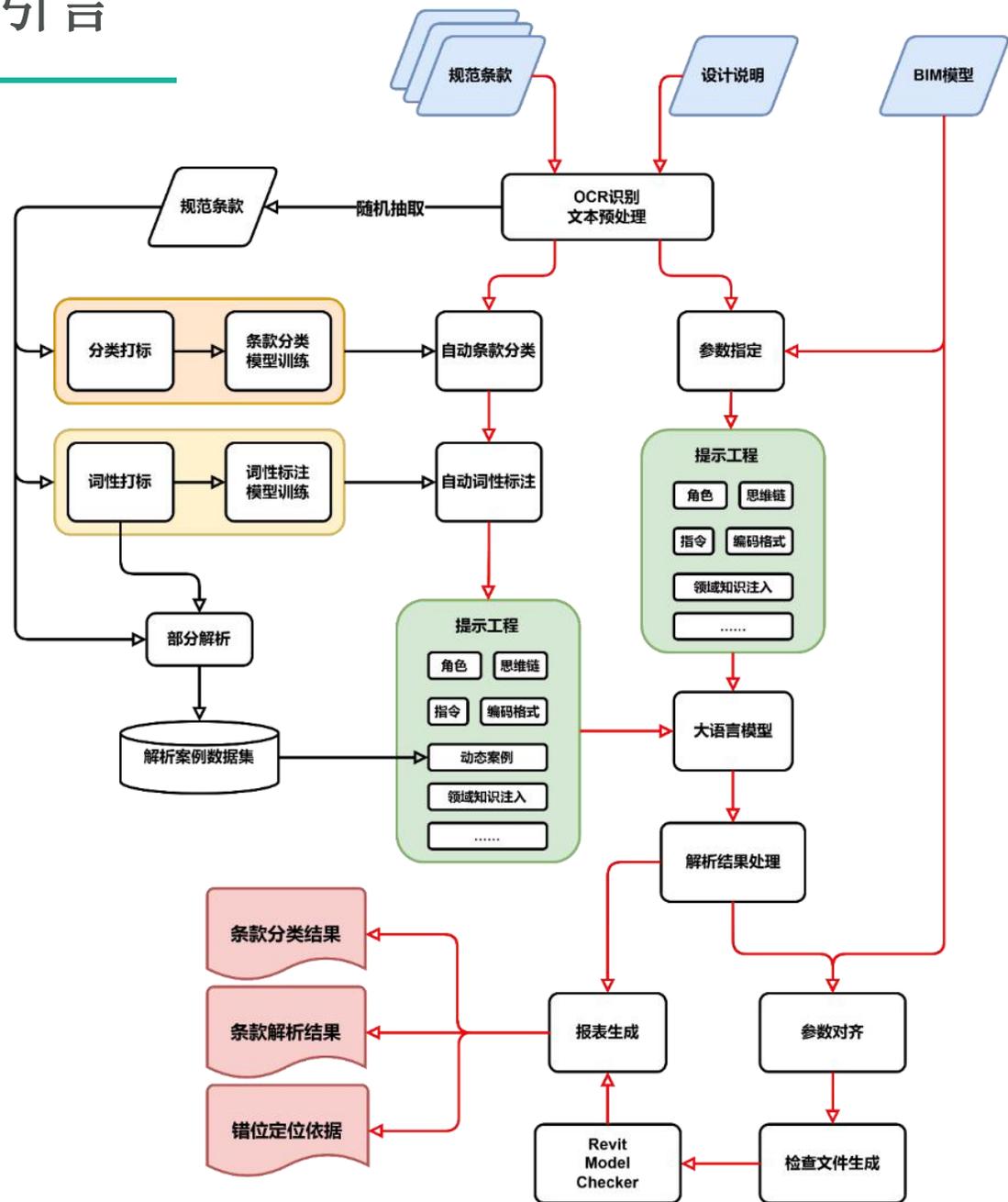


### 引出问题

**不同大语言模型跟不同提示策略如何搭配, 才能进一步帮助我们推进自动的建筑合规性检查呢?**

# 引言

## 研究背景和意义/文献综述/技术路线



## 研究内容:

本文拟采用大语言模型、深度学习、NLP等技术，提出一种全新的规则自动解析方法并构建基于大语言模型的建筑自动合规性检查的全流程可拓展应用框架。本研究主要的研究内容包括：

- (1) 基于此研究方向中词性标注数据集难以互通问题，提出了一种**普适的条款语义结构**；
- (2) 将最先进的大语言模型应用于暖通领域的规范解析，建立并验证基于大语言模型的**建筑自动合规性检查流程框架**。
- (3) 对**主流大语言模型**在自动合规性检查任务中的表现进行**试验测试**，通过具体的评价标准得出此任务下的普适结论以及多个**大语言模型+提示策略的推荐组合**。
- (4) 基于Revit Model Checker软件**开发**出一套少人力工作的**全自动建筑合规性检查系统**。为设计师提供更为全面、可解释性更强的模型检查报表。

2/7

## 规范条款自动分类

- 哪些条款可以被自动化检查？
- 能不能自动把他们分出来？

### 分类标准:

#### 并非所有的规范条文都适合用于模型检查

“1.1 为贯彻国家有关法律法规和方针政策，改善公共建筑的室内环境，提高能源利用效率，促进可再生能源的建筑应用，降低建筑能耗，制定本标准。”

——《公共节能设计标准》



旨在引导设计师进行设计，而不涉及具体的检查对象。同样，还有一些条文涉及到具体操作步骤、公式等条款

4.3.3 在选配集中供暖系统的循环水泵时，应计算集中供暖系统耗电输热比（EHR-h），并应标注在施工图的设计说明中。集中供暖系统耗电输热比应按下式计算：

$$\text{EHR-h} = 0.003096 \Sigma (G \times H / \eta_b) / Q \leq A(B + \alpha \Sigma L) / \Delta T \quad (4.3.3)$$

式中：EHR-h——集中供暖系统耗电输热比；

G——每台运行水泵的设计流量（m<sup>3</sup>/h）；

H——每台运行水泵对应的设计扬程（m）；

$\eta_b$ ——每台运行水泵对应的设计工作点效率；

Q——设计热负荷（kW）；

$\Delta T$ ——设计供回水温差（℃）；

A——与水泵流量有关的计算系数，按本规范表 4.3.9-2 选取；

B——与机房及用户的水阻力有关的计算系数，一级泵系统时 B 取 17，二级泵系统时 B 取 21；

$\Sigma L$ ——热力站至供暖末端（散热器或辐射供暖分集水器）供回水管道的总长度（m）；

$\alpha$ ——与  $\Sigma L$  有关的计算系数；

当  $\Sigma L < 400\text{m}$  时  $\alpha = 0.0115$ 。

#### 专家型条款

定义：这类条款包含了规范所涵盖的前缀指导性思想和操作知识性文本，或者包含了常规BIM模型中难以定义的参数。通常涉及工程实践经验，需要专家进行人工介入。

举例：为贯彻国家有关法律法规和方针政策，改善公共建筑的室内环境，提高能源利用效率，促进可再生能源的建筑应用，降低建筑能耗，制定本标准。

#### 定义型条款

定义：该类型条文规定本规范内容中可能需要用到的专业名词知识性文本。

举例：建筑体形系数 shape factor 建筑物与室外空气直接接触的外表面积与其所包围的体积的比值，外表面积不包括地面和不供暖楼梯间内墙的面积。

#### 属性型条款

定义：该类型条文包含常规BIM模型中所定义的模型参数，比对运算方式直接方便。

举例：甲类公共建筑单一立面窗墙面积比小于 0.40 时，透光材料的可见光透射比不应小于 0.60；

#### 间接型条款

定义：这类条款包含了常规BIM模型中未定义的参数，但所需对比参数可以进行进一步运算或其他方式获取，通常需要进行模型数据增强。

举例：各城市的建筑热工设计分区应按表3.1.2确定。

# 规范条款自动分类

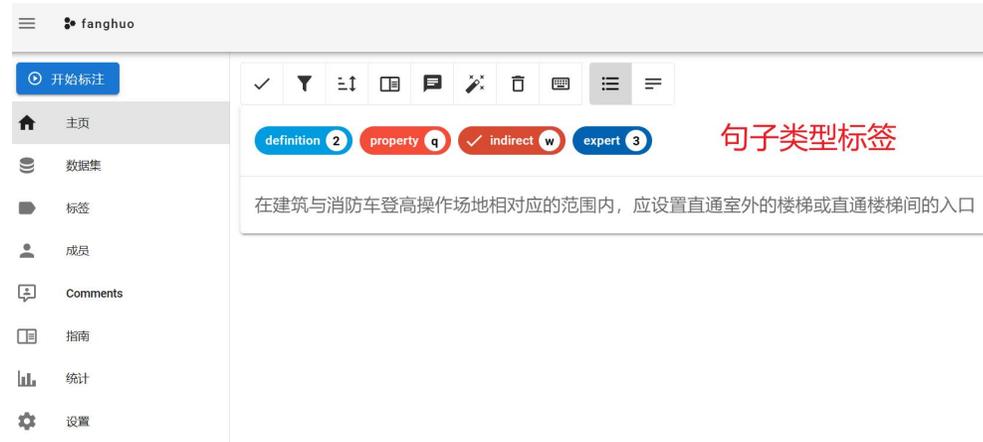
## 规范预处理:

Github第三方库CnSTD  
(<https://github.com/breezedeus/CnSTD/tree/master>)



2 术语  
2.0.1 透光幕墙 transparent curtain wall  
可见光可直接透射入室内的幕墙。  
2.0.2 建筑体形系数 shape factor  
建筑物与室外空气直接接触的外表面积与其所包围的体积的比值,外表面积不包括地面和不供暖楼梯间内墙的面积。  
2.0.3 单一立面窗墙面积比 single facade window to wall ratio  
建筑某一个立面的窗洞口面积与该立面的总面积之比,简称窗墙面积比。  
2.0.4 太阳得热系数 (SHGC) solar heat gain coefficient  
通过透光围护结构(门窗或透光幕墙)的太阳辐射室内得热量与投射到透光围护结构(门窗或透光幕墙)外表面上的太阳辐射量的比值。太阳辐射室内得热量包括太阳辐射通过透明围护结构透射的热量及太阳辐射被构件吸收再传入室内的热量两部分。

## 标签标注:

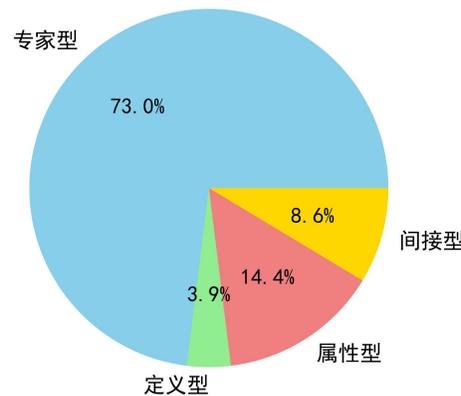


使用doccano平台标注

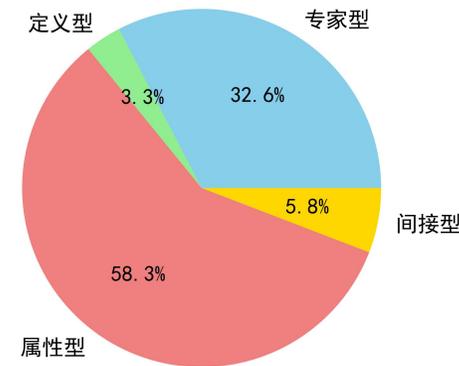


▼ 点击查看条文说明  
10.7 本规范涉及及推广,但也根据各类建筑、设备的防火内容和性能要求,试验方法等全部包括其中,仅对重要性的建筑防火问题的基本的安全需求作了规定,设计采用的产品、材料要有国家有关产品、材料标准的规范,采取的防火技术和措施还要符合国家其他有关工程建设技术标准的规定。  
2术语、符号  
2.1 术语  
2.1.1 高层建筑 high-rise building 建筑高度大于27m的住宅建筑和建筑高度大于24m的非单层厂房、仓库和其他民用建筑。注:建筑高度的计算应符合本规范附录A的规定。  
2.1.2 裙房 podium 在高层建筑主体投影范围外,与建筑主体相连且建筑高度不大于24m的附属建筑。  
2.1.3 重要公共建筑 important public building 发生火灾可能造成重大人员伤亡、财产损失和严重影响社会稳定的公共建筑。  
2.1.4 商业服务网点 commercial facilities 设置在住宅建筑的首层或首层及二层,每个分隔单元建筑面积不大于300m<sup>2</sup>的商店、邮政、快递、理发店等小型营业性用房。  
2.1.5 高架仓库 high rack storage 货架高度大于7m且采用机械化操作或自动化控制的货架仓库。

### 节能条款分类



### 防火条款分类



- 《建筑设计防火规范》(GB50016-2014)
- 《公共建筑节能设计标准》(GB50189-2015)
- 1097例条款

1. 《建筑设计防火规范》中,属性型和间接型条款的比例高达64.1%

2. 《公共建筑节能设计标准》中这两类条款的比例不足25%。

防火规范中的约束条款通常定义更为明确简洁,对于构建的要求往往能具体到具体属性对应的数据要求。对于一个规范来说,其行文和标准越明确,可进行自动检查的部分比例就会更高。

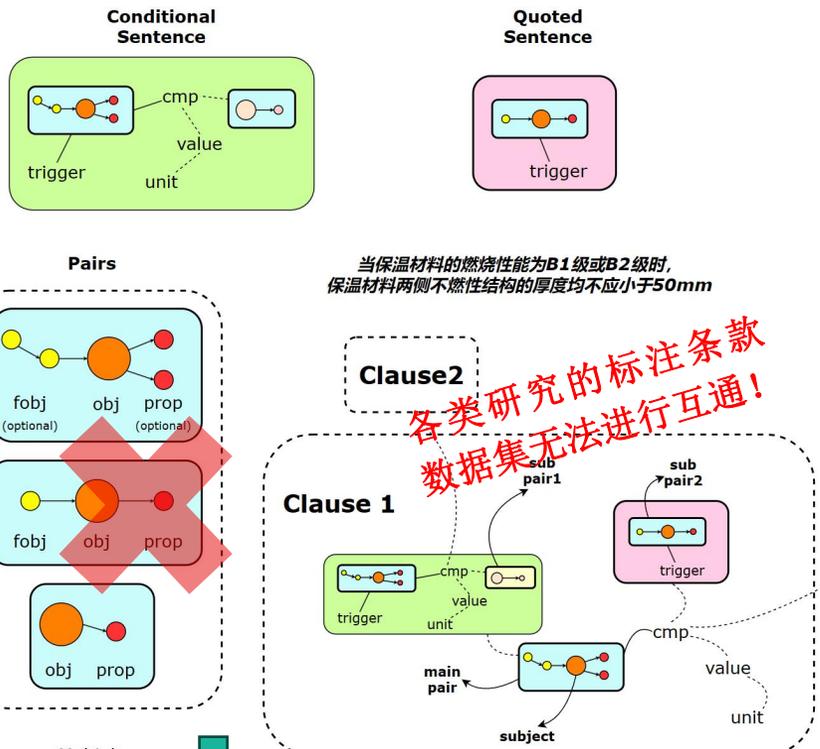


3/7

## 条款自动词性标注

- 大模型解析意味着更普适的语义结构？
- 标注也可以自动化。

## 以往基于规则解析的语义结构:



当保温材料的燃烧性能为B1级或B2级时,  
保温材料两侧不燃性结构的厚度均不应小于50mm

各类研究的标注条款  
数据集无法进行互通!

- 元素定义过于关注条款内部，元素越多，规则越多；
- 忽略了子句之间的关系。直接导致了即使是最先进的ACC系统在复杂句和长难句上的表现都不尽如人意；

## 大语言模型解析情境下:

- 大语言模型在信息提取（Information Extraction）任务中体现出很大的潜力。需要的标签会更少甚至是不需要标签；
- 标签的作用：
  - 更加偏向于句子内子句之间的联系；
  - 为解析能力较弱的大语言模型提供参考；
  - 为解析提供断句依据（如FCU、窗墙比）

### GPT-3.5-无标签

**模型输入**

你的任务是解析现有的中文建筑条款，从提取出条款的主句、条件子句、引用子句。并按照以下格式输出：  
'main\_sentence:[xxx],conditional\_sentence:[xxx],quoted\_sentence:[xxx]'

\*\*\*\*案例\*\*\*\*  
输入：高层厂房和甲、乙、丙类多层厂房的疏散楼梯应采用封闭楼梯间或室外楼梯  
输出：main\_sentence:[(高层-厂房-疏散楼梯-应采用-封闭楼梯间)/(甲、乙、丙类多层-厂房-疏散楼梯-应采用-封闭楼梯间)/(高层-厂房-疏散楼梯-应采用-室外楼梯)](甲、乙、丙类多层-厂房-疏散楼梯-应采用-室外楼梯)],conditional\_sentence:[null],quoted\_sentence:[null]

输入：严寒地区甲类公共建筑各单一立面窗墙面积比(包括透光幕墙)均不宜大于0.60  
输出：

**模型输出**

main\_sentence:[甲类公共建筑-各单一立面窗墙面积比(包括透光幕墙)-均不宜大于0.60]  
conditional\_sentence:[严寒地区]  
quoted\_sentence:[null]

### GPT-3.5-有标签

**模型输入**

你的任务是解析现有的中文建筑条款，从提取出条款的主句、条件子句、引用子句。并按照以下格式输出：  
'[xxx],quoted\_sentence:[xxx]'

\*\*\*\*标签知识\*\*\*\*  
在展示给你输出结构之前，我会给你一些已经设定好的学习材料，希望你在进行解析工作时能记住这些元素的作用！  
标签解释  
1. Obj(object): xxx  
.....

\*\*\*\*案例\*\*\*\*  
输入：高层厂房和甲、乙、丙类多层厂房的疏散楼梯应采用封闭楼梯间或室外楼梯  
输出：main\_sentence:[(高层-厂房-疏散楼梯-应采用-封闭楼梯间)/(甲、乙、丙类多层-厂房-疏散楼梯-应采用-封闭楼梯间)/(高层-厂房-疏散楼梯-应采用-室外楼梯)](甲、乙、丙类多层-厂房-疏散楼梯-应采用-室外楼梯)],conditional\_sentence:[null],quoted\_sentence:[null]

输入：严寒地区甲类公共建筑各单一立面窗墙面积比(包括透光幕墙)均不宜大于0.60  
输出：

**模型输出**

输入：[严寒地区/obj][甲类/obj][公共建筑/obj]各单一立面/obj窗墙面积比/obj(包括透光幕墙)均不宜大于/cmp][0.60/value]  
输出：

### GPT-4-无标签

**模型输入**

你的任务是解析现有的中文建筑条款，从提取出条款的主句、条件子句、引用子句。并按照以下格式输出：  
'main\_sentence:[xxx],conditional\_sentence:[xxx],quoted\_sentence:[xxx]'

\*\*\*\*案例\*\*\*\*  
输入：高层厂房和甲、乙、丙类多层厂房的疏散楼梯应采用封闭楼梯间或室外楼梯  
输出：main\_sentence:[(高层-厂房-疏散楼梯-应采用-封闭楼梯间)/(甲、乙、丙类多层-厂房-疏散楼梯-应采用-封闭楼梯间)/(高层-厂房-疏散楼梯-应采用-室外楼梯)](甲、乙、丙类多层-厂房-疏散楼梯-应采用-室外楼梯)],conditional\_sentence:[null],quoted\_sentence:[null]

输入：严寒地区甲类公共建筑各单一立面窗墙面积比(包括透光幕墙)均不宜大于0.60  
输出：

**模型输出**

main\_sentence:[(严寒地区-甲类公共建筑-各单一立面-窗墙面积比-不宜大于-0.60)]  
conditional\_sentence:[null]  
quoted\_sentence:[null]

**模型输出**

main\_sentence:[(严寒地区-甲类-公共建筑-单一立面-窗墙面积比-不宜大于-0.60)]  
conditional\_sentence:[null]  
quoted\_sentence:[null]

# 条款自动词性标注

## 普适（简化）标签：

标签	说明
obj	条款中出现的建筑元素，通常是一个建筑构件，元素之间的相互连接可能是并列也可能是从属游关系
cmp	比较谓词，表示比较的判断词，被比较的对象可以是数值也可以是obj
value	具体数值，表示某个构建属性的值
qtg	属于trigger的一种，是一种表示引用的触发标签
ctg	属于trigger的一种，是一种表示条件的触发标签

## 数据集制作：

地面辐射供暖面层材料的热阻不宜大于0.05m<sup>2</sup>·K/W。

•obj •obj •obj•cmp •value

建筑空间高度大于等于10m、且体积大于10000m<sup>3</sup>时，宜采用辐射供暖供冷或分层空气调节系统。

•obj•obj •cmp •value •obj •value •cmp •obj  
•cmp •ctg •obj

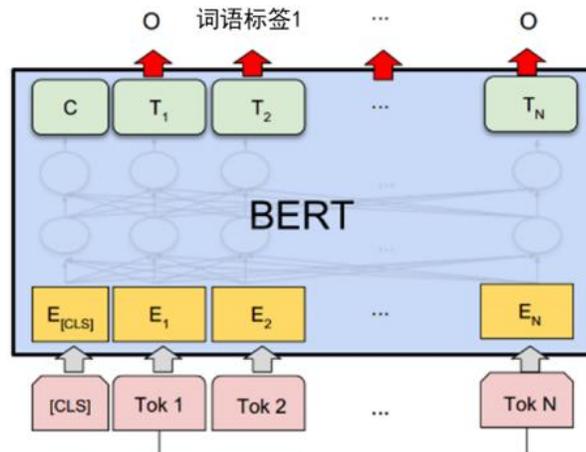
首层外门的最小净宽度不应小于1.20m。

•obj•obj •obj •cmp •value

甲类公共建筑单一立面窗墙面积比大于等于0.40时，透光材料的可见光透射比不应小于0.40。

•obj•obj •obj •obj •cmp •value •obj •obj •cmp •value  
•ctg

## 自动标注微调模型：



## 模型训练结果： 单个句子

标签	准确度	召回率	F1分数	验证数量
B-obj	0.949	0.987	0.967	375
I-obj	0.975	0.993	0.984	1689
B-cmp	0.976	0.960	0.968	126
I-cmp	0.985	0.995	0.990	202
B-qtg	0.667	0.167	0.267	12
I-qtg	0.545	0.429	0.480	28
B-ctg	0.916	0.970	0.942	101
I-ctg	0.905	0.974	0.938	78
B-value	0.962	0.962	0.962	131
I-value	0.962	0.966	0.964	444
O	0.959	0.838	0.895	365
加权平均/合计	<b>0.962</b>	<b>0.963</b>	<b>0.962</b>	<b>3633</b>

```
seq: 送风高度小于或等于5m时，送风温差不宜小于5℃
pred: [送风高度/obj][小于或等于/cmp][5m/value][时/ctg], [送风温差/obj][不宜小于/cmp][5℃/unit]

seq: 散热器宜明装
pred: [散热器/obj][宜/cmp][明装/obj]

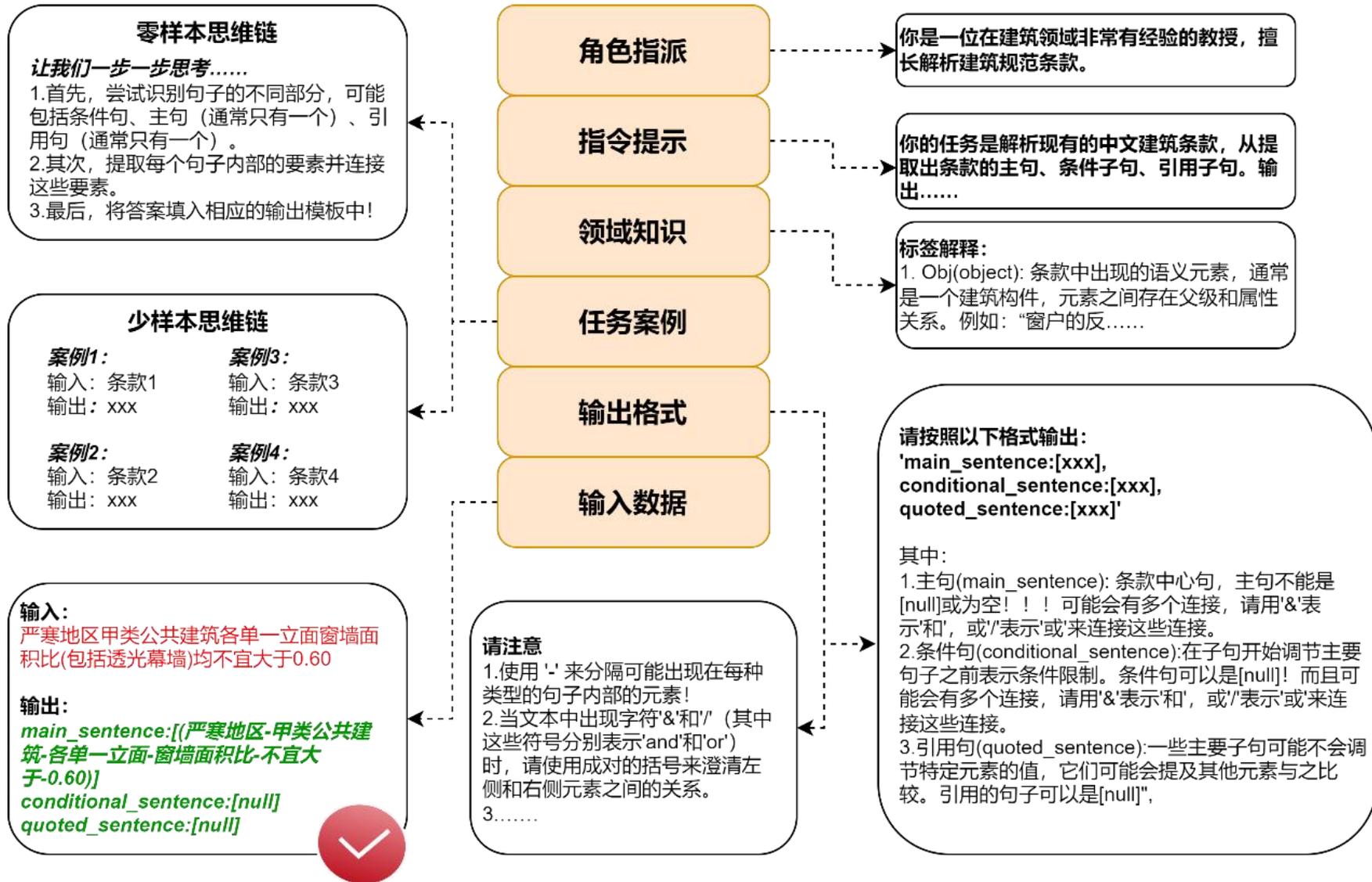
seq: 乙类公共建筑外窗有效通风换气面积不宜小于窗面积的30%
pred: [乙类/obj][公共建筑/obj][外窗/obj][有效通风换气面积/obj][不宜小于/cmp][窗面积/obj]的[30%/value]

完成标签预测，打标耗时：3.01 s
```

4/7

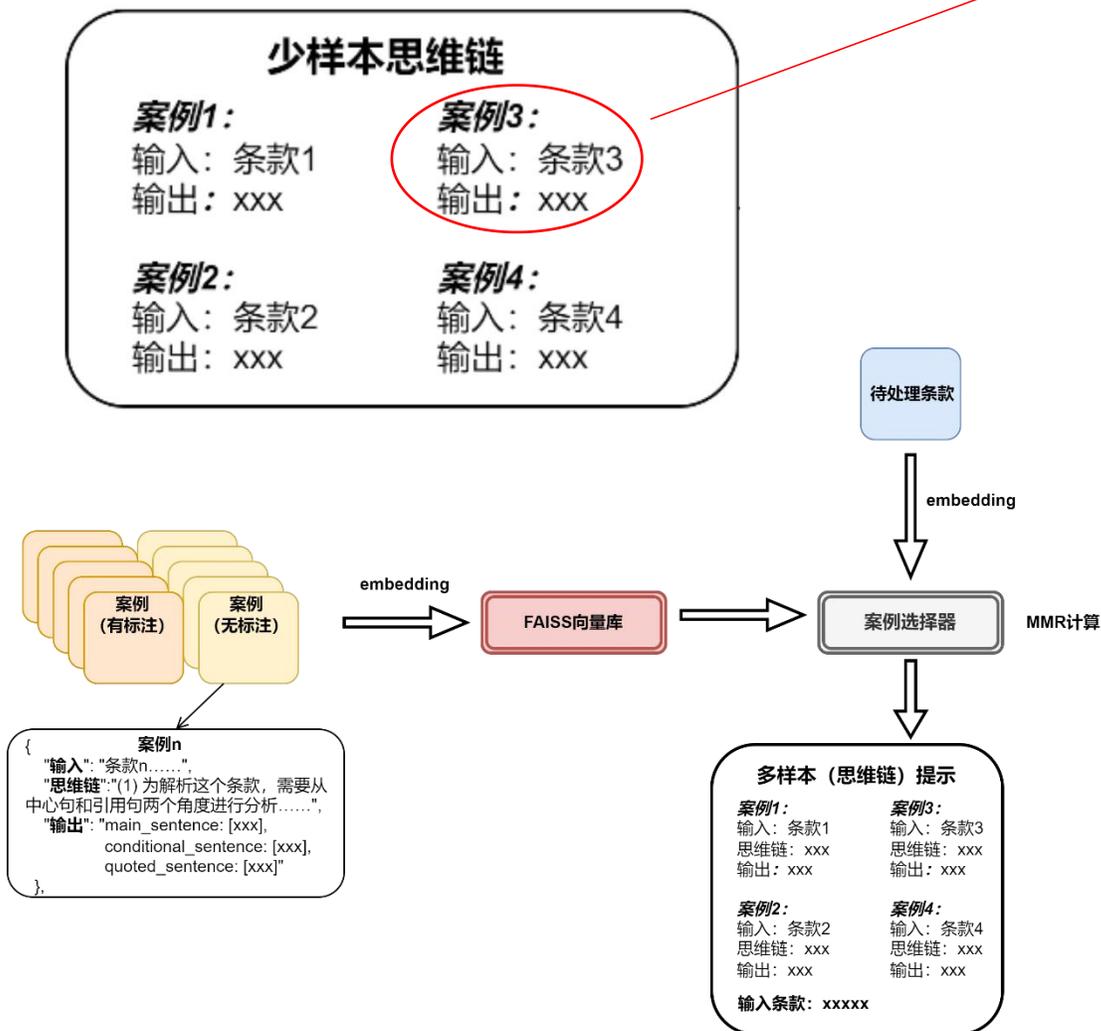
## 基于大语言模型提示工程的条款 解析方法

- 该任务下如何组建提示？
- 不同的大语言模型和什么样的提示组合效果更好？



### 动态案例选择:

### 少样本思维链案例:



- 输入:** “在名义工况和规定条件下, 锅炉的热效率不应低于表4.2.的数值。”
- 思维链:**
  - 为解析这个条款, 需要从条件句、引用句和中心句三个角度进行分析
  - 中心句 (Main Sentence): 中心句阐述了该条款中的主要规定, 本条款最终的目的是规定锅炉的热效率。因此, 中心句在“锅炉的热效率不应低于表4.2.5的数值”这句话中。根据判断, “表4.2.5的数值”为引用。因此, 中心句为: “锅炉-热效率-不应低于-`quote`”
  - 条件句 (Conditional Sentence): 表明这个条款是在特定操作条件下适用, 在这里为“名义工况&名义规定”。
  - 引用句 (Quoted Sentence): 引用的是具体的表或条款或标准, 表示本条款存在引用内容, 在本条款中指代“表4.2.5-数值”。
- 输出:** "main\_sentence: [锅炉-热效率-不应低于-`quote`],  
conditional\_sentence: [名义工况&名义规定], quoted\_sentence: [表4.2.5-数值]"

# 基于大语言模型提示工程的条款解析方法

## 提示词构建/组合实验设计/实验结果

### 主流大模型:

模型	是否开源	参数量 (亿)	发布时间
GPT-3.5	×	175	2022年11月
GPT-4	×	1750	2023年3月
ERNIE-Bot 4.0	×	2600	2023年10月
Qianfan-Chinese-Llama-2-13B	×	130	2023年3月
Baichuan2-7B-Chat	√	70	2023年9月
通义千问	×	720	2023年9月
ChatGLM2-6B	√	60	2023年3月
Ernie-bot 3.5	×	100	2023年3月

### 策略组合:

实验编号	角色	指令	标签	输出编码	标签说明	少样本案例	零样本思维链	少样本思维链
1	√	√	√	√				
2	√	√	√	√	√			
3	√	√	√	√		√		
4	√	√	√	√			√	
5	√	√	√	√				√
6	√	√	√	√	√	√	√	
7	√	√	√	√	√		√	√
8	√	√	√	√	√	√	√	
9	√	√	√	√			√	√
10	√	√		√				
11	√	√		√		√		
12	√	√		√			√	
13	√	√		√				√
14	√	√		√		√	√	
15	√	√		√			√	√
16	√	√				√	√	

通过云服务大模型平台所提供的API在Python环境下进行调用。

由于实验数据组数较大，本课题随机筛选测试集20条规范条款，作为每一组策略下的待解析对象。

- 在阿里云灵积模型平台 (<https://dashscope.console.aliyun.com/model>) 使用通义千问大模型;
- 在openai (<https://platform.openai.com/>) 使用GPT-4、GPT-3.5大模型;
- 在百度智能云千帆平台 (<https://console.bce.baidu.com/qianfan/modelcenter/model/buildIn/list>) 使用文心4大模型、Ernie-bot-3.5大模型、Qianfan-Chinese-Llama-2-13B大模型和ChatGLM2-6B大模型

### 评价结果:

#### 评价指标:

- 指令遵循→True or False
- 格式遵循→True or False
- 分类正确→True or False
- 连接正确→True or False
- 主观评分→分数

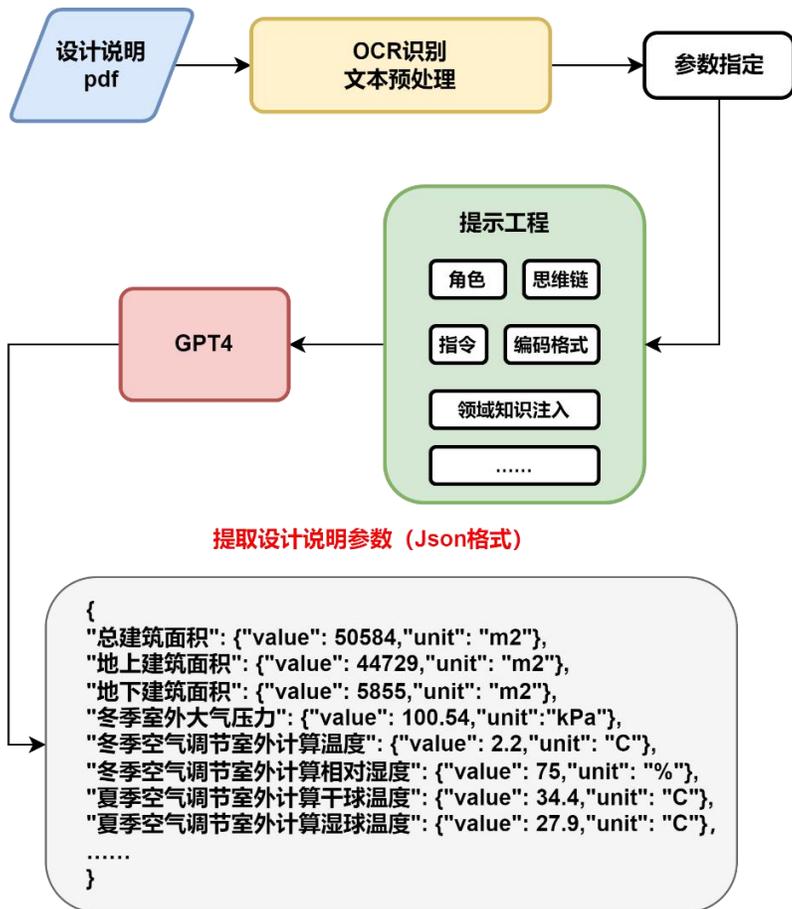
模型表现	代表模型	角色	指令	标签	输出编码	标签说明	少样本案例	零样本思维链	少样本思维链
一般	千帆	√	√		√		√		
良好	Ernie-Bot-3.5、ChatGLM	√	√	√	√	√	√	√	
优秀	GPT-3.5、通义千问、GPT-4、文心4	√	√	√	√		√	√	

#### 基础结论:

- 文心4超越GPT-4，在4项指标中达到最佳；
- 参数量较小的大模型通常在最开始的“指令遵循”中表现即不佳，从而导致后续的解析正确率一落千丈；
- “少样本案例”通常能够作为普适且最为必要的提示策略；
- 对于基础性能较差的大语言模型（百川大模型在解析条款的结果中表现尤其不佳，无法作为比较和推荐的依据，故被舍去），应当以千帆大模型的实验组11进行配置并适当添加零样本思维链，即：角色指派+指令明确+输出编码+少样本案例+（零样本思维链）。
- 对于基础性能良好的大语言模型，应当以实验组7进行配置并根据模型逻辑推理能力适当用少样本案例替换少样本思维链，即：角色指派+指令明确+标签+标签说明+输出编码+少样本思维链（案例）+零样本思维链
- 对于基础性能优异的大语言模型，应当以实验组11进行配置并根据情况适当删除标签说明，即：角色指派+指令明确+输出编码+少样本案例+零样本思维链

模型	指令遵循 (%)	格式遵循 (%)	分类正确 (%)	解析正确 (%)	主观评分 (分)
GPT-3.5	93.13	96.25	90.94	85.00	90.44
GPT-4	93.75	94.06	87.19	93.13	95.47
ERNIE-Bot-4.0	98.75	98.13	97.50	91.25	96.38
Qianfan- Llama-2-13B	69.06	48.06	67.19	43.13	61.10
Baichuan2-7B-Chat	9.38	7.19	8.13	5.31	7.72
ERNIE-Bot-3.5	90.00	92.19	88.44	80.31	85.69
ChatGLM2-6B	64.69	72.50	62.81	58.75	65.00
通义千问	94.75	95.31	93.31	87.19	93.66

实验组	指令遵循%	格式遵循%	分类正确%	解析正确%	主观评分
1	72.78	63.89	72.22	58.33	68.28
2	83.33	73.33	80.00	60.00	75.83
3	78.89	83.33	76.67	73.33	79.56
4	66.67	57.22	63.89	53.33	60.39
5	70.0	71.11	69.44	67.78	68.11
6	86.67	86.11	82.22	76.11	83.89
7	64.44	67.22	63.89	63.33	63.28
8	84.44	85.00	82.78	75.00	81.83
9	66.11	74.44	65.00	63.33	64.39
10	75.00	66.67	71.67	60.00	68.67
11	88.89	88.89	85.00	78.33	86.44
12	83.89	72.78	83.89	66.11	77.72
13	72.22	77.78	70.00	65.00	71.44
14	86.11	87.78	85.56	77.22	85.89
15	73.89	86.11	75.00	71.67	75.89
16	85.00	85.00	85.00	76.67	82.94



### 系统提示:

以下是基于第6组提示策略形成的最终提示, 完整提示详见附录x:

\*\*\*\*角色\*\*\*\*

你是一位经验丰富的建筑领域暖通方向的审图专家。擅长从已经完成的设计文档和图纸中获取建筑信息和暖通系统设计相关信息。

\*\*\*\*指令说明\*\*\*\*

你的任务是从一份中文暖通系统'设计说明'中获取设计过程中的重要参数信息。并按照要要求输出JSON文本。

注意!!! 该设计文档是从png图片通过OCR识别转化成的文本文档, 其中的一些单位在图像识别过程中会出现错误, 请你自行修正:

.....

```
1 {  
2   "总建筑面积": {  
3     "value": 50584,  
4     "unit": "m2"  
5   },  
6   "地上建筑面积": {  
7     "value": 44729,  
8     "unit": "m2"  
9   },  
10  "地下建筑面积": {  
11    "value": 5855,  
12    "unit": "m2"  
13  },  
14  "建筑层数": {  
15    "value": 3,  
16    "unit": "层"  
17  },  
18  "学术交流中心大楼": {  
19    "value": 13,  
20    "unit": "层"  
21  },  
22  "裙房": {  
23    "value": 1,  
24    "unit": "层"  
25  },  
26  "地下建筑": {  
27    "value": 1,  
28    "unit": "层"  
29  },  
30  },  
31  "建筑高度": {  
32    "value": 62.1,  
33    "unit": "m"  
34  },  
35  "学院大楼": {  
36    "value": 79.3,  
37    "unit": "m"  
38  },  
39  "学术交流中心大楼": {  
40    "value": 62.1,  
41    "unit": "m"  
42  },  
43  },  
44 }
```

一、设计说明

(一)设计依据

1. 上海市某高校学院办公建筑平面图2. 国家标准和规范1)《民用建筑供暖通风与空气调节设计规范》(GB50736-2012)2)《建筑设计防火规范》(GB 50016-2014)3)《暖通空调制图标准 GT\*750114-20104)《公共建筑节能设计标准》(GB50189-20155)《建筑照明设计标准》GCB50034-2013) 6)《通风与空调工程施工质量验收规范》(GB50243-20133.其他有关规定以及甲方提出的设计要求

(二)工程概况

上海市某高校学院办公楼为框架结构, 总建筑面积50584平方米, 地上部分建筑面积为44729平方米, 地下部分建筑面积为5855平方米。地上部分由裙房及两栋高度小于100米的高层建筑组成。建筑裙房3层, 功能为门厅、陈列厅等; 两栋主楼功能分别为: 1) 学院大楼, 地上17层, 高度为79.3米, 功能为办公室、会议室等; 2) 另一栋为学术交流中心大楼, 地上13层, 高度为62.1米, 功能为工作室。地下建筑有1层, 为地下车库、变电所及其他配套用房。

(三)设计范围

1.冷热源系统设计(采用复合式地源热泵系统)2. 房1至3层及塔楼标准层暖通空调系统设计3. 地上和地下通风及防排烟系统设计

(四)室外设计参数

冬季室外大气压力100.54kPa夏季室外大气压力102.54kPa冬季空气调节室外计算温度2.2C冬季空气调节室外计算相对湿度75%夏季空气调节室外计算干球温度34.4 C夏季空气调节室外计算湿球温度27.9°C

(五)室内设计参数

设计参数夏季冬季房间类型温度C湿度%温度C湿度%噪声(dB)A办公室26602045K45工作室26602045K45会议室26601845s4多功能厅26601845K45阅览室26602045K45报告厅26601845K45陈列厅2651845K45大厅26601845545咖啡厅26601845K45餐厅2660184535走廊26601845550注:其他类型房间的室内参数均如上表所示

(六)空调计算负荷

夏季空调计算冷负荷为2863.1W, 冬季空调计算热负荷为1735kW。

(七)冷热源方案及选择

1.本工程的冷热源采用复合式地源热泵系统。经计算, 该建筑冷负荷大于热负荷, 如果仅采用土壤作为冷为热源将会使得夏季排热量高于冬季取热量, 长时间运行的结果是导致土壤温度升高, 失去吸热能力。故本系统采用并联冷却塔的方式平衡土壤夏季热量不均造成的问题, 土壤埋管换热器按照冬季供热能力设计确定, 夏季剩余冷量由冷却塔承担。2. 选用两台螺杆式地源热泵机组和一台螺杆式冷水机组。地源热泵机组和螺杆式冷水机组的制冷/热量分别为



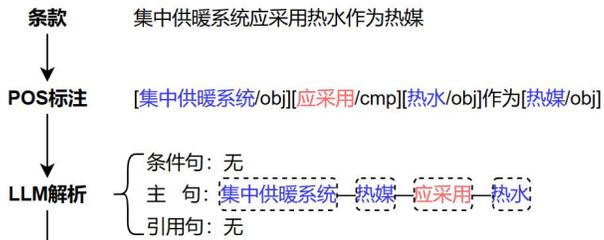
5/7

## 自动合规性检查代码生成方法

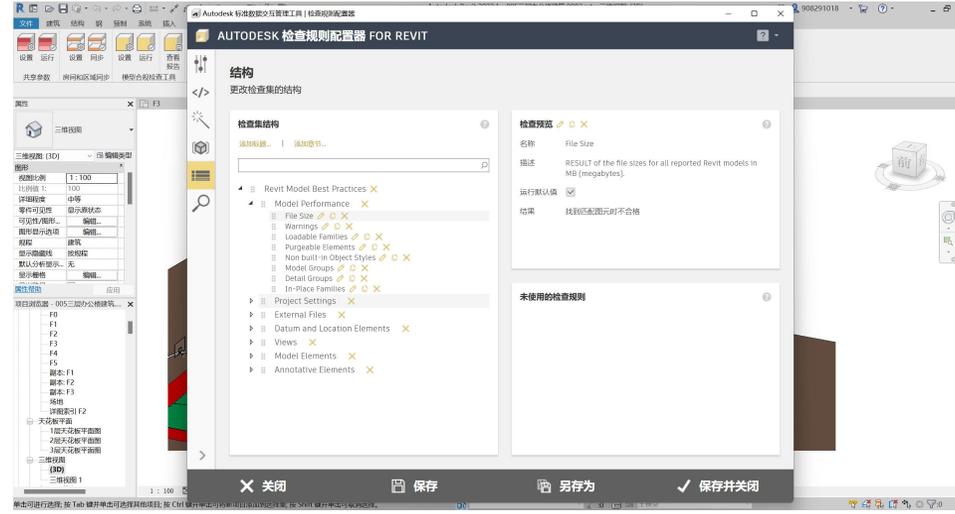
- 模型检查本质是一个数据核对的工作。
- 逻辑→检查代码生成（基于Revit Model Checker）

# 自动合规性检查代码生成方法

# Revit Model Checker/自动检查代码生成



```
<Check
CheckName="集中供暖系统应采用热水作为热媒"
ResultCondition="FailMatchingElements">
  <Filter
  Operator="And"
  Category="Category"
  Property="OST_MechanicalEquipment"
  Condition="included"
  Value="True"/>
  <Filter
  Operator="And"
  Category="Parameter"
  Property="热媒"
  Condition="Contains"
  Value="热水"/>
</Check>
```



类别/参数名	正则表达式	说明
OST_Doors	门\$	门
OST_Stairs	楼梯\$	楼梯
OST_Windows	窗\$ ^窗 透光材料	窗、幕墙
OST_Walls	墙\$ ^墙	墙
OST_Floors	楼板\$	楼板
OST_MassShade	遮阳\$	遮阳
OST_Sheets	(甲 乙)?类 严寒 夏热冬(冷 暖)?温和	气候区
OST_Rooms	(房 空)?间	房间
OST_DuctTerminal	(送 回)?风	风管、风口
OST_MechanicalEquipment	散热器 锅炉 供暖系统	设备、系统

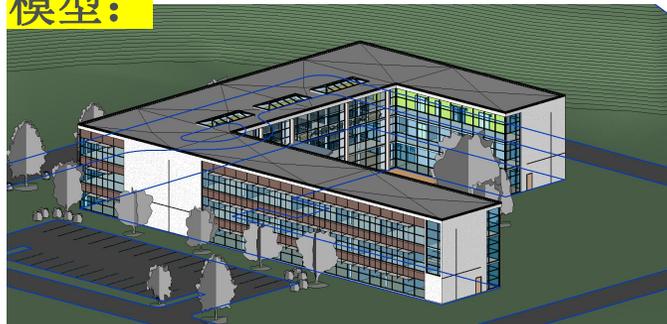
符号	正则表达式	说明
≤	小于或等于 不大于 高于 多于 超过	小于等于
≥	大于或等于 不小于 低于 少于	大于等于
>	大于 超过 高于	大于
<	小于 低于	小于
≠	不等于 避免 (不 无法 严禁)采用 非 除	不等于
=	等于 为采?用 按照?符合 执行	等于
has no	(不 严禁)(有 设置?) 具备) 无	不含
has	(置 有?) (增 铺 搭)设 安装 具备 采取	包含
not in	(不 除 严禁)在	不在
in	在	在

6/7

## BIM模型规范自动审查验证与分析

- 条款→分类→标注→解析→代码→报表

### 模型:



来源: Revit自带模型  
 建筑面积5121.26m<sup>2</sup>  
 办公室、会议室、休息区等



来源: 实际工程  
 建筑面积8872.17m<sup>2</sup>  
 办公室、会议室、宿舍、食堂等



来源: 实际工程  
 建筑面积16805.43 m<sup>2</sup>  
 办公室、会议室、礼堂、控制室、等待区等

**条款:** 15条来自于验证数据集条款。

序号	说明	内容	分类	评价
1	原文	为贯彻国家有关法律法规和方针政策, 改善公共建筑的室内环境, 提高能源利用效率, 促进可再生能源的建筑应用, 降低建筑能耗, 制定本标准。	专家型	正确
2	原文	建筑体形系数shape factor建筑物与室外空气直接接触的外表面积与其所包围的体积的比值, 外表面积不包括地面和不供暖楼梯间内墙的面积。	定义型	正确
3	原文	参照建筑reference building进行围护结构热工性能权衡判断时, 作为计算满足标准要求的全年供暖和空气调节能耗用的基准建筑。	定义型	正确
4	原文	独栋建筑面积小于或等于300m <sup>2</sup> 的建筑, 应为乙类公共建筑。	定义型	正确
5	原文	建筑总平面设计及平面布置应合理确定能源设备机房的位置, 缩短能源供应输送距离。同一公共建筑的冷热源机房宜位于或靠近冷热负荷中心位置集中设置。	专家型	正确
6	原文	严寒地区甲类公共建筑各单一立面窗墙面积比(包括透光幕墙)均不宜大于0.60	属性型	正确
	标注	[严寒地区/obj][甲类/obj][公共建筑/obj]各[单一立面/obj][窗墙面积比/obj](包括透光幕墙)均[不宜大于/cmp][0.60/value]		
	解析	<b>主句:</b> [严寒地区-甲类-公共建筑-各单一立面-窗墙面积比-均不宜大于-0.60], <b>条件句:</b> [null], <b>引用句:</b> [null]		
7	原文	甲类公共建筑单一立面窗墙面积比小于0.40时, 透光材料的可见光透射比不应小于0.60	属性型	正确
	标注	[甲类/obj][公共建筑/obj][单一立面/obj][窗墙面积比/obj][小于/cmp][0.40/value][时/ctg], [透光材料/obj]的[可见光透射比/obj][不应小于/cmp][0.60/value]		
	解析	<b>主句:</b> [透光材料-可见光透射比-不应小于-0.60], <b>条件句:</b> [甲类-公共建筑-单一立面-窗墙面积比-小于-0.40], <b>引用句:</b> [null]		

## 条款（续）：

序号	说明	内容	分类	评价
13	原文	210层以下建筑外窗的气密性不应低于6级；	属性型	错误
	标注	[210层/value][以下/ctg][建筑/obj][外窗/obj]的[气密性/obj][不应低于/cmp][6级/value]		
	解析	主句: [外窗-气密性-不应低于-6-级], 条件句: [建筑-低于-210-层], 引用句: [null]		
14	原文	耐火等级为三级的公共建筑，其柱的耐火极限不低于2h	属性型	正确
	标注	[耐火等级/obj][为/cmp][三级/value]的[公共建筑/obj], [其/obj][柱/obj]的[耐火极限/obj][不低于/cmp][2h/value]		
	解析	主句: [公共建筑-柱-耐火极限-不低于-2-h], 条件句: [公共建筑-耐火等级-为-三-级], 引用句: [null]		

## 问题分析：



1. 标注结果显示其中的“以下”是作为“ctg”即“条件触发标签”，但却理解成了“cmp”即“比较”谓词的属性，这可能和模型训练过程中此类词语的训练标签较为集中导致模型在这种场景下欠拟合导致；
2. 把“[建筑外窗/obj]”错误地分成了“[建筑/obj][外窗/obj]”进而导致了该条款后续的解析错误。
3. 模型对条款原意进行了曲解。出现这个问题的原因可能时由于“ctg”。其内涵被曲解成了当满足建筑层数低于210层时，建筑外窗的气密性应当满足一定要求。

# BIM模型规范自动审查验证与分析

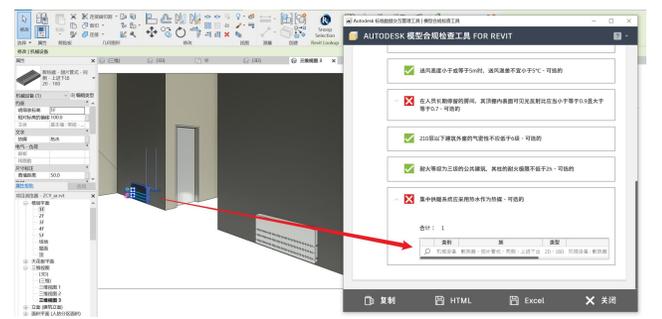
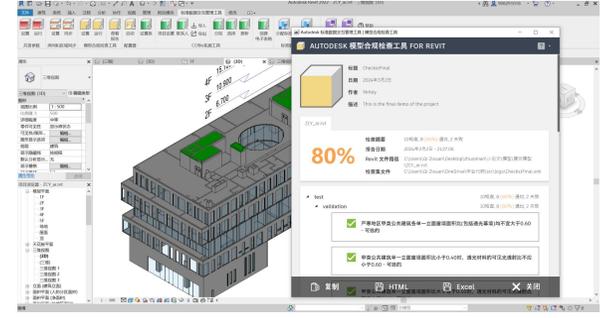
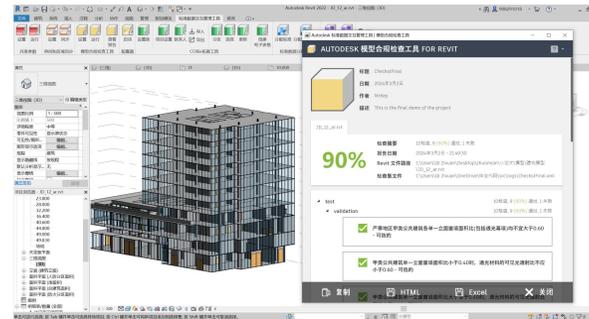
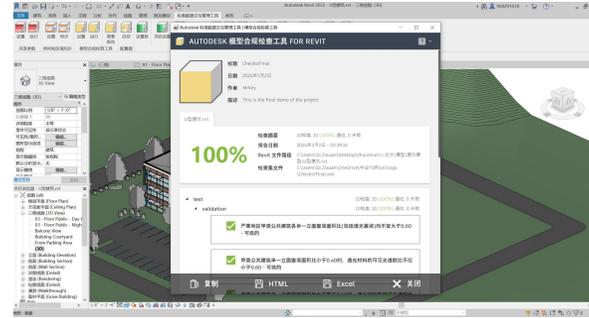
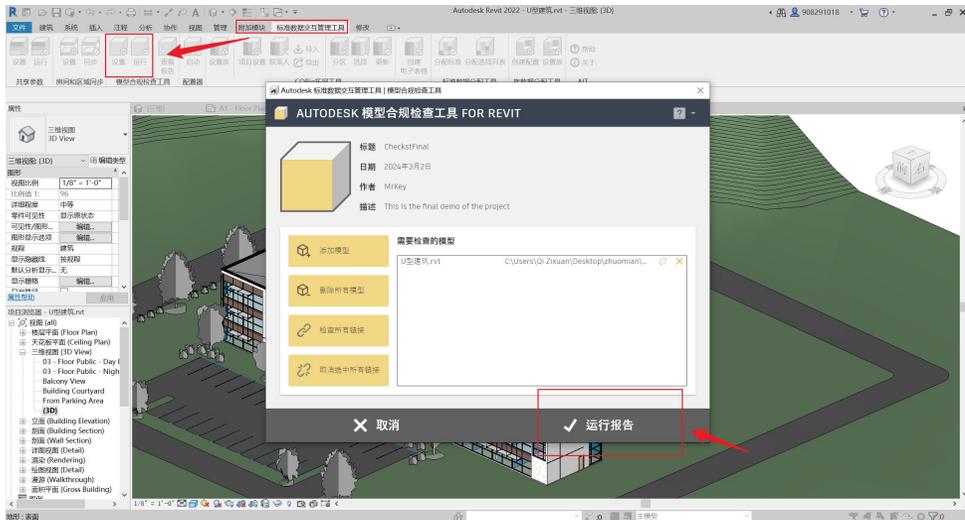
# 案例概览/规范解析结果/自动检查结果

## 可视化页面（错误案例）+最终报告：

```

1 <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2 <CSettings AllowRequired="True" Name="ChecksFinal" Author="MrKey" Description="This is the final demo of the project" Image="" LastModified=
3 "13558558739839281">
4 <Heading ID="bd42d07-4c55-4bcf-b13a-9ef1152a89" HeadingText="411;--Unused-4gt;" Description="Checks that have been created but not placed in
5 the structure yet." IsChecked="True">
6 <Section ID="9189d47-6ee7-44e5-9113-70c9a59864" SectionName="611;--Unused-4gt;" Title="Checks that have not been placed" IsChecked="True"
7 Description="" />
8 </Heading>
9 <Heading ID="95505cc1-9564-4625-b975-bb2ded2311b" HeadingText="test" Description="" IsChecked="True">
10 <Section ID="0f7d444f-23da-4dd0-9fe2-95df2e250b1" SectionName="validation" Title="" IsChecked="True" Description="">
11 <Check ID="4731f9b5-41c1-4ba6-bce2-1df6624791d" CheckName="严寒地区甲类公共建筑各单一立面窗墙面积比(包括透光幕墙)均不宜大于0.60" Description=
12 "" FailureMessage="" ResultCondition="FailMatchingElements" CheckType="Custom" IsRequired="False" IsChecked="True">
13 <Filter ID="1f811be8-5247-4d53-bd50-9218b1345ae" Operator="And" Category="Category" Property="OST_Walls" Condition="Included" Value="True"
14 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
15 <Filter ID="2ebc770a-d593-4e2a-92a4-17ee6a87821" Operator="And" Category="Parameter" Property="窗墙比" Condition="GreaterThan" Value="0.6"
16 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
17 </Check>
18 <Check ID="6211484-43dd-40cb-91e8-e92076981d46" CheckName="甲类公共建筑单一立面窗墙面积比小于0.40时,透光材料的可见光透射比不应小于0.60"
19 Description="" FailureMessage="" ResultCondition="FailMatchingElements" CheckType="Custom" IsRequired="False" IsChecked="True">
20 <Filter ID="7d6445d8-343b-4ba3-978-9d52278a5cd" Operator="And" Category="Category" Property="OST_Walls" Condition="Included" Value="True"
21 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
22 <Filter ID="4906877-648b-4344-84f7-3af1c5115bbd" Operator="And" Category="Parameter" Property="窗墙比" Condition="LessThan" Value="0.4"
23 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
24 <Filter ID="5922427d-1191-402a-950c-8069039524b" Operator="And" Category="Category" Property="OST_Windows" Condition="Included" Value="True"
25 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
26 <Filter ID="2be0705-30b4-4e1b-bd86-03d628099ef" Operator="And" Category="Parameter" Property="ANALYTICAL_VISUAL_LIGHT_TRANSMITTANCE"
27 Condition="LessThan" Value="0.6" CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
28 </Check>
29 <Check ID="3f4d0ba0-b76e-4db9-abd7-9f4336aad27" CheckName="甲类公共建筑单一立面窗墙面积比大于等于0.40时,透光材料的可见光透射比不应小于0.40"
30 Description="" FailureMessage="" ResultCondition="FailMatchingElements" CheckType="Custom" IsRequired="False" IsChecked="True">
31 <Filter ID="a29375f8-433a-4981-b6f0-53f1a18e" Operator="And" Category="Category" Property="OST_Walls" Condition="Included" Value="True"
32 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
33 <Filter ID="b33d2a23-39b9-4ef2-8c5d-e08ac85e8f9" Operator="And" Category="Parameter" Property="窗墙比" Condition="GreaterOrEqual" Value="0.4"
34 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
35 <Filter ID="7d5baa7-0bbe-40e4-898a-d1eb89c7c5" Operator="And" Category="Category" Property="OST_Windows" Condition="Included" Value="True"
36 CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
37 <Filter ID="1e0c66e1-2392-4213-b747-5d5a5ed53a" Operator="And" Category="Parameter" Property="ANALYTICAL_VISUAL_LIGHT_TRANSMITTANCE"
38 Condition="LessThan" Value="0.4" CaseInsensitive="False" Unit="None" UnitClass="None" FieldTitle="" UserDefined="False" Validation="None" />
39 </Check>
40 <Check ID="f5205a27-1e4c-4775-81e5-642ea474b1b7" CheckName="乙类公共建筑外窗有效通风换气面积不宜小于窗面积的30%" Description="" FailureMessage=

```



序号	内容	结果	错误	计数	分类	标注	解析	程度词
1	为贯彻落实国家有关法律法规和方针政策,改善公共建筑的室内环境,提高能源利用效率,促进可再生能源的建筑应用,降低建筑能耗,制定本标准。	人工介入			专家型			
2	建筑体形系数shape factor建筑物与室外空气直接接触的外表面积与其所包围的体积的比值。外表面积不包括地面和不采暖楼梯间凸出部分的面积				定义型			
3	参照建筑reference building进行围护结构热工性能判断时,作为计算满足标准要求的全年供暖和空气调节能耗用的基准建筑。				定义型			
4	单栋建筑面积小于或等于300m²的建筑,应为乙类公共建筑。				定义型			
5	建筑节能设计平面和屋面合理节能设备机房的位,应能确保提供相应配。同一公共建筑的冷热源机房宜位于或靠近冷热负荷中心位置集中设置。	人工介入			专家型			
6	严寒地区甲类公共建筑各单一立面窗墙面积比(包括透光幕墙)均不宜大于0.60	通过	0		属性型	[严寒地区/obj]甲类/obj]公共建筑/obj]有(单一立面/obj]窗墙面积比/obj]比透光幕墙均不宜大于/cmp[0.60/value]	main_sentence: [严寒地区-甲类-公共建筑-各单一立面-窗墙面积比-均不宜大于-0.60] conditional_sentence: [严寒地区-公共建筑-单一立面-窗墙面积比-小于-0.60] quoted_sentence: [null]	不宜
7	甲类公共建筑单一立面窗墙面积比小于0.40时,透光材料的可见光透射比不应小于0.60	通过	0		属性型	[甲类/obj]公共建筑/obj]单一立面/obj]窗墙面积比/obj]小于/cmp[0.40/value]时/ctg. [透光材料/obj]的可见光透射比/obj]不应小于/cmp[0.60/value]	main_sentence: [透光材料-可见光透射比-不应小于-0.60] conditional_sentence: [甲类-公共建筑-单一立面-窗墙面积比-小于-0.40]	不应
8	甲类公共建筑单一立面窗墙面积比大于等于0.40时,透光材料的可见光透射比不应小于0.40	通过	0		属性型	[甲类/obj]公共建筑/obj]单一立面/obj]窗墙面积比/obj]大于等于/cmp[0.40/value]时/ctg. [透光材料/obj]的可见光透射比/obj]不应小于/cmp[0.40/value]	main_sentence: [透光材料-可见光透射比-不应小于-0.40] conditional_sentence: [甲类-公共建筑-单一立面-窗墙面积比-大于等于-0.40] quoted_sentence: [null]	不应
9	乙类公共建筑外窗有效通风换气面积不宜小于窗面积的30%	通过	0		间接型	[乙类/obj]公共建筑/obj]外窗/obj]有效通风换气面积/obj]不宜小于/cmp[窗面积/obj]的30%/value]	main_sentence: [乙类-公共建筑-外窗-有效通风换气面积-不宜小于-窗面积-30%] conditional_sentence: [null] quoted_sentence: [null]	不宜
10	散热器宜明装	通过	0		属性型	[散热器/obj]宜/cmp[明装/obj]	main_sentence: [散热器-宜-明装] conditional_sentence: [null] quoted_sentence: [null]	宜
11	送风高度小于或等于5m时,送风温差宜小于5℃	通过	0		属性型	[送风高度/obj]小于等于/cmp[5m/value]时/ctg. [送风温差/obj]不宜小于/cmp[5℃/value]	main_sentence: [送风高度-小于或等于-5m] conditional_sentence: [送风温差-不宜小于-5℃] quoted_sentence: [null]	不宜
12	在人员长期停留的房间,其顶棚内表面可见光反射比应大于等于0.9且	通过	0		属性型	[在/ctg]人员长期停留的房间/obj]. [其/obj]顶棚/obj]的可见光反射比/obj]应大于等于/cmp[0.9/value]	main_sentence: [房间-顶棚-内表面-可见光反射比-应当大于等于-0.9] conditional_sentence: [null]	应

A teal diamond shape is positioned on the left side of the slide, partially overlapping a dark grey horizontal bar. Inside the diamond, the number '7/7' is written in white, bold, sans-serif font.

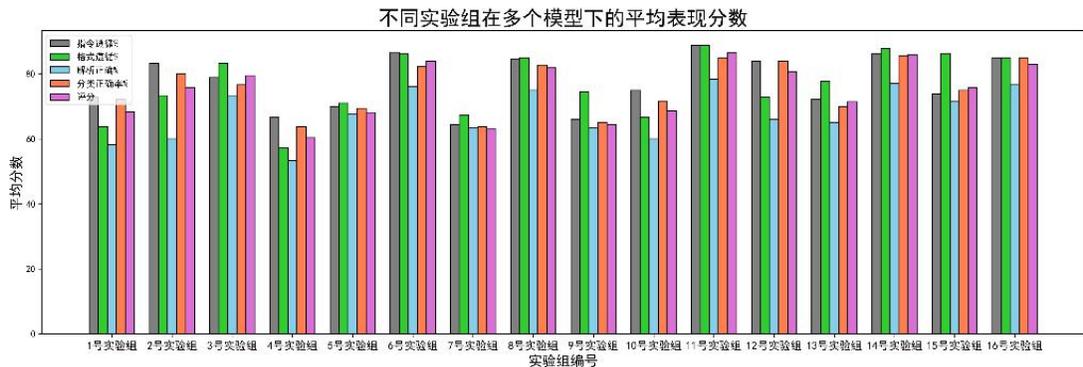
7/7

总结与展望

# 总结

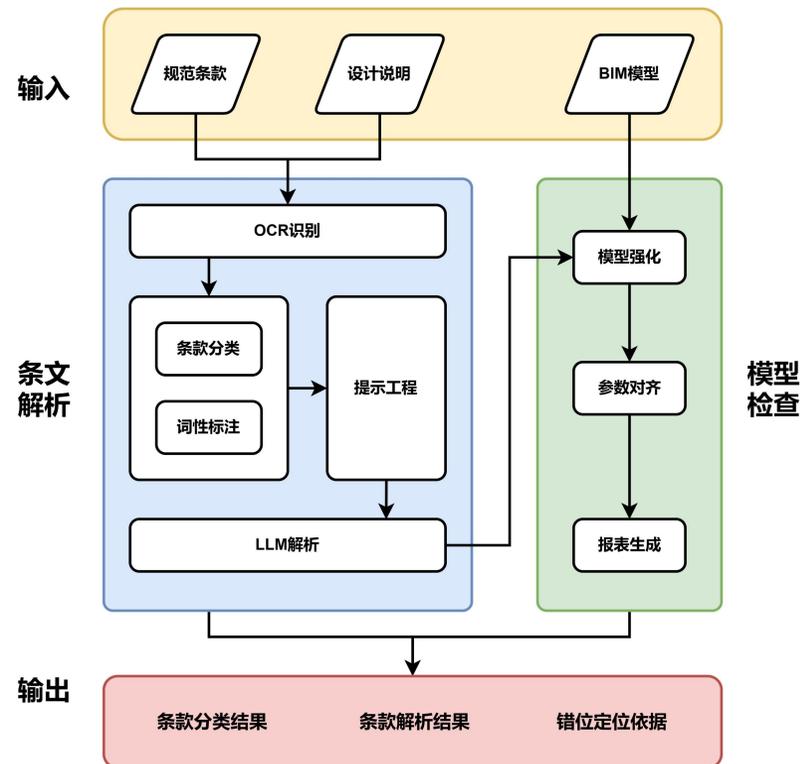
## 主要研究内容和贡献:

- (1) 基于此研究方向中词性标注数据集难以互通问题，提出了一种更为普适的**条款语义简化结构**：大语言模型以其强大的理解能力和推理能力在信息提取任务中表现出色。因此，对于大语言模型来说，并不需要为其提供多余的规则用于条款句子的解析，仅仅只需要简单的物件标签和子句关系标签帮助大语言模型规避专有名词和连接子句之间逻辑即可，而这几类标签在以往的研究者是普遍存在的，从而解决了各研究数据集难以互通的问题。
- (2) 将最先进的大语言模型应用于暖通领域的规范解析，建立并验证基于大语言模型的**建筑自动合规性检查流程框架**；
- (3) 针对本课题所涉及的中文建筑条款解析任务，进行了**大模型+策略组合实验**。得到表现优良的模型+提示策略组合。
- (4) 本研究通过Revit Model Checker插件开发了一套全自动**BIM模型检查工具**，并选取了三栋体量各异的BIM建筑模型对该系统进行了测试，验证了该方法的可行性。



# 总结/展望

标签	说明
obj	条款中出现的建筑元素，通常是一个建筑构件，元素之间的相互连接可能是并列也可能是从属游关系
cmp	比较谓词，表示比较的判断词，被比较的对象可以是数值也可以是obj
value	具体数值，表示某个构建属性的值
qtg	属于trigger的一种，是一种表示引用的触发标签
ctg	属于trigger的一种，是一种表示条件的触发标签



# 总结

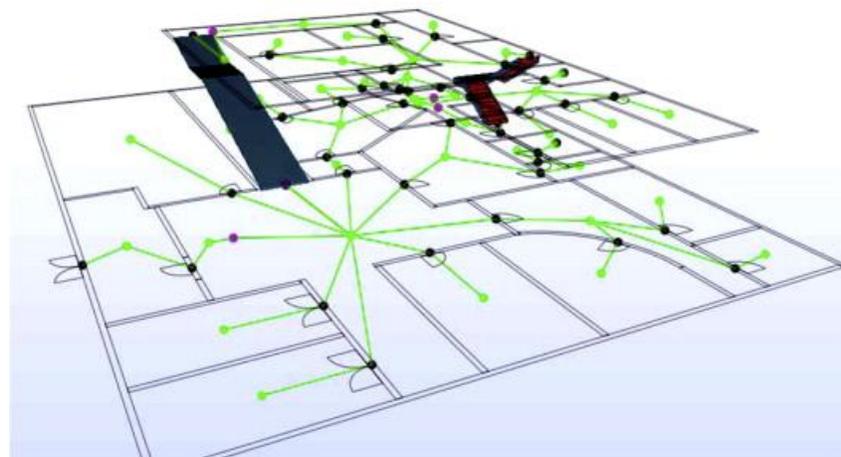
## 本方向需要被继续深挖的点:

- (1) 从条款解析到模型检查中间依然需要大量的**元素对齐**工作，并且因为部分条款本身表达并不包含“数值对比”，导致许多条款的语义逻辑在插件检查中无法对应到具体的运算逻辑中，这是在近几年此方向研究中最需要被解决的问题。**(最迫切解决的问题)**
- (2) **建筑自动合规性检查流程的全黑箱化**：本研究仅仅只是将黑盒模型的输出推进到结构化的语义表达。然而模型检查最终需要的是检查系统可运行的代码文件（在本研究中就是Revit Model Checker插件的XML输入文件）。随着大语言模型逻辑推理能力不断增强，这个流程将会被很快黑箱化。
- (3) **建筑模型数据增强方法**：本研究使用了设计说明作为外挂的项目设计参数来源。即使如此，当下的模型质量依然不支持全面的自动合规性检查（数据缺失建筑规范条款中所涉及的**许多参数往往需要进行额外的计算**（例如：最短逃生距离、窗墙比、体型系数等）。
- (4) **基于开源协议和平台的模型审查工具**：本研究所使用的Revit Model Checker是Autodesk公司所推出的商业软件。该软件首先在建筑构件的定义结构上极为复杂，因此在条款元素映射的过程中经常出现无法和某类构件代码进行对应的情况，从而降低了系统通用性；其次，该软件运算逻辑较为简单，无法对间接型引用条款进行处理。
- (5) **更加多样的检查交互方式**：**除了用自然语言文本以及模型数据作为此类任务的输入**，设计师需要更加直观的实时交互方式对模型进行审查，从而提高模型审查的可解释性以及准确性。

# 总结/展望

类别/参数名	正则表达式	说明
OST_Doors	门\$	门
OST_Stairs	楼梯\$	楼梯
OST_Windows	窗\$ ^窗 透光材料	窗、幕墙
OST_Walls	墙\$ ^墙	墙
OST_Floors	楼板\$	楼板
OST_MassShade	遮阳\$	遮阳
OST_Sheets	(甲 乙)?类 严寒 夏热冬(冷 暖)? 温和	气候区
OST_Rooms	(房 空)?间	房间
OST_DuctTerminal	(送 回)?风	风管、风口
OST_MechanicalEquipment	散热器 锅炉 供暖系统	设备、系统

不方便!





同濟大學  
TONGJI UNIVERSITY

College of Mechanical Engineering

谢谢评委老师的时间  
请老师们指正

Email: [2130276@tongji.edu.cn](mailto:2130276@tongji.edu.cn)