

基于 LLM 智能体的智能家居场景规则自动生成*

刘力玮^{1,2}, 陈伟^{1,2,3}, 魏峻^{1,2}, 朱家鑫^{1,2,3,4}, 王伟^{1,2,3}, 吴国全^{1,2,3}, 汪涛^{1,2}



¹(中国科学院 软件研究所, 北京 100190)

²(中国科学院大学, 北京 100049)

³(中国科学院大学南京学院, 江苏 南京 211135)

⁴(中科南京软件技术研究院, 江苏 南京 211135)

通信作者: 陈伟, E-mail: wchen@otcaix.iscas.ac.cn

摘要: 在智能家居系统中, 场景自动化规则可根据用户设定的触发条件, 自动执行一组设备操作, 实现如“回家模式”“睡眠模式”等复杂的设备联动行为. 然而, 终端用户在创建定制化规则时常面临挑战, 特别是在实现多设备协同和满足复杂用户意图时表现尤为明显. 一方面, 用户普遍缺乏专业知识, 难以清晰表达需求或将其转化为形式化规则; 另一方面, 现有自动化规则生成方法普遍存在两方面局限: 其一, 缺乏对复杂场景的建模能力, 难以支持涉及多设备协同的高级自动化需求; 其二, 过于依赖用户提供明确的、结构化的指令, 忽略了用户意图的隐性表达. 为解决上述问题, 提出 HomeMind——以大语言模型 (large language model, LLM) 为核心, 融合模式挖掘与语义理解能力的智能体系统. HomeMind 的工作流程包括 3 个核心步骤: 首先, 采用无监督的频繁事件模式挖掘方法, 自适应处理不同长度的事件序列, 以更精准地识别事件模式; 其次, 为每个模式提取周与日两个尺度的时间特征, 显著增强其时间上下文信息; 最后, 通过上下文增强的提示词机制, 引导 LLM 基于思维链推断用户意图, 生成语义明确的场景规则. 最终, HomeMind 将生成的候选规则呈现给用户, 辅助其理解并筛选最契合生活习惯的自动化方案, 并支持与智能家居平台接口集成, 实现规则一键部署. 在真实世界数据集和合成数据集上对 HomeMind 进行全面评估, 实验结果表明其在意图推断和规则生成准确性方面均优于现有基线方法.

关键词: 触发-动作编程; LLM 智能体; 频繁模式挖掘; 物联网; 智能家居

中图法分类号: TP311

中文引用格式: 刘力玮, 陈伟, 魏峻, 朱家鑫, 王伟, 吴国全, 汪涛. 基于 LLM 智能体的智能家居场景规则自动生成. 软件学报. <http://www.jos.org.cn/1000-9825/7663.htm>

英文引用格式: Liu LW, Chen W, Wei J, Zhu JX, Wang W, Wu GQ, Wang T. Automatic Generation of Smart Home Scenario Rules Based on LLM Agents. Ruan Jian Xue Bao/Journal of Software (in Chinese). <http://www.jos.org.cn/1000-9825/7663.htm>

Automatic Generation of Smart Home Scenario Rules Based on LLM Agents

LIU Li-Wei^{1,2}, CHEN Wei^{1,2,3}, WEI Jun^{1,2}, ZHU Jia-Xin^{1,2,3,4}, WANG Wei^{1,2,3}, WU Guo-Quan^{1,2,3}, WANG Tao^{1,2}

¹(Institute of Software, Chinese Academy of Sciences, Beijing 100190, China)

²(University of Chinese Academy of Sciences, Beijing 100049, China)

³(University of Chinese Academy of Sciences, Nanjing, Nanjing 211135, China)

⁴(Nanjing Institute of Software Technology, Nanjing 211135, China)

Abstract: In smart home systems, scenario automation rules can automatically execute a set of device operations based on user-defined trigger conditions, enabling complex coordinated behaviors such as “home mode” or “sleep mode”. However, end users often face challenges when creating customized rules, particularly when coordinating multiple devices and addressing complex intentions. On the one

* 基金项目: 国家自然科学基金 (62472412); 中国科学院软件研究所重大项目 (ISCAS-ZD-202302)

收稿时间: 2025-08-20; 修改时间: 2025-12-05; 采用时间: 2026-03-06; jos 在线出版时间: 2026-06-03

hand, users typically lack the technical expertise to clearly articulate their needs or translate them into formal rules. On the other hand, existing methods for generating automation rules have two primary limitations. First, they lack the capability to model complex scenarios, making it difficult to support advanced automation needs involving multi-device coordination. Second, they overly rely on users providing explicit, structured instructions, overlooking the implicit expression of user intentions. To address these issues, this study proposes HomeMind, an intelligent agent system powered by large language models (LLMs) that integrates pattern mining and semantic understanding capabilities. The workflow of HomeMind comprises three core steps. First, an unsupervised method for mining frequent event patterns is used to adaptively process event sequences of varying lengths, enabling more accurate identification of event patterns. Second, temporal features at both weekly and daily scales are extracted for each pattern, significantly enhancing its temporal context. Finally, a context-augmented prompt mechanism is leveraged to guide the LLM in inferring user intentions through chain-of-thought reasoning and generating semantically clear scenario automation rules. Ultimately, HomeMind presents candidate rules to users, facilitating the understanding and selection of automation solutions best suited to their lifestyles, and supports integration with smart home platforms for one-click rule deployment. Comprehensive evaluations of HomeMind are conducted on both real-world and synthetic datasets. Experimental results demonstrate that HomeMind outperforms existing baseline methods in both intention inference and rule generation accuracy.

Key words: trigger-action programming (TAP); LLM agent; frequent pattern mining; Internet of Things (IoT); smart home

智能家居 (smart home) 作为物联网 (Internet of Things, IoT) 的典型应用场景, 通过互联设备的协同工作实现日常任务的自动化, 从而提升用户的生活便利性和舒适度. 当前, 智能家居自动化的主流实现技术基于触发-动作编程 (trigger-action programming, TAP) 范式^[1], 其核心逻辑采用“IF [事件] THEN [动作]”的规则形式定义设备行为关联. 这类自动化规则 (即 TAP 规则) 通过事件驱动机制自动执行预设动作, 有效简化了智能家居系统的配置复杂度, 已被 IFTTT^[2]、SmartThings^[3]、米家^[4]等商业平台以及 Home Assistant^[5]、OpenHAB^[6]等开源系统广泛采用.

然而, 终端用户在手动创建定制化的场景规则时仍面临挑战. 其核心难点在于, 用户需同时准确理解自身需求与设备底层逻辑 (包括状态监测、触发器设置及动作执行^[1]), 并建立需求与设备行为之间的映射关系. 以“离家安全”场景的自动化规则为例, 用户需要协调门锁、灯光、摄像头和运动传感器等多设备的联动行为. 此外, 用户需求通常以模糊的自然语言表达 (如“离家时确保房屋安全”), 其语义较为抽象, 难以直接转化为具体的触发条件和执行动作, 导致用户意图与机器可执行逻辑之间存在语义鸿沟. 人工创建的场景规则常偏离实际需求, 因而多数用户倾向于依赖平台提供的预设模板. 然而, 这些模板普遍存在个性化不足和扩展性受限的问题.

现有研究主要从两个方向探索 TAP 规则的自动生成: 一是基于日志的隐式模式挖掘^[7-9], 二是自然语言到结构化规则的转换^[10,11]. 前者通过分析历史设备操作序列推断用户意图, 但由于缺乏语义理解能力, 难以将序列模式转化为符合 TAP 规则语法的触发-动作结构, 因而通常局限于“单触发-单动作”的简单规则建模, 对长期周期性规律 (例如以周为时间周期的模式) 识别能力不足, 且易受随机操作噪声的干扰. 后者利用 LLM 的语义理解和推理能力, 将自然语言指令转换为 TAP 规则, 部分研究进一步构建基于 LLM 的交互式智能体以辅助规则生成^[10,11]. 然而, 此类方法高度依赖用户提供明确的指令, 当用户意图模糊或表达不完整时, 生成规则的质量会显著降低.

针对上述局限性, 本文提出一种融合事件模式挖掘与 LLM 语义理解的解决方案: 一方面, 通过从历史日志中自动挖掘高价值事件模式, 提炼能够反映用户行为偏好的规律, 从而提取隐式意图; 另一方面, 借助 LLM 的语义理解能力, 在无需用户提供精确指令的情况下, 将事件模式与相关上下文注入 LLM 的推理链, 以降低幻觉发生率并提升规则生成准确性. 为此, 本文提出 HomeMind, 一个用于自动生成智能家居场景规则的 LLM 智能体系统.

HomeMind 具备多工具调用能力, 通过顺序决策协调数据预处理、频繁模式挖掘、时间特征提取、用户意图推断和 TAP 规则生成等任务. 其场景规则生成流程包括 3 个阶段: 首先, 基于无监督频繁模式挖掘, 从历史事件日志中提取高频事件模式, 捕捉用户的行为习惯与隐性需求; 其次, 为每个模式构建多粒度时空特征, 包括日/周时间维度与空间维度, 以增强上下文表征能力; 最后, 通过基于思维链 (chain-of-thought, CoT) 的上下文增强提示策略, 引导 LLM 推断用户意图, 生成符合用户需求的场景自动化规则.

我们基于 GPT-4o 构建了 HomeMind 原型系统, 并在真实与合成智能家居数据集上进行了全面的实验评估. 结果表明, HomeMind 生成的 TAP 规则兼具可解释性与实用性. 量化评估显示, HomeMind 在意图推断任务中取得

了 88.75% 的 Accuracy@20 指标, 在 TAP 规则生成任务中达到 86.25% 的 Precision@20 和 54.29% 的 Recall@20. 相较于 Trace2TAP^[8]和 SGen^[9]等基准方法, HomeMind 在规则覆盖率与场景适应性指标上均展现出显著优势.

本文的核心贡献包括以下 3 个方面.

(1) 提出一种自适应划分智能家居事件日志并提取用户行为模式的挖掘方法, 结合多粒度时空特征增强模式语义. 通过嵌入智能体流程的模式挖掘工具, 提供结构化、准确的时空上下文, 为后续意图推断与规则生成奠定基础.

(2) 设计并构建 HomeMind, 一个具备上下文增强和思维链推理能力的 LLM 智能体系统. 该系统能够理解用户行为模式, 准确推理其意图, 并生成结构清晰、语义明确的 TAP 规则.

(3) 在真实和合成智能家居数据集上对 HomeMind 进行系统评估, 实验结果表明其在意图推断和规则生成方面的性能优于现有基线方法.

本文第 1 节介绍本研究所依赖的基础知识, 阐述 TAP 规则生成的研究现状及挑战. 第 2–7 节详细描述 HomeMind 的总体方法和技术实现. 第 8 节通过实验验证 HomeMind 在意图推断与规则生成方面的有效性与优势. 第 9 节分析 HomeMind 的局限性及其潜在的有效性威胁. 第 10 节综述与 TAP 规则生成相关的研究工作. 第 11 节对全文进行总结.

1 问题分析

1.1 基础知识

智能家居系统基于事件驱动架构, 通过持续感知环境变化并触发响应实现自动化运行. 系统将自身及环境的动态变化记录为时间戳事件序列 (即事件轨迹), 其核心构成要素可分为: (1) 操作事件: 用户主动触发的设备控制行为 (如开启灯光、调节空调温度); (2) 状态变更事件: 设备感知的环境反馈状态变化 (如检测到人员移动、门窗开启). 以下为事件的形式化定义.

定义 1. 事件 (event). 事件是智能家居环境中表征操作行为或状态变化的原子单元, 定义为五元组 $\text{event}=(\text{object}, \text{service}, \text{parameter}, \text{time}, \text{location})$:

- 对象 (object): 事件源实体, 通常是一个智能家居设备, 通过设备类型、唯一 ID 等元数据标识.
- 服务 (service): 设备提供的功能和服务接口 (如 switch、setTemperature、detectMotion).
- 参数 (parameter): 服务调用的具体参数 (如 on、25°C、detected).
- 时间 (time): 事件发生的时间戳.
- 位置 (location): 事件发生的空间位置 (例如, 客厅 (living room)、卧室 (bedroom)).

抽象事件 (abstract event) 简化为三元组 (object, service, parameter), 保留语义核心, 剔除具体事件的时空上下文. 为简化表述, 后文在不引起歧义的情况下均采用此形式. 例如: (light, switch, on) 表示开灯的操作事件. (motion-Sensor, detectMotion, detected) 表示运动检测传感器触发的状态变更事件.

定义 2. 场景 (Scenario). 在智能家居系统中, 场景指一组具有时空关联性的事件序列, 共同表征特定的用户意图或行为模式. 实现场景自动化是智能家居系统的核心目标, 依赖于将事件序列转化为符合语义和触发逻辑的 TAP 规则, 其形式化表示为:

$$\text{IF triggers occur, THEN perform actions} \quad (1)$$

其中, 触发器 (triggers) 是指启动场景的初始事件集合, 而动作 (actions) 则是为响应触发器而执行的满足用户意图的设备操作序列.

定义 3. 事件模式 (Event Pattern). 事件模式由一组语义关联的抽象事件集合 $\{e_1, e_2, \dots, e_k\}$ 构成, 表征潜在场景的共性特征. 例如, 事件模式 $\{(\text{curtain_motor}, \text{control}, \text{close}), (\text{fan}, \text{set_mode}, \text{sleep}), (\text{bed}, \text{pressure_sensor}, \text{detected})\}$ 可表征一个“午睡”场景. 与之对应, 事件模式实例表示该模式在某一时刻的具体发生, 包含具体的时空上下文信息.

1.2 问题分析

近年来, LLM 的发展推动了智能家居自动化技术的进步. 凭借其强大的自然语言处理能力, LLM 能够解析用

户以自然语言表达的自动化需求,并将其转化为可执行的 TAP 场景规则.此外,LLM 在数据处理方面的优势为从历史设备日志中挖掘潜在事件模式、分析其语义关联并生成自动化规则提供了新的可能.在此背景下,基于 LLM 的智能家居自动化研究主要沿两条技术路径发展:(1) 用户自然语言需求到 TAP 规则的端到端生成;(2) 基于历史设备日志的用户意图推断与规则合成.然而,这两类方法在处理模糊输入和非结构化行为数据时均存在显著局限性.为系统评估这些局限性,本文基于 GPT-4o 开展了以下两项实证研究:(1) 分析提示词对 LLM 生成场景规则可靠性的影响;(2) 验证在缺乏事件模式支持的情况下,LLM 直接基于原始事件日志进行意图推断的有效性.

1.2.1 实证研究 1: 提示词对场景规则自动化生成的影响

现有系统(如 Sasha^[10]、ChatIoT^[11])依赖 LLM 的语义理解能力实现规则生成,但其性能受提示词具体程度的显著影响.例如,“当我离开家时,锁门并关灯”这类明确提示能生成准确的规则,而模糊表达如“在我离开时保证家庭安全”可能导致规则不完整或不准确.

为评估 LLM 在不同具体程度提示词下的表现,我们重点考察其在意图推断和 TAP 规则生成方面的能力.具体而言,本文分析了模型在从明确指定到高度抽象的提示范围内的响应效果.本研究基于一个包含 32 种代表性用户意图的合成数据集(详见第 8.1 节),借鉴 Sasha 方法^[10],将提示词按具体程度划分为 3 个层级:(1) 明确指定:提示明确包含触发条件、目标设备及期望操作;(2) 部分指定:提示提供部分上下文信息(如时间和位置)和一般性意图,但缺少关键操作细节;(3) 完全不指定:提示仅包含抽象意图(如“离家安全”),未提供任何可执行要素.

我们采用 ChatIoT^[11]的提示词结构,配置了具备预定义系统角色与推理流程的 LLM 智能体.针对 32 种用户意图中的每一种,我们在每个具体程度层级下分别构造提示词,并提供可用设备及其功能的上下文信息.图 1 展示了“离家安全”意图在 3 种提示词类型下的示例.为消除随机性影响,每个提示词执行 4 次生成过程,生成结果与人工标注的规则基准进行对比.实验结果表明,规则生成的准确率随提示词具体程度的下降而急剧降低,明确提示的规则生成准确率为 96.88%,部分指定降至 53.13%,完全不指定进一步降至 33.59%.

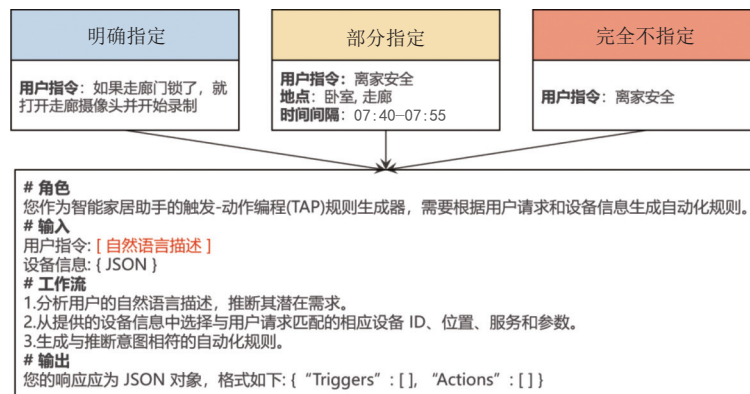


图 1 意图“离家安全”在不同具体程度层级下的提示词

通过对错误案例的深入分析,我们发现准确率下降的主要原因包括以下两点.首先,现有基于 LLM 的端到端生成方法,本质上仍属于一种高阶的“精准提示工程”,其性能高度依赖于输入提示与模型预设任务模式之间的匹配程度.一旦用户意图表述模糊或不完整,提示信息与模型内隐的“规则生成逻辑”之间便出现语义断层,导致系统性能急剧下滑.其次,面对模糊指令时,LLM 往往表现出保守的生成倾向,即优先输出最通用、最安全的规则组合,这种倾向导致模型频繁遗漏符合场景特定需求的关键动作.例如,在“离家安全”场景中,基准规则通常包含“锁门、关灯、开启摄像头”等多个维度操作,而 LLM 在模糊提示下往往仅生成“锁门”这一最显性的动作,忽略了环境控制与安防监测等重要组成部分,致使生成的规则功能残缺.

综上所述,本实证研究表明,完全依赖用户自然语言输入并与 LLM 直接映射的现有方法,存在对意图表述精确性的强依赖性.由于 LLM 自身的内在机制限制及其在模糊语境下的保守生成策略,该方法难以持续输出完整、

准确且契合场景细微需求的 TAP 规则, 这构成了该技术路径目前的一个瓶颈。

1.2.2 实证研究 2: 基于原始事件日志的 LLM 意图推断性能评估

为评估模式挖掘在意图推断及自动化生成中的作用, 我们在 ARAS^[12]和 Kasteren^[13]两个真实智能家居数据集上开展了案例研究。我们从每个数据集中随机抽取了 100 个标注有特定活动 (如上床睡觉、离开家) 的时间区间, 针对每个区间, 将对应的传感器状态变化序列转化为结构化事件描述, 作为 LLM 的输入提示词, 提示词的具体结构如图 2 所示。值得注意的是, LLM 在推断过程中未提供任何挖掘得到的事件模式或上下文抽象信息, 仅依赖原始事件序列进行意图推理。

```
# 任务信息
您是一个智能家居行为分析专家, 给定一段智能家居用户的设备操作日志片段, 以及一系列的用户意图,
您需要判断出哪个意图最能反映出该片段里用户的整体行为模式。

# 输入
设备操作日志片段: {2008-03-10 09:47:25 hallway toilet_door change_status close,
2008-03-10 09:47:28 hallway toilet_door change_status open, ...}
意图列表: {1.离家, 2.入厕, 3.洗澡, 4.睡觉, ...}

# 输出
请只输出最可能的单个活动名称, 您的输出结果应为: '总体意图: XXX.'
```

图 2 基于原始事件日志的 LLM 意图推断性能评估的提示词

实验结果显示, 在仅包含 7 种活动类型的 Kasteren 数据集上, LLM 的意图推断准确率为 62%, 表现一般。在更复杂、包含 26 种活动类型的 ARAS 数据集上, 准确率进一步下降至 28%。进一步对失败案例进行分析后, 我们发现准确率低下的原因主要包括两点。首先, 原始日志主要由低层次传感器触发构成, 其中大量动作与核心意图无关 (如用户走动导致多次运动传感器触发), LLM 很难从冗长且噪声密集的事件序列中自动筛选出具有判别力的关键事件, 导致推断结果不稳定。其次, 单次提示词仅包含一个局部序列, 模型无法自主获取周期性信息 (如“每个工作日 22:00 关闭客厅灯”), 而这些周期性规律恰恰是推断“睡觉”等习惯性意图的关键依据。

综上所述, 这一实证研究说明, 直接将原始的历史数据抛给 LLM, 并不能有效解决用户隐式意图的推断问题。相反, 它凸显了一个关键的前置需求: 必须在 LLM 推理之前, 对历史事件流进行提炼、压缩和结构化建模, 将低级别的传感器触发和设备操作转化为高级别的行为模式, 才能支撑可靠的意图理解与自动化规则生成。

1.3 研究动机

基于上述研究与分析, 我们总结出以下 4 个关键观察, 这些观察直接启发了 HomeMind 的设计。

观察 1: 用户意图隐式编码于设备事件序列中。用户通常周期性地执行一系列设备操作, 反映其日常习惯与潜在意图。例如, 频繁出现的事件序列 {(curtain_motor, control, close), (fan, set_mode, sleep), (bed, pressure_sensor, detected)} 可能指向“午睡”或“夜间睡眠”场景。事件序列的频繁共现特性为从日志数据中推理用户意图提供了统计基础。

观察 2: 不同事件序列可能映射到相同意图。由于用户行为的随机性和环境上下文的变化, 不同的事件序列可能体现相同的意图。图 3 以“午睡”场景为例, 展示了这些变化如何导致对应同一意图的异构事件序列。例如, 前两个序列的事件顺序差异可能源于用户与风扇和窗帘的距离不同; 而最后一个序列中缺少风扇事件, 则可能由于天气凉爽, 无需开启风扇。

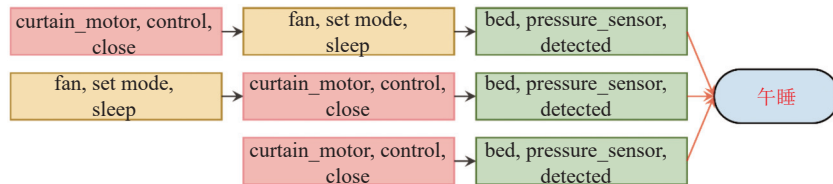


图 3 行为随机性和上下文变化如何影响相同意图下事件顺序的示意图

观察 3: 日志中的事件顺序不等同于 TAP 规则的逻辑结构. 日志记录反映用户的实际操作顺序, 与 TAP 规则所需的逻辑结构可能存在差异. 例如, 图 3 中的第 2 个序列记录用户依次设置风扇睡眠模式、关闭窗帘、躺在床上, 但实际的 TAP 规则应以床压力传感器的检测为触发器, 将风扇与窗帘操作作为动作. 这表明, 事件序列需经过语义转换才能映射为逻辑合理的 TAP 规则.

观察 4: 时空上下文有助于提升意图推断的准确性. 相同的事件序列在不同时间或空间下可能对应不同意图. 例如, 图 3 中的第 2 个序列在 12:30 可能表示“午睡”, 而在 23:00 更可能为“夜间睡眠”. 空间信息同样提供关键线索: 设备所在位置通常与意图类型密切相关, 如“夜间睡眠”通常仅涉及卧室内的设备事件.

综上所述, 我们提出 HomeMind, 一种结合频繁事件模式挖掘与上下文增强提示词的 LLM 智能体系统. 前者克服了 LLM 直接分析原始事件日志的局限性, 后者利用 LLM 的语义理解能力, 同时借助已识别的事件模式和时空上下文, 辅助 LLM 更准确地推断用户意图并生成语义合理的 TAP 规则.

2 HomeMind 总体架构

图 4 展示了 HomeMind 的整体架构, 涵盖 LLM、意图知识记忆模块、数据处理与规则生成工具链 (包括数据预处理、频繁模式挖掘、时间特征提取、用户意图推断与 TAP 规则生成) 以及用于规则展示与部署的外部接口模块. 基于此架构, 场景规则的生成流程主要包括 5 个阶段: (1) HomeMind 对原始事件日志进行预处理, 转化为结构化且语义更丰富的信息格式; (2) 从结构化日志中挖掘频繁事件模式, 识别可能反映用户意图的核心行为序列; (3) 针对挖掘的事件模式, 提取长期时间 (如周) 与短期时间 (如天) 尺度下的时间特征, 增强模式的时间上下文; (4) 采用上下文增强的提示词机制, 引导 LLM 结合事件模式和时间特征推断用户意图, 生成语义明确的 TAP 规则; (5) 最后, 通过可视化界面展示推荐的 TAP 规则, 包括其语义描述、设备操作和时间规律, 方便用户理解和选择, 选定的规则自动转换为目标智能家居平台兼容的可执行脚本并完成部署. 后文图 5 进一步展示了最核心的数据处理与规则生成工具链的详细执行流程, 标注了各子模块的输入输出、关键算法、工具调用顺序以及 LLM 在其中扮演的决策角色.

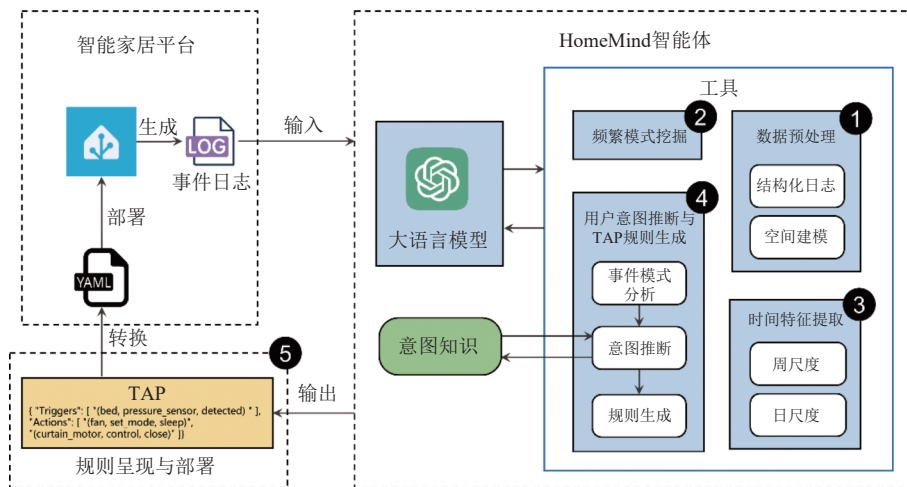


图 4 HomeMind 总体架构

3 数据预处理

(1) 结构化事件日志

智能家居事件日志格式多样, 导致直接使用原始日志训练 LLM 在语义理解和推理能力上受限^[12-14]. 为提升模型对事件语义的理解能力, 本文设计了数据预处理工具, 将原始日志转换为统一的结构化表示, 构建语义清晰、

上下文明确的输入格式, 以支持后续的意图推断与规则生成.

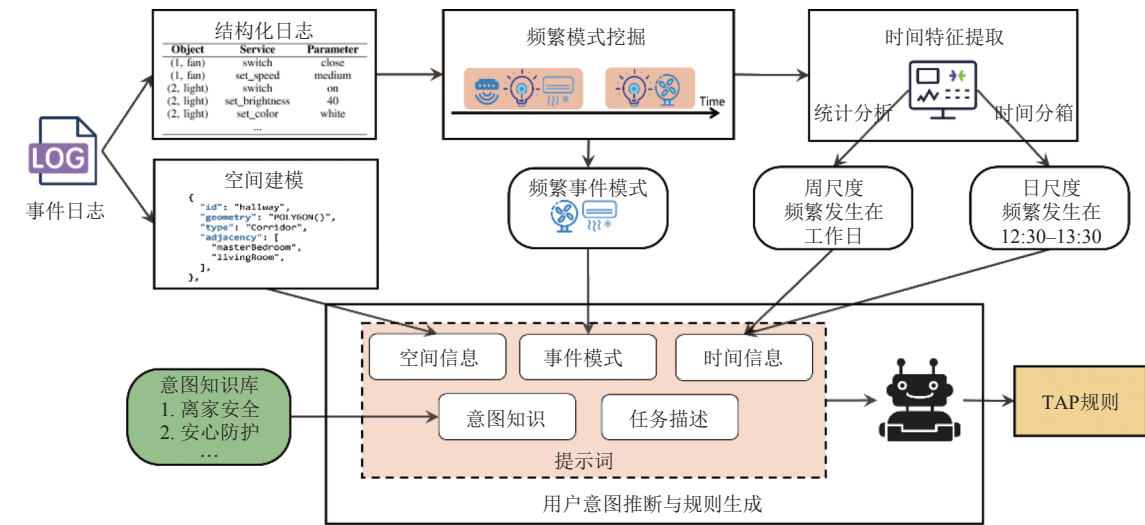


图 5 数据处理与规则生成工具链执行流程

结构化转换面临两方面挑战: 首先, 不同数据源的命名方式、字段格式和记录粒度不一致, 缺乏统一标准; 其次, 多数日志仅包含设备 ID 或传感器编号, 缺乏可供推理的语义信息。

为应对上述挑战, 本文基于第 1.1 节中的事件定义, 对日志进行标准化处理, 并引入语义增强机制。具体而言, 利用预构建的映射表对原始日志条目进行语义补全, 补齐每个事件的关键属性, 包括设备类型 (device)、所在位置 (location)、功能类别 (service) 以及具体参数 (parameter) 等。这些语义信息通常在原始日志中缺失或仅以编码形式存在。为此, 我们通过解析智能家居系统文档、设备使用手册以及平台配置文件, 构建了覆盖常见设备和操作的映射规则。例如, 在 Kasteren 数据集^[13]中, ID 为 17 的传感器可映射为语义明确的设备标签 (17, 冰箱), 从而显著提升事件表达的可读性和建模能力。表 1 展示了若干结构化日志转换示例。

表 1 结构化事件日志示例

Time	Location	Object	Service	Parameter
2024-04-01 20:16:43	secondBedroom	(1, door sensor)	open_close_detection	close
2024-04-01 20:35:00	bathroom	(2, light)	switch	on
2024-04-01 20:35:06	bathroom	(2, light)	set brightness	80
2024-04-01 20:35:15	bathroom	(3, washing machine)	switch	on

(2) 空间建模

为使 LLM 有效利用空间上下文信息, 本文设计了空间建模工具, 基于 IndoorGML 标准^[15]对室内空间布局进行结构化建模, 捕获空间的拓扑关系与语义属性。每个空间单元被建模为一个节点, 并附加以下属性:

- 空间 ID: 空间单元的唯一标识符。
- 几何形状: 以平面图左下角为原点的二维多边形, 定义空间物理边界。
- 类型: 空间功能标签 (如客厅、走廊等)。
- 邻接关系: 与当前空间拓扑相连的其他空间单元列表。
- 设备: 安装于当前空间内的智能设备列表。

如图 6 所示, 基于 IndoorGML 的空间建模增强了室内空间表示的结构化程度, 为 LLM 智能体的推理提供了关键的上下文支持。LLM 能够基于此来识别空间之间的邻接关系、推断用户移动轨迹, 并理解不同空间在家庭活动中的功能角色, 从而提升意图推断和规则生成的准确性。

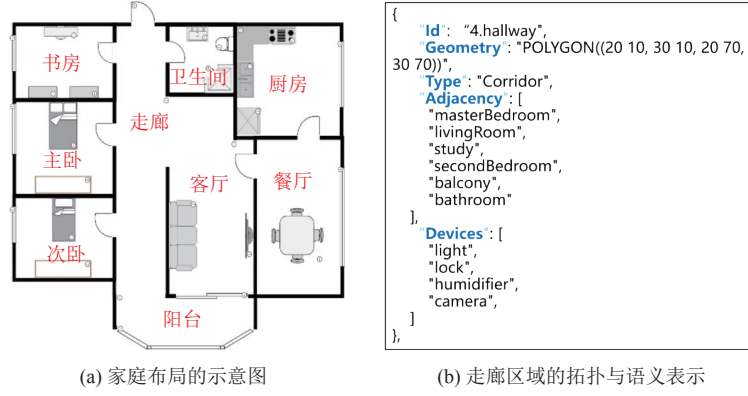


图6 基于 IndoorGML 的空间建模示意图

4 频繁模式挖掘

直接将结构化事件日志输入 LLM 进行分析面临诸多挑战: 其一, 推理过程计算开销高, 难以在大型数据集上高效扩展; 其二, 受限于模型的最大 token 长度, 日志需分割为较小片段, 而不当的切分可能破坏上下文连续性, 导致模式提取错误或不完整. 为此, 本文设计了频繁模式挖掘工具, 优先从结构化事件日志中提取高频且关联性强的频繁事件模式, 捕捉用户行为特征和潜在意图, 为规则生成提供支持.

工具采用无监督的频繁模式挖掘算法 FP-Growth^[16], 基于事件共现频率, 识别具有代表性的频繁事件模式. 为确保模式的语义相关性和出现频率, 我们设定了最小支持度阈值, 并根据日志覆盖的时间跨度动态调整. 例如, 在覆盖四周的日志中, 若目标是挖掘每周至少出现一次的模式, 则支持度阈值设为 4, 以反映周期性行为特征.

频繁模式挖掘的核心挑战在于合理划分事件序列, 以提取语义完整的日志片段. 传统方法常采用固定时间窗口 (如每 10 min)^[9,17] 或固定事件数量 (如每 10 个事件)^[18] 进行分割, 但这类静态策略难以适应智能家居场景下行为时长的动态变化, 易导致语义断裂. 如图 7 所示, 固定窗口和固定长度划分可能将语义关联的事件错误拆分至不同片段, 丢失场景的整体语义.

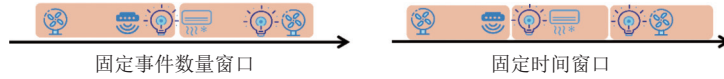


图7 固定窗口分割方法导致的语义断裂

为应对这一问题, 我们提出了一种基于相邻事件距离的自适应分割策略. 该策略以相邻事件的时间间隔 (事件距离) 为依据, 动态调整分割阈值 $dist$. 当两事件的距离超过该阈值时, 判断场景或用户意图可能发生切换, 启动新的日志片段. 该阈值根据近期历史事件之间的间隔更新, 其计算公式如下:

$$dist = \alpha \times mean + \beta \times std \quad (2)$$

其中, $mean$ 与 std 分别为过去 N 个相邻事件间隔的平均值与标准差, α 与 β 为可调超参数. 日志切分数量直接影响事件相关性分析和频繁模式挖掘效果, 且不同数据集的事件密度和分布特征上差异显著, 因此需根据数据特性调优参数. 合成数据集上的敏感性分析表明, 参数 N 对切分数量影响较小, 而 α 和 β 的增大显著减少切分数量. 例如, 当 β 从 0.2 增至 0.5 时, 日志片段数从 1530 减少至 1205. 实践中, 我们将 N 、 α 和 β 分别设置为 100、0.9 和 0.2, 该策略有效保持事件的时序连贯性, 适应用户行为的动态变化.

在完成片段分割后, 我们应用 FP-Growth 算法提取频繁事件模式, 不同于传统方法主要依据支持度进行模式筛选, 本文关注频繁模式在意图推断与规则生成场景中的表达价值. 对于后续推理而言, 尽管某些模式出现频率较高, 但若其在事件构成上高度相似或冗余 (例如仅设备组合略有差异的同类行为), 并不能有效拓展对用户多样化意图的理解. 为此, 本文进一步引入以下两步精炼过程, 以提升模式的质量与多样性.

- 冗余过滤: 移除高度冗余模式, 避免重复捕获相同用户意图. 具体而言, 若模式 a 中超过 70% 的事件被模式 b 所覆盖, 则视为冗余, 优先保留包含更多不同事件的模式, 以增强行为区分性.

- Top-K 模式选择: 根据模式复杂度 (事件种类数量) 和支持度 (出现频率) 联合排序, 优先选择复杂度更高的模式, 若复杂度相同, 则选择支持度更高的模式. 最终选出前 K 个兼备多样性与代表性的模式 (本文设置 $K=20$), 作为意图推断与规则生成的输入.

综上所述, 本研究中的频繁模式挖掘部分在目标与角色上与传统应用存在本质不同: 在分割策略上, 从固定机械分割转向自适应分割, 以保障输入片段的语义完整性; 在模式筛选上, 从频率驱动转向频率与语义多样性并重, 以提供更具有多样性的行为模式; 在设计目标上, 其输出的事件模式并非直接作为最终规则, 而是作为经过提炼的结构化中间语义表示, 这些模式与后续提取的时空特征一同构成增强的上下文, 旨在弥合低层事件流与高层用户意图之间的语义鸿沟, 从而更有效地支撑 LLM 进行深度推理与精准的规则生成.

5 时间特征提取

用户习惯和情境上下文共同影响事件模式, 使其可能在一周或一天的不同时间段多次出现. 每一次出现称为事件模式实例, 其特征包括开始时间、结束时间及所在时间区间. 通过统计这些实例在各时间段的分布, 可揭示事件模式的时间特性.

理解事件模式的时间特征对推断用户意图与生成场景规则至关重要. 不同事件模式在时间周期和尺度上呈现多样化的分布规律, 为此, 本文设计了时间特征提取工具, 为目标频繁事件模式提取时间特征. 考虑到智能家居规则通常在周和日周期内表现出显著规律性, 本研究聚焦于周和日两种时间尺度的特征提取, 同时支持根据需求扩展至月或年等更长周期.

5.1 周尺度的时间特征分析

为了提取频繁事件模式的周尺度时间特征, 从以下两个维度进行分析: (1) 模式在工作日与周末的出现偏好; (2) 模式是否集中出现在一周中的特定日期. 具体方法是, 统计事件模式实例在一周 7 天内的分布频次, 计算其在每个日期 i 上的出现率 r_{occ} 以量化上述特征. 出现率定义如下:

$$r_{\text{occ}}(i) = \frac{n_{\text{occ}}(i)}{n_{\text{total}}}, i \in \{\text{Mon.}, \text{Tue.}, \dots, \text{Sun.}\} \quad (3)$$

$$r_{\text{occ}}(\text{weekend}) = r_{\text{occ}}(\text{Sat.}) + r_{\text{occ}}(\text{Sun.}) \quad (4)$$

当事件模式在工作日或周末的出现比例超过预设阈值 (如 60%) 时, 分别标记为“工作日频繁”或“周末频繁”. 若某日期的出现率超过阈值 (如在周六的出现率超过 60%), 则将其标记为“该日期频繁”.

5.2 日尺度的时间特征分析

本节旨在刻画事件模式在日尺度时间周期上的分布特征, 重点识别其高频活跃的时间区间. 为此, 我们将一天划分为多个时间区段 (称为时间箱, time bins), 通过统计分析确定包含最多模式实例的时间箱, 以其起止时间描述模式的日尺度时间特性.

5.2.1 基于事件密度的时间分箱策略

时间分箱的核心挑战在于合理确定时间区段的长度. 已有研究^[19]多采用固定长度的时间区段 (如每小时), 但这种静态划分难以适应不同模式在持续时间与活跃频率上的差异. 例如, 平均持续 4 min 的模式适合 5 min 的时间箱, 而持续 10 min 的模式可能无法被完整覆盖. 此外, 即使模式具有周期性, 其每日发生时间常存在波动, 固定长度的时间箱可能因此遗漏部分实例. 以“离家安全”模式为例, 其实例多集中在工作日 7:50–8:10 开始、8:20–8:40 结束, 若使用 30 min 的固定时间箱 (如 8:00–8:30), 可能无法涵盖所有实例.

在第 4 节频繁模式挖掘阶段, 我们采用了基于事件距离阈值的启发式分割策略, 以适应长周期、稀疏型的智能家居日志数据. 这种粗粒度策略在降低计算开销的同时, 能够较好地保留事件的整体结构. 然而, 模式挖掘得到的高频行为模式数量有限, 其日尺度分布特征通常较为集中且边界明确, 因此需要更精细的分割策略.

为此, 本文提出一种基于事件密度的时间分箱策略, 通过分析模式实例中所有事件的密度分布, 自动确定各时间区段的边界. 具体而言, 我们采用核密度估计 (kernel density estimation, KDE)^[20], 将事件密度建模为一条平滑曲线, 从而突出事件在时间轴上的集中区域. 图 8 展示了该策略的整体流程, 主要包括以下 3 个步骤.

- 事件投影: 将所有模式实例的事件按天组织, 并将其时间戳 (以分钟为单位) 映射至统一日时间轴 (范围 0–1439, 对应 00:00–23:59). 任意时刻 $hh:mm:ss$ 被转换为分钟值 $hh \times 60 + mm$, 作为 KDE 的输入数据点.
- 密度分布建模: 应用高斯核密度估计 (Gaussian KDE), 生成平滑的事件密度曲线, 突出事件在各时段的集中程度.
- 时间分箱: 在 KDE 曲线中识别局部极小值作为潜在分割点, 为避免过度切分, 过滤掉曲线中幅度变化较小的次要波动, 仅保留与邻近峰值相比下降显著的极小值作为区段边界.

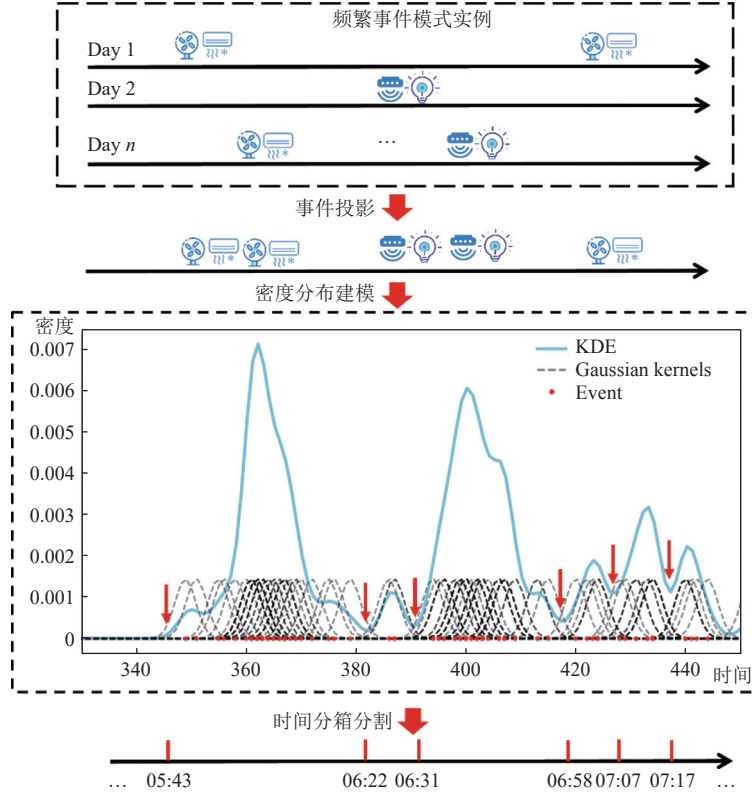


图 8 基于事件密度的时间分箱策略

如图 8 所示, 5:30–7:30 时间段内 KDE 曲线呈现多个峰谷, 反映用户活动在此期间的转换规律. 该分箱策略确保时间区段合理划分, 准确捕捉模式实例的时序特性.

5.2.2 时间特征向量构建

第 2 个挑战在于如何表征事件模式实例的时间特征. 为支持后续基于时间相似性的聚类分析, 我们为每个模式实例构建时间特征向量. 时间特征向量的维度等于前述时间分箱策略 (见第 5.2.1 节) 确定的时间箱数量, 每一维表示实例在对应时间箱内的持续时间比例. 形式化定义如下: 设共有 n 个时间箱, 其边界为 $\{b_0, b_1, \dots, b_{n-1}\}$, 第 i 个时间箱为区间 $[b_i, b_{i+1})$ ($0 \leq i \leq n-1$). 对于一个起止时间为 t_{start} 和 t_{end} 的模式实例, 其向量每个维度的值计算如下:

$$w_i = \begin{cases} \frac{\min(t_{\text{end}}, b_{i+1}) - \max(t_{\text{start}}, b_i)}{t_{\text{end}} - t_{\text{start}}}, & \text{if } \min(t_{\text{end}}, b_{i+1}) > \max(t_{\text{start}}, b_i) \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

其中, 每个 w_i 是第 i 维的特征值, 表示模式实例在第 i 个时间箱的持续时间比例. 最终, 向量 $w = [w_0, w_1, \dots, w_{n-1}]$ 即为该模式实例的时间特征向量. 如图 9 所示, 对于一个包含 3 个实例的事件模式, 计算得到的时间特征向量组成一个五维特征矩阵. 以持续时间为 07:05–07:15 的模式实例为例, 其各个维度的特征值分别为: $\frac{0}{07:15-07:05}, \frac{0}{07:15-07:05}, \frac{0}{07:15-07:05}, \frac{0.2}{07:15-07:05}, \frac{0.8}{07:15-07:05}$, 因此其对应的时间特征向量为 $[0, 0, 0, 0.2, 0.8]$.

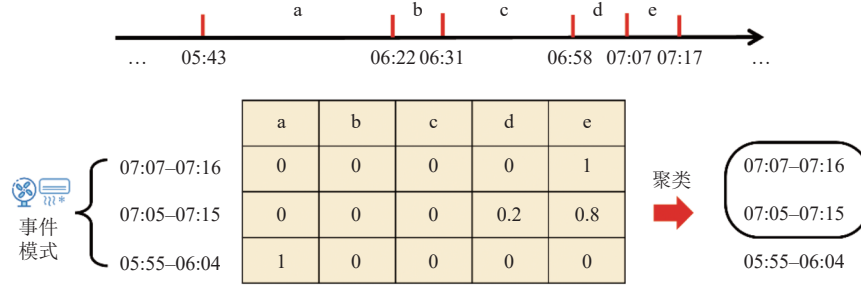


图 9 时间特征向量构建与聚类

5.2.3 时间特征提取

为刻画事件模式的日尺度时间特性, 我们识别与每个模式相关的显著时间箱, 即聚集显著比例事件模式实例的时间窗口, 代表模式频繁发生的时间区间.

首先, 我们对每个事件模式所有实例的时间特征向量应用 K-means 聚类进行分组. 为确定最优簇数量 K , 我们采用自适应策略, 通过计算轮廓系数在多个候选 K 值中选取最优值. 如图 9 所示, 前两个向量在特征空间中的距离较近, 因此被归入同一簇.

然后, 我们计算簇内所有向量在各时间箱的平均权重, 来衡量各个时间箱在该簇内的重要性. 例如, 图 9 中的簇包含两个实例, 其时间特征向量分别为 $[0, 0, 0, 0, 1]$ 和 $[0, 0, 0, 0.2, 0.8]$. 因此, 每个时间箱的平均权重计算为: $\frac{0+0}{2}, \frac{0+0}{2}, \frac{0+0}{2}, \frac{0+0.2}{2}, \frac{1+0.8}{2}$, 即 $[0, 0, 0, 0.1, 0.9]$.

最后, 识别那些平均权重高于阈值的时间箱作为显著时间箱, 该阈值定义为簇内所有时间箱平均权重的均值与标准差之和. 例如, 在图 9 的示例中, 均值为 $\frac{0+0+0+0.1+0.9}{5} = 0.2$, 标准差约为 0.35, 因此阈值为 $0.2+0.35=0.55$. 因此, 仅权重为 0.9 的时间箱 (07:07–07:17) 被判定为显著时间箱. 通过汇总所有显著时间箱, 便可得到该事件模式的日尺度时间特征. 例如, 本例中的模式时间特征为: “频繁发生于 07:07 至 07:17”.

6 用户意图推断与 TAP 规则生成

基于第 3–5 节提取的包含时间和空间上下文的频繁事件模式, HomeMind 执行用户意图推断与 TAP 规则生成任务. 流程分为 3 个连续阶段: 事件模式分析、用户意图推断和 TAP 规则生成, 对应“识别发生事件 (what)”“理解背后意图 (why)”以及“决定如何执行 (how)”的自然推理顺序.

(1) 提示词设计

为确保各阶段生成稳定、一致的输出, 本文为 HomeMind 设计了结构清晰的提示模板 (见图 10), 每个阶段均包含明确的任务说明和标准化的输入格式, 以引导 LLM 准确理解任务目标并生成规范结果. 三阶段推理流程统一采用思维链提示策略, 引导模型按步骤进行解释性推理.

- 事件模式分析. 模型分析给定频繁事件模式在时间与空间维度上的特性, 结合每个事件的发生时间与空间位置, 解释其在用户行为中的功能角色.

- 用户意图推断. 基于事件模式分析结果, 模型推断用户的潜在意图. 为降低生成发散性与幻觉风险, 并确保对相同语义场景的多次输出一致, 我们引入一个涵盖智能家居常见用户意图的预定义知识库, 引导模型优先在该

语义空间内归类. 同时, 通过提示词设计, 赋予模型生成新意图的能力. 知识库和提示词设计的细节将在后续内容中详细描述.

● TAP 规则生成. 根据推断的意图, 模型生成结构化的 TAP 规则, 明确自动化的触发条件和响应动作. 生成的规则可转换为自动化脚本, 支持如 Home Assistant 等主流智能家居平台的部署与执行.

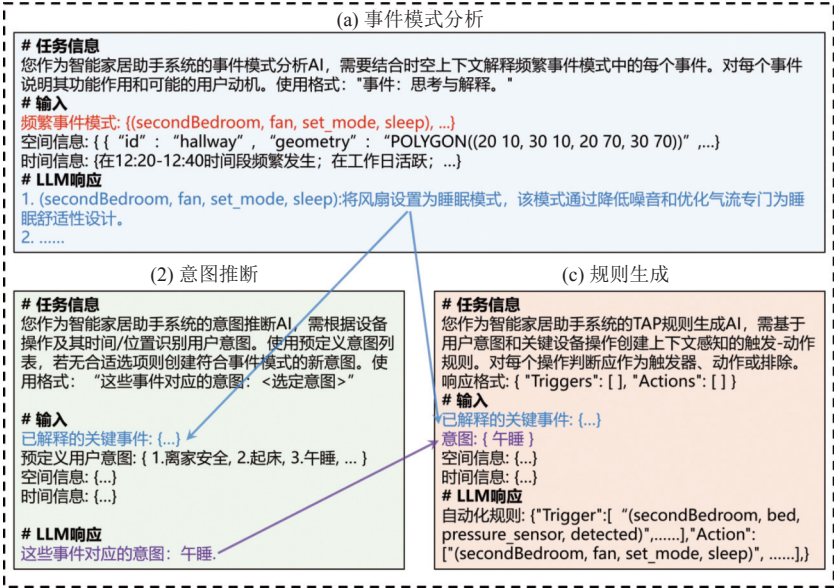


图 10 各个推理阶段的系统提示词结构

(2) 用户意图知识库

为辅助 LLM 进行意图推断, 本文构建了一个可扩展的智能家居意图知识库, 涵盖与自动化相关的常见用户意图类型. 知识库的初始构建基于多个公开数据集 (如 ARAS^[12]、CASAS^[14]、IFTTT^[21]) 及开源智能家居语料库^[13,22,23] 中的标注活动标签. 这些数据源提供了丰富的用户与智能设备的交互场景, 支持提取现实生活中的常见意图模式.

为确保知识库的高相关性和实用性, 我们过滤掉与设备操作无关的活动标签. 例如, 在 ARAS 数据集中, 刮胡子、换衣服、聊天等活动通常不触发设备操作或对应自动化规则, 因而被排除. 我们重点保留可从事件模式中推断并映射到自动化场景中的高层次意图类型, 包括日常例程 (如离家安全、起床、夜间睡眠) 以及环境控制 (如加湿、空气净化、采暖) 等场景. 最终, 知识库定义了图 11 所示的 32 种意图类型.

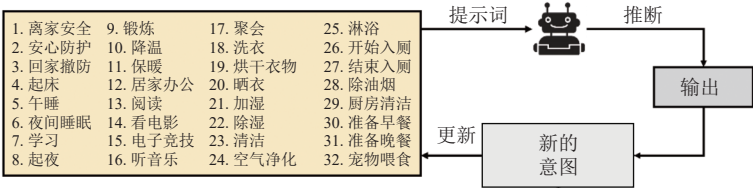


图 11 可维护和更新的用户意图知识库

此外, HomeMind 支持通过与 LLM 的反馈循环持续优化和扩展知识库. 系统通过提示词设计, 引导 LLM 在识别新行为模式或罕见场景时提出新意图建议. 经人工审核确认后, 合理的新意图被纳入知识库, 从而提升系统对不同家庭、用户习惯和场景的适应性.

7 面向终端用户的规则呈现与自动部署

在生成 TAP 规则后, HomeMind 将候选规则呈现给用户, 支持用户理解、比较和筛选适合其生活方式的自动

化方案. 为此, 本文设计了一个可视化的规则描述界面 (见图 12), 展示候选规则的详细信息, 并支持用户进行偏好选择和确认. 确认后的规则自动转换为目标智能家居平台支持的部署格式, 实现自动化方案的快速部署.

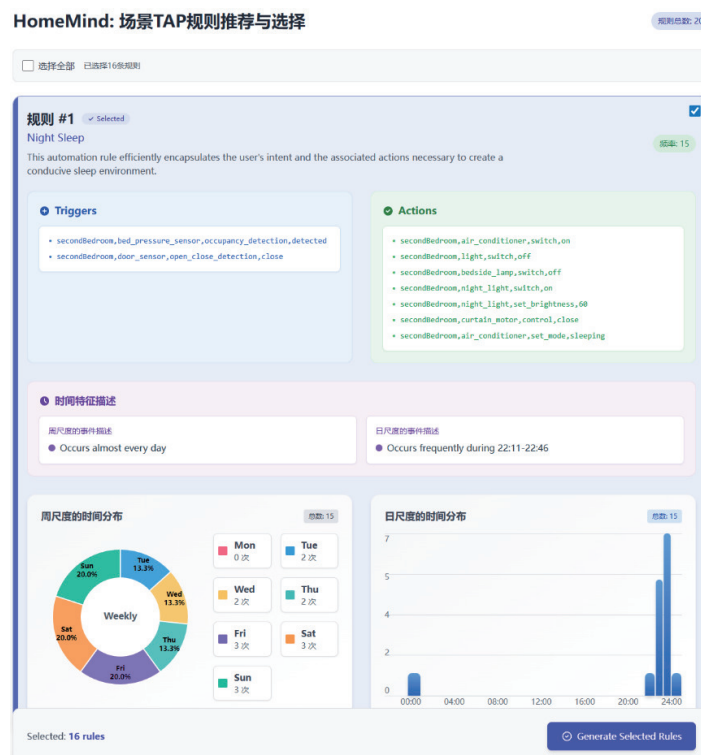


图 12 TAP 规则展示与选择界面

该界面包含以下关键组成部分.

- 意图类型: 展示规则对应的用户意图 (如“夜间睡眠”和“离家安全”), 便于用户从语义层面快速识别其功能目标.
- 功能描述: 简要说明规则在日常生活中的应用场景, 帮助用户理解其作用.
- 触发条件与执行动作: 详细列出规则的触发事件和相应的设备操作.
- 时间特征分析与分布: 可视化呈现规则关联的事件模式在周和日尺度上的分布, 展示触发频率和活跃时间段, 帮助用户评估规则适用性.

在规则确认环节, 用户可根据界面中提供的详细信息, 选择符合其偏好和场景需求的自动化规则. HomeMind 随后将选定规则转换为目标智能家居平台支持的标准化格式. 为实现这一过程, 本文设计了一个规则格式转换模块, 将结构化的 TAP 规则映射为符合 Home Assistant 语法规则的 YAML 脚本. 该模块基于预定义的模板库进行转换, 每类设备操作 (如开关控制、亮度调节、模式设定等) 对应特定的 YAML 模板. 模板包含参数占位符 (如 `{{entity_id}}`、`{{service}}`、`{{data}}` 等), 在转换过程中根据规则中的实体、服务类型和参数值自动填充. 整个转换过程依赖于从 Home Assistant 接口获取的实体映射表和服务映射表, 确保生成脚本与用户真实设备环境一致. 如后文图 13 所示, 以图 12 中展示的“夜间睡眠”规则为例, 系统可自动生成结构完整、语义明确的 Home Assistant 自动化配置, 实现了从用户意图推断、规则生成到部署执行的闭环, 提升了智能家居自动化的可用性与部署效率.

8 实验评估

本文围绕以下研究问题对 HomeMind 进行了系统性评估.

- RQ1: HomeMind 能否准确从智能家居事件日志中推断用户意图并生成正确的 TAP 规则?

- RQ2: HomeMind 中各核心工具和模块对 TAP 规则生成的性能有何影响?
- RQ3: 不同 LLM 的选择对生成规则的准确性、一致性与语义质量有何影响?

<pre> - id: "Night Sleep" alias: "Rule 1 for Night Sleep" trigger: - platform: state entity_id: secondBedroom.bed_pressure_sensor to: "detected" - platform: state entity_id: secondBedroom.door_sensor to: "close" </pre>	<pre> action: - service: switch.turn_on target: entity_id: secondBedroom.air_conditioner_switch - service: curtain_motor.control target: entity_id: secondBedroom.curtain_motor_control data: value: "close" - ... </pre>
---	--

图 13 TAP 规则转换成的 HomeAssistant 的自动化脚本

8.1 实验环境设置

(1) 真实世界数据集. 本文采用两个广泛使用的智能家居数据集对 HomeMind 进行评估: ARAS^[12]和 Kasteren^[13]. 其中, ARAS 数据集记录了两个家庭连续 30 天的细粒度活动与传感器日志, 涵盖多样化的日常行为场景; Kasteren 数据集则包含单人家庭的标注传感器序列, 常用于活动识别任务. 在意图推断评估中, 我们针对每个挖掘的频繁事件模式, 收集其全部实例所对应的活动标签, 并将这些标签视为该模式的意图基准 (ground truth). 由于这两个数据集均未提供 TAP 规则基准, 因此无法直接评估规则生成的准确性.

(2) 合成数据集. 为建立包含可靠基准的完整评估标准, 我们基于虚拟智能家居布局 (见图 4) 构建了合成数据集. 该布局包含 73 个智能设备, 涵盖 44 种设备类型, 支持 104 项带参数的独特服务, 模拟真实家庭的空间分布. 在此布局下, 我们手动设计了 35 个核心事件序列, 每个序列对应 32 种预定义用户意图中的一种 (其中有 3 种意图各对应两个候选序列). 这些序列被作为频繁模式基准, 嵌入为期 28 天、共计 5400 个事件的合成日常日志中. 为模拟现实世界的不确定性, 我们引入受控扰动, 包括事件重排、替换、删除及随机噪声注入. 以上涉及的所有数据集均已开源发布.

(3) TAP 规则基准的构建. 本文基于上述 35 个预定义的核心事件序列手动构建了 TAP 规则, 作为合成数据集的规则基准. 考虑到同一用户意图可能对应多个合理的事件序列, 我们并未强制模式与规则之间一一对应, 而是定义了如下的规则构建约束, 确保规则在具备语义正确性的同时保持结构多样性.

- 关键事件: 每条规则需包含反映意图核心语义的关键事件. 例如, 对于“午睡”意图, 触发器应包含 (bed, pressure_sensor, detected), 而动作应包含 (fan, set_mode, sleep) 和 (curtain_motor, control, close), 以营造舒适的睡眠环境.
- 传感器仅用于触发器: 涉及传感器的事件 (如压力或运动传感器) 仅允许出现在规则的触发器部分, 因为传感器是被动设备, 无法由用户主动控制.
- 设备逻辑顺序: 规则需遵循有效的设备使用逻辑和时间依赖关系. 例如, 在打开烤箱之前设置其模式 (例如, 触发器: (oven, set_mode, bake); 动作: (oven, switch, on)) 不符合实际操作流程, 应视为无效.

仅当生成的 TAP 规则同时满足上述全部约束条件时, 才视为正确, 这确保了规则在语法正确的基础上具有语义一致性和设备可执行性.

8.2 性能评估 (RQ1)

本文从两个角度评估 HomeMind 的性能: (1) 意图推断准确性. 我们评估 HomeMind 在各数据集上基于频繁事件模式推断用户意图的准确性, 重点判断推断出的结果是否与合成数据集的预定义意图标签或真实世界数据集的活动标签匹配. 该任务本质为多类别分类问题, 因此采用 $\text{Accuracy}@K$ 作为评估指标. 鉴于两个真实数据集中的事件模式数量较少, 我们设置 $K=10$; 对于合成数据集, 则分别评估 $K=10$ 和 $K=20$ 的表现. 由于现有基线方法不支持意图推断, 该项评估仅适用于 HomeMind. (2) TAP 规则生成质量. 在合成数据集上, 我们采用 $\text{Precision}@K$ 与 $\text{Recall}@K$ ($K=10, 20$) 来评估 HomeMind 生成 TAP 规则的质量, 衡量生成规则与手动定义规则基准的匹配程度. 由

于合成数据集中共定义了 35 条规则基准, 因此 Recall@10 和 Recall@20 的理论上限分别为 28.57% (10/35) 与 57.14% (20/35).

(1) 基线方法. 本文将 HomeMind 与以下 3 种具有代表性的 TAP 规则生成方法进行对比.

- 频繁模式挖掘 (frequent pattern mining, FPM): 该方法基于固定规则模板, 将频繁事件模式直接转换为 TAP 规则, 首个事件作为触发器, 其余事件按时间顺序作为动作.

- Trace2TAP^[8]: Trace2TAP 需要预先指定动作事件作为输入, 随后根据时间相关性选取最高的 3 个事件作为触发器. 为公平比较, 我们提供了正确的动作, 仅评估其触发器预测性能.

- SGen^[9]: SGen 在滑动时间窗口内挖掘事件轨迹中的频繁事件对, 并基于共现关系构建规则.

(2) 结果分析. 实验使用 GPT-4o 作为 HomeMind 默认的 LLM 推理后端. 为降低模型输出的随机性, 每个输入模式重复测试 4 次, 取平均得分.

表 2 展示了 3 个数据集上的意图推断评估结果. HomeMind 在所有数据集上均实现了 70% 以上的准确率, 表明其能够有效地将事件模式与用户意图相匹配. 相比合成数据集, HomeMind 在真实世界数据集上的准确率略低, 主要原因是传感器数量的限制. 例如, ARAS 数据集包含了 26 种意图标签, 但仅有 20 个传感器, 导致部分事件模式可能同时对应多个意图标签. 此外, 合成数据集中的推理错误主要源于 LLM 的幻觉或在相似意图 (如“阅读”和“学习”) 间混淆.

表 2 HomeMind 的意图推断结果 (%)

数据集	Accuracy@10	Accuracy@20
Kasteren	70.00	—
ARAS	75.00	—
合成数据集	80.00	88.75

图 14 展示了各方法在合成数据集上的 TAP 规则生成性能. HomeMind 利用 GPT-4o 为每个输入模式生成 4 条候选规则, 并取平均结果. 相较所有基线方法, HomeMind 在 Precision@K 和 Recall@K 上均表现出显著优势, 表明其具备生成准确、多样且符合用户意图的 TAP 规则的能力. 这一优势得益于 HomeMind 融合了时间、空间与意图信息的上下文增强推理能力, 有效弥合了用户执行日志与结构化场景规则之间的语义鸿沟, 而非仅依赖于原始的事件轨迹.

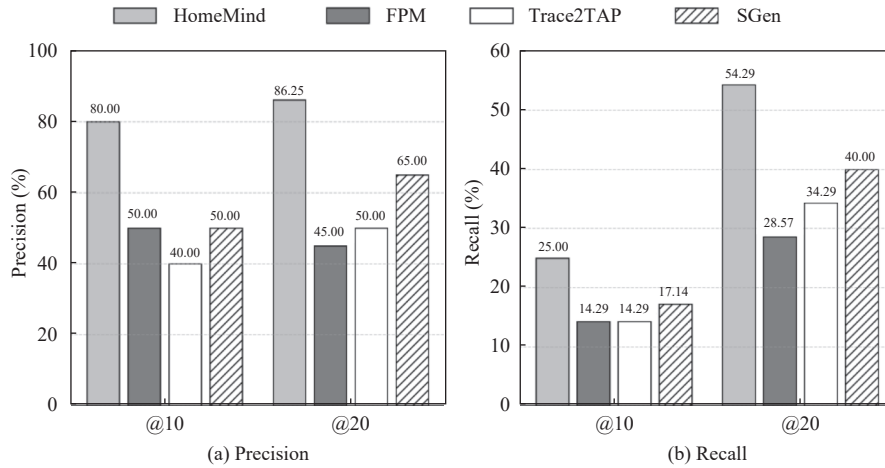


图 14 合成数据集上的规则生成性能

8.3 工具影响性分析 (RQ2)

为评估 HomeMind 中各工具和上下文模块对整体性能的贡献, 我们设计并实施了一项消融实验, 通过从提

示词中移除关键上下文信息,分析各工具和模块的独立作用和影响性.具体而言,我们构建了以下3种简化变体:(1)无意图推断:跳过意图推断,直接基于频繁事件模式生成TAP规则;(2)无空间信息:从输入提示中移除空间位置信息;(3)无时间特征:从输入提示中移除时间特征信息.我们在合成数据集上使用GPT-4o对所有变体进行评估.为降低输出随机性,每个输入事件模式重复测试4次,取平均结果.评估指标包括Precision@10、Precision@20、Recall@10和Recall@20.

图15展示了完整HomeMind与各消融变体的性能对比.完整HomeMind取得了最佳性能,Precision@10为80.00%,Precision@20为86.25%,Recall@10和Recall@20分别为25.00%和54.29%.在所有消融项中,移除意图推断所导致的性能下降最为显著,表明意图推断在生成语义一致和实用的规则中发挥了关键作用.移除空间信息与移除时间特征同样导致性能下降,说明两者对区分相似事件模式、引导LLM进行准确意图推断及规则生成具有重要作用.

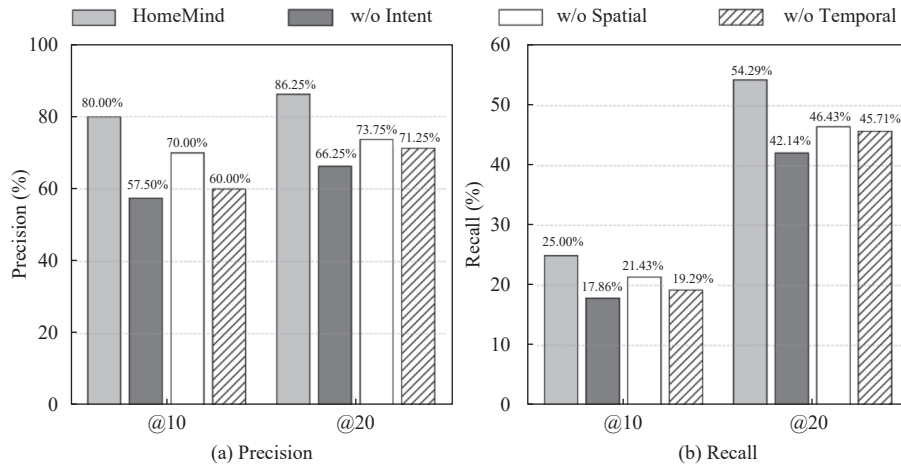


图15 HomeMind 及其消融变体的性能

进一步分析表明,当前LLM在理解部分意图时仍存在局限,因此进行意图推断并在提示词中显式提供意图标签是必要的.例如,“电竞娱乐”意图对应的事件模式包括关闭书房门、拉上窗帘、开启音响和显示器设置游戏模式等操作.在没有正确推断出该意图的情况下,LLM常误将“开启音响”识别为触发条件,忽略了其作为辅助设备的角色,最终生成不符合语义的错误规则.在该场景中,包含意图推断过程的生成准确率为4/4,而不包含意图推断过程时准确率仅为0/4.

空间和时间信息对于引导LLM理解用户行为意图、避免生成幻觉内容也发挥了重要作用.空间信息有助于LLM建立设备关联,避免遗漏关键条件.例如,在“空气净化”场景中,规则基于客厅空气质量传感器检测到PM_{2.5}>40后,触发走廊的空气净化器开启.若提示词中缺少空间信息,LLM可能误判设备无关联,从而删除触发条件,生成无效规则.时间特征则有助于区分相似但语义不同的行为意图.例如,“起夜”和“夜间睡眠”两个场景都以床压力传感器触发作为规则的触发器,但意图不同.若缺乏时间上下文,LLM常将“起夜”误判为“夜间睡眠”,生成不合适的规则.

此外,为进一步验证用户意图知识库(详见第6节)对规则生成质量的影响,本文设计了补充实验,将HomeMind与其不包含意图知识库的变体进行对比.结果如后文表3所示,在移除意图知识库后,模型在各项指标上均出现一定程度的下降.通过对生成错误的案例进行分析,我们发现性能下降主要源于某些特定意图在缺少知识库支持时难以被准确识别.例如,在面对一系列夜间传感器事件(如床压力传感器触发、卫生间灯光开启)时,若缺乏先验知识,LLM可能无法联想到“起夜”这一典型场景,进而错误地将其归类为“夜间睡眠”等其他意图,最终导致触发条件与动作设置均出现偏差.

综上所述,消融实验表明,意图推断和上下文建模等工具和模块显著提升了TAP规则生成的准确性和适用性.

8.4 不同 LLM 对规则生成的影响 (RQ3)

为评估不同 LLM 在意图推断与 TAP 规则生成中的表现差异, 我们对比了 4 种先进模型: GPT-4o^[24]、Claude 3.5 Haiku^[25]、Gemini 2.0 Flash^[26]和 LLaMA 4 Scout^[27]. 实验中, 每个模型针对每个事件模式均执行 4 次推断与规则生成, 并取平均值以降低随机性. 所有模型均采用其默认温度配置, 模拟真实应用场景下的典型使用条件.

表 3 意图知识库对 HomeMind 规则生成的影响 (%)

方法	Precision@10	Precision@20	Recall@10	Recall@20
HomeMind	80.00	86.25	25.00	54.29
缺少意图知识库的HomeMind	72.50	80.00	22.86	50.71

(1) 结果分析. 表 4 展示了各模型在意图推断任务中的准确率表现, 图 16 展示了它们在 TAP 规则生成任务中的整体性能. 整体来看, 不同模型在各项指标上的表现大体接近, 验证了 HomeMind 在不同 LLM 后端下的普适性与稳定性, 无需依赖特定模型即可可靠运行.

表 4 不同 LLM 的意图推断结果 (%)

模型	Accuracy@10	Accuracy@20
GPT-4o	80.00	88.75
Claude 3.5 Haiku	75.00	82.50
Gemini 2.0 Flash	72.50	81.25
LLaMA 4 Scout	62.50	78.75

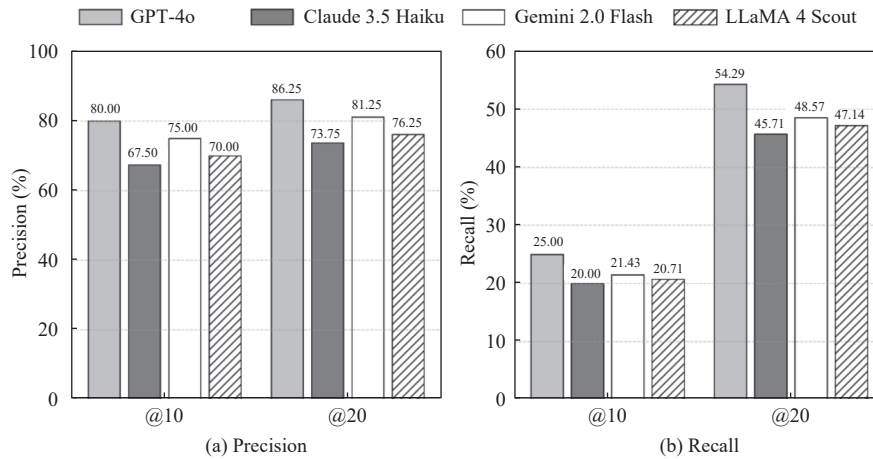


图 16 不同 LLM 的规则生成性能

(2) 行为分析. 尽管整体性能接近, 不同 LLM 在推理和生成行为上呈现独特特性.

- 泛化与创造性推理: GPT-4o 和 LLaMA 4 Scout 偶尔会在预定义意图集合之外推断出新的复合意图. 例如, 面对包含“准备晚餐”和“回家撤防”等事件序列时, 这些模型会合成出“回家吃晚饭”这一新意图类别, 展现出较强的泛化与组合推理能力. 相比之下, Claude 3.5 Haiku 和 Gemini 2.0 Flash 更倾向于严格遵循已知意图类别.

- 输出一致性: Claude 3.5 Haiku 在多轮提示下表现出最高的一致性, 生成几乎相同的规则, 这种一致性有助于智能家居环境中规则的可靠部署.

- 逻辑约束遵循性: Claude 3.5 Haiku 和 LLaMA 4 Scout 有时会生成违反设备逻辑约束的规则. 例如, 某些规则在尚未启动烤箱前便尝试设置烤箱模式 (如触发器: (oven, set_mode, bake), 动作: (oven, switch, on)), 这种顺序违背了设备的实际操作逻辑和时间依赖关系.

9 讨论

9.1 有效性威胁

HomeMind 性能的主要威胁源于真实世界智能家居数据集的局限性. 现有大多数公开数据集多聚焦于底层传感器的触发事件, 缺乏对多样化设备及服务级别事件的全面记录. 为缓解这一问题, 本文人工构建了高度模拟真实家庭设备交互的合成数据集, 基于常见家庭日常活动精心设计设备序列; 同时, 我们还通过人工增强方式扩充了两个广泛使用的真实数据集 (ARAS 和 Kasteren) 的语义信息, 使其更适用于自动化场景.

此外, LLM 生成结果的发散性和幻觉风险构成另一威胁. 为应对此问题, 本文采取了两方面的缓解策略: 一是精心设计系统提示词, 提供清晰、结构化的指令与上下文信息, 引导模型按预期推理; 二是提供预定义的意图集合, 约束模型在可控的标签空间内进行分类与生成, 降低生成不合理意图的概率.

9.2 局限性

- 冷启动问题. HomeMind 依赖历史设备日志以挖掘事件模式并推断用户意图. 在新部署环境或新添加设备尚无足够历史数据时, 系统难以提取有效模式, 限制了其在早期阶段生成高质量 TAP 规则的能力.

- 聚焦单一用户场景. HomeMind 目前聚焦于单一用户场景的建模, 针对个体用户的行为模式与意图推断. 在多用户共存环境中, 用户间意图重叠、行为冲突与协同的判别仍是重要挑战.

10 相关工作

现有的生成智能家居自动化规则的方法可大致分为两类: (1) 基于自然语言描述生成规则; (2) 从日志中挖掘设备使用模式. 此外, 近期迅速发展的 LLM 智能体为规则生成提供了新的能力支撑和设计范式, 成为推动自动化系统更高层次智能的关键基础设施.

10.1 基于自然语言描述的规则生成

此类方法根据用户提供的自然语言指令生成 TAP 规则, 主要分为基于传统机器学习方法与基于 LLM 的方法.

- 基于传统机器学习的方法. 早期方法主要依赖句法与语义解析, 将用户描述映射为结构化规则 (如 IFTTT 配方). Quirk 等人^[28]提出了一种基于抽象语法树的语义解析器, 可在弱监督数据上进行训练; Beltagy 等人^[29]将规则生成建模为使用语言特征的结构化预测任务; Dong 等人^[30]和 Chen 等人^[31]引入了注意力机制, 可以更好地捕捉触发语句与动作语句间的依赖关系. RecipeGen^[32]通过将 TAP 生成建模为语言建模任务, 引入基于 Transformer 的模型以提升语义一致性. 此外, 部分工作要求用户自行将自然语言描述转换为领域特定语言 (DSL)^[33,34]. 然而, 这些方法普遍依赖预定义的规则模板, 限制了其灵活性与泛化能力.

- 基于 LLM 的方法. 近期有部分研究工作利用 LLM 来生成更具适应性的规则. Sasha^[10]采用模块化流程, 依次执行目标澄清、设备筛选、动作规划与反馈精炼; ChatIoT^[11]提供了一套零代码解决方案, 结合上下文感知提示与强化学习进行个性化定制, 并通过蒸馏多模态大模型实现动态感知扩展. 然而, 这些方法依然依赖于用户清晰指定的意图与明确的输入描述, 在模糊或欠规范化场景下的适用性仍然有限.

10.2 基于设备使用日志的模式挖掘

此类工作通过分析历史设备使用日志, 旨在预测后续动作或直接合成自动化规则.

- 基于深度学习的活动预测. 智能家居场景下的深度学习方法^[35-40]主要聚焦于活动预测, 但普遍面临将高层预测结果转化为可执行 TAP 规则的挑战. 近期研究通过建模细粒度的用户-设备交互行为, 尝试弥合这一差距. SmartSense^[41]利用基于查询的 Transformer 捕捉行为序列中的上下文依赖关系; SmartUDI^[42]和 DeepUDI^[43]结合胶囊图网络、关系门控图神经网络及历史注意力机制, 有效学习用户日常、并发意图及多层次周期性行为; ASGen^[19]融合序列上下文、事件语义与外部领域知识, 合成个性化设备操作序列, 进一步向自动化规则生成迈进. 然而, 这些方法主要聚焦于根据当前状态下的事件或上下文作为触发条件, 预测接下来的动作. 这种方法更偏向于短期的下一步动作预测, 缺乏对全局行为模式的建模与抽象, 难以生成结构化、可显式表达和复用的 TAP 规则.

● 规则合成. 部分研究则尝试直接从历史事件轨迹中合成可重用的 TAP 规则. Helion^[7]通过连接现有的触发器-动作对构建长距离的 T-A-T-A 序列, 并利用 N-gram 模型捕捉常见自动化模式; Trace2TAP^[8]结合符号推理与约束求解, 从用户轨迹中生成基础的 TAP 规则; TAGen^[44]和 SGen^[9]则进一步通过挖掘智能家居日志中的频繁与周期性模式来构建规则. 尽管这些方法在特定场景下表现出一定效果, 但普遍依赖于稳定的时间周期假设, 而用户的探索性或非规律行为往往会打破这一前提. 此外, 这些方法多聚焦于单一触发与单一动作的规则结构, 限制了其在处理复杂自动化需求时的适用性与灵活性.

10.3 大语言模型智能体

近年来, 基于 LLM 的智能体系统发展迅速, 其核心目标是赋予模型自主执行复杂任务的能力. 这类智能体通常具备任务分解、策略规划、工具调用与上下文感知推理等能力, 能够完成多阶段或多子任务协同的复杂流程.

CoT 是最早应用于智能体设计的关键技术之一^[45], 它通过在提示词中引入多个逐步推理的示例, 引导模型按照逻辑链条展开复杂思维. 进一步的工作如零样本 CoT^[46], 则仅需在提示中加入如“让我们一步一步思考”这样的语句, 即可激发模型的推理能力. 此外, 还有研究提出将 CoT 扩展为树状结构, 使每个中间步骤可派生出多个并行的子步骤^[47], 从而提升推理的层次性和表达能力. 随着智能体能力的增强, 环境交互反馈机制被引入推理流程中. 例如, ReAct 智能体^[48]将环境反馈整合至每个推理步骤中, 使智能体能够基于观察结果动态调整推理路径和计划策略. 同时, 人类反馈也被用于策略优化, 如鼓励模型主动提问以获取更多偏好或上下文, 从而提升行为决策的适应性与准确性.

在此基础上, 工具调用能力成为 LLM 智能体的一项核心特性. 智能体可调用外部 API、访问知识库或执行代码, 从而突破模型参数固有知识的限制. 例如, Gorilla^[49]和 ToolLLM^[50]专门对模型进行了训练, 使其能够更高效地解析 API 文档并执行正确的调用. 目前, LLM 智能体已广泛应用于各类复杂的规划与推理任务中. 它们可以与数字环境进行交互, 如调用 API^[51,52]、浏览网页^[53]、执行代码^[54]; 也可以作用于物理环境, 包括操控实体对象^[55]、执行实验任务^[56]或进行自主行为决策^[57].

在本文中, 我们提出的 HomeMind 智能体基于上下文增强的提示词机制, 将从日志中挖掘的频繁事件模式、时间与空间特征以及用户意图进行有效整合, 引导 LLM 进行多阶段的链式推理并合成结构化的场景规则. 该智能体架构不仅继承了 CoT 推理与环境反馈机制的优势, 还具备任务分解与规则部署等能力, 成功构建了从设备日志到高质量自动化场景规则的智能闭环.

11 总 结

随着智能设备规模的快速增长和用户意图的复杂化, 终端用户构建适配的智能家居自动化规则面临诸多挑战. 本文提出了 HomeMind, 一个以 LLM 为核心, 集成感知、推理和协同能力的智能体系统, 用于自动生成复杂智能家居场景的 TAP 规则. 该系统从设备日志中挖掘频繁事件模式, 结合多尺度的时间与空间上下文进行语义增强, 构建上下文增强的提示词机制, 引导 LLM 通过思维链推理用户意图并生成语义明确的 TAP 规则. 实验结果表明, HomeMind 在意图推断准确率与场景规则生成质量上均显著优于现有基线方法, 验证了其有效性.

References

- [1] Ur B, Ho MPY, Brawner S, Lee J, Mennicken S, Picard N, Schulze D, Littman ML. Trigger-action programming in the wild: An analysis of 200000 IFTTT recipes. In: Proc. of the 2016 CHI Conf. on Human Factors in Computing Systems. San Jose: ACM, 2016. 3227–3231. [doi: 10.1145/2858036.2858556]
- [2] IFTTT. <https://ifttt.com/>
- [3] SmartThings. <https://www.samsung.com.cn/smartthings/>
- [4] MiHome (in Chinese). <http://home.mi.com/>
- [5] Home Assistant. <https://www.home-assistant.io/>
- [6] OpenHAB. <https://www.openhab.org/>
- [7] Manandhar S, Moran K, Kafle K, Tang RH, Poshyvanyk D, Nadkarni A. Towards a natural perspective of smart homes for practical

- security and safety analyses. In: Proc. of the 2020 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2020. 482–499. [doi: [10.1109/SP40000.2020.00062](https://doi.org/10.1109/SP40000.2020.00062)]
- [8] Zhang LF, He WJ, Morkved O, Zhao V, Littman ML, Lu S, Ur B. Trace2TAP: Synthesizing trigger-action programs from traces of behavior. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2020, 4(3): 104. [doi: [10.1145/3411838](https://doi.org/10.1145/3411838)]
 - [9] Liu LW, Chen W, Wang T, Wang W, Wu GQ, Wei J. Generating scenario-centric TAP rules for smart homes by mining historical event logs. In: Proc. of the 2023 IEEE Int'l Conf. on Web Services. Chicago: IEEE, 2023. 21–27. [doi: [10.1109/ICWS60048.2023.00015](https://doi.org/10.1109/ICWS60048.2023.00015)]
 - [10] King E, Yu HX, Lee S, Julien C. Sasha: Creative goal-oriented reasoning in smart homes with large language models. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2024, 8(1): 12. [doi: [10.1145/3643505](https://doi.org/10.1145/3643505)]
 - [11] Gao Y, Xiao KJ, Li F, Xu WF, Huang JM, Dong W. ChatIoT: Zero-code generation of trigger-action based IoT programs. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2024, 8(3): 103. [doi: [10.1145/3678585](https://doi.org/10.1145/3678585)]
 - [12] Alemdar H, Ertan H, Incel OD, Ersoy C. ARAS human activity datasets in multiple homes with multiple residents. In: Proc. of the 7th Int'l Conf. on Pervasive Computing Technologies for Healthcare. Venice: IEEE, 2013. 232–235.
 - [13] van Kasteren T, Noulas A, Englebienne G, Kröse B. Accurate activity recognition in a home setting. In: Proc. of the 10th Int'l Conf. on Ubiquitous Computing. Seoul: ACM, 2008. 1–9. [doi: [10.1145/1409635.1409637](https://doi.org/10.1145/1409635.1409637)]
 - [14] CASAS. 2024. <https://casas.wsu.edu/datasets/>
 - [15] IndoorGML. <https://www.indoorgml.net/>
 - [16] Han JW, Pei J, Yin YW. Mining frequent patterns without candidate generation. In: Proc. of the 2000 ACM SIGMOD Int'l Conf. on Management of Data. Dallas: ACM, 2000. 1–12. [doi: [10.1145/342009.335372](https://doi.org/10.1145/342009.335372)]
 - [17] Quigley B, Donnelly M, Moore G, Galway L. A comparative analysis of windowing approaches in dense sensing environments. MDPI Proc., 2018, 2(19): 1245. [doi: [10.3390/proceedings2191245](https://doi.org/10.3390/proceedings2191245)]
 - [18] Krishnan NC, Cook DJ. Activity recognition on streaming sensor data. Pervasive and Mobile Computing, 2014, 10: 138–154. [doi: [10.1016/j.pmcj.2012.07.003](https://doi.org/10.1016/j.pmcj.2012.07.003)]
 - [19] Song YP, Bian YH, Wang XR, Cai ZM. Learning from user-driven events to generate automation sequences. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2023, 7(4): 177. [doi: [10.1145/3631427](https://doi.org/10.1145/3631427)]
 - [20] Parzen E. On estimation of a probability density function and mode. The Annals of Mathematical Statistics, 1962, 33(3): 1065–1076. [doi: [10.5109/13049](https://doi.org/10.5109/13049)]
 - [21] Yu HX, Hua J, Julien C. Analysis of IFTTT recipes to study how humans use Internet-of-Things (IoT) devices. In: Proc. of the 19th ACM Conf. on Embedded Networked Sensor Systems. Coimbra: ACM, 2021. 537–541. [doi: [10.1145/3485730.3494115](https://doi.org/10.1145/3485730.3494115)]
 - [22] iResearch Consulting Group. China Smart Home (AloH) Development White Paper 2023. In: Proc. of the iResearch April Research Report Conf. 2023. Beijing: iResearch Consulting Group, 2023. 408–448 (in Chinese).
 - [23] iResearch Consulting Group. Personalized and Seamless Intelligence: China Whole-house Smart Home Industry Research White Paper. In: iResearch Industry Research Report Series (No. 3, 2022). Beijing: iResearch Consulting Group, 2022. 1302–1346 (in Chinese).
 - [24] GPT-4o. 2024. <https://openai.com/index/hello-gpt-4o/>
 - [25] Claude 3.5 Haiku. 2024. <https://www.anthropic.com/claude/haiku/>
 - [26] Gemini 2.0 Flash. 2024. <https://deepmind.google/models/gemini/flash/>
 - [27] LLaMA 4 Scout. 2025. <https://www.llama.com/models/llama-4/>
 - [28] Quirk C, Mooney R, Galley M. Language to code: Learning semantic parsers for if-this-then-that recipes. In: Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing (Vol. 1: Long Papers). Beijing: ACL, 2015. 878–888. [doi: [10.3115/v1/P15-1085](https://doi.org/10.3115/v1/P15-1085)]
 - [29] Beltagy I, Quirk C. Improved semantic parsers for if-then statements. In: Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Berlin: ACL, 2016. 726–736. [doi: [10.18653/v1/P16-1069](https://doi.org/10.18653/v1/P16-1069)]
 - [30] Dong L, Lapata M. Language to logical form with neural attention. In: Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Berlin: ACL, 2016. 33–43. [doi: [10.18653/v1/P16-1004](https://doi.org/10.18653/v1/P16-1004)]
 - [31] Chen XY, Liu C, Shin R, Song D, Chen MC. Latent attention for if-then program synthesis. In: Proc. of the 30th Int'l Conf. on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016. 4581–4589.
 - [32] Yusuf INB, Jiang LX, Lo D. Accurate generation of trigger-action programs with domain-adapted sequence-to-sequence learning. In: Proc. of the 30th IEEE/ACM Int'l Conf. on Program Comprehension. ACM, 2022. 99–110. [doi: [10.1145/3524610.3527922](https://doi.org/10.1145/3524610.3527922)]
 - [33] Wang B, Zhang Y, Geng JN, Li XY. SSRules: Make it easier to write and check automation rules for smart home systems. Ruan Jian Xue Bao/Journal of Software, 2021, 32(12): 3728–3750 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6098.htm> [doi: [10.13328/j.cnki.jos.006098](https://doi.org/10.13328/j.cnki.jos.006098)]

- [34] Bian H, Chen XH, Jin Z, Zhang M. Approach to generating TAP rules in IoT systems based on environment modeling. *Ruan Jian Xue Bao/Journal of Software*, 2021, 32(4): 934–952 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6224.htm> [doi: 10.13328/j.cnki.jos.006224]
- [35] Krishna K, Jain D, Mehta SV, Choudhary S. An LSTM based system for prediction of human activities with durations. *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2017, 1(4): 147. [doi: 10.1145/3161201]
- [36] Maharjan R, Shiraishi K, Yamamoto T, Yamamoto Y, Ohshima H. Development of IoT monitoring device and prediction of daily life behavior. In: *Proc. of the 21st Int'l Conf. on Information Integration and Web-based Applications & Services*. Munich: ACM, 2019. 584–588. [doi: 10.1145/3366030.3366123]
- [37] Mohamed M, El-Kilany A, El-Tazi N. Future activities prediction framework in smart homes environment. *IEEE Access*, 2022, 10: 85154–85169. [doi: 10.1109/ACCESS.2022.3197618]
- [38] Almeida A, Bermejo U, Bilbao A, Azkune G, Aguilera U, Emaldi M, Dornaika F, Arganda-Carreras I. A comparative analysis of human behavior prediction approaches in intelligent environments. *Sensors*, 2022, 22(3): 701. [doi: 10.3390/s22030701]
- [39] Wang W, Li JB, Li Y, Dong XS. Predicting activities of daily living for the coming time period in smart homes. *IEEE Trans. on Human-machine Systems*, 2023, 53(1): 228–238. [doi: 10.1109/THMS.2022.3176213]
- [40] Dunne R, Morris T, Harper S. A semantic blocks model for human activity prediction in smart environments using time-windowed contextual data. *Journal of Reliable Intelligent Environments*, 2023, 9(1): 87–95. [doi: 10.1007/s40860-022-00194-1]
- [41] Jeon H, Kim J, Yoon H, Lee J, Kang U. Accurate action recommendation for smart home via two-level encoders and commonsense knowledge. In: *Proc. of the 31st ACM Int'l Conf. on Information and Knowledge Management*. Atlanta: ACM, 2022. 832–841. [doi: 10.1145/3511808.3557226]
- [42] Xiao JY, Zou QS, Li Q, Zhao D, Li K, Weng ZX, Li RY, Jiang Y. I know your intent: Graph-enhanced intent-aware user device interaction prediction via contrastive learning. *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2023, 7(3): 136. [doi: 10.1145/3610906]
- [43] Xiao JY, Zou QS, Li Q, Zhao D, Li K, Tang WX, Zhou RJ, Jiang Y. User device interaction prediction via relational gated graph attention network and intent-aware encoder. In: *Proc. of the 2023 Int'l Conf. on Autonomous Agents and Multiagent Systems*. London: Int'l Foundation for Autonomous Agents and Multiagent Systems, 2023. 1634–1642.
- [44] Liu LW, Chen W, Liu L, Zhang KK, Wei J, Yang Y. TAGen: Generating trigger-action rules for smart homes by mining event traces. In: *Proc. of the 19th Int'l Conf. on Service-Oriented Computing*. Springer, 2021. 652–662. [doi: 10.1007/978-3-030-91431-8_41]
- [45] Wei J, Wang XZ, Schuurmans D, Bosma M, Ichter B, Xia F, Chi EH, Le QV, Zhou D. Chain-of-thought prompting elicits reasoning in large language models. In: *Proc. of the 36th Int'l Conf. on Neural Information Processing Systems*. New Orleans: Curran Associates Inc., 2022. 24824–24837.
- [46] Kojima T, Gu SS, Reid M, Matsuo Y, Iwasawa Y. Large language models are zero-shot reasoners. In: *Proc. of the 36th Int'l Conf. on Neural Information Processing Systems*. New Orleans: Curran Associates Inc., 2022. 22199–22213.
- [47] Yao SY, Yu D, Zhao J, Shafraan I, Griffiths TL, Cao Y, Narasimhan K. Tree of thoughts: Deliberate problem solving with large language models. In: *Proc. of the 37th Int'l Conf. on Neural Information Processing Systems*. New Orleans: Curran Associates Inc., 2023. 11809–11822.
- [48] Yao SY, Zhao J, Yu D, Du N, Shafraan I, Narasimhan KR, Cao Y. ReAct: Synergizing reasoning and acting in language models. In: *Proc. of the 2023 Int'l Conf. on Learning Representations*. OpenReview.net, 2023.
- [49] Patil SG, Zhang TJ, Wang X, Gonzalez JE. Gorilla: Large language model connected with massive APIs. In: *Proc. of the 38th Int'l Conf. on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2024. 126544–126565.
- [50] Qin YJ, Liang SH, Ye YN, Zhu KL, Yan L, Lu YX, Lin YK, Cong X, Tang XR, Qian B, Zhao SH, Hong L, Tian RC, Xie RB, Zhou J, Gerstein M, Li DH, Liu ZY, Sun MS. ToolLLM: Facilitating large language models to master 16000+ real-world APIs. *arXiv:2307.16789*, 2023.
- [51] Ge YQ, Hua WY, Mei K, Ji JC, Tan JT, Xu SY, Li ZL, Zhang YF. OpenAGI: When LLM meets domain experts. In: *Proc. of the 37th Int'l Conf. on Neural Information Processing Systems*. New Orleans: Curran Associates Inc., 2023. 5539–5568.
- [52] Schick T, Dwivedi-Yu J, Dessi R, Raileanu R, Lomeli M, Hambro E, Zettlemoyer L, Cancedda N, Scialom T. Toolformer: Language models can teach themselves to use tools. *Proc. of the 37th Int'l Conf. on Neural Information Processing Systems*. New Orleans: Curran Associates Inc., 2023. 68539–68551.
- [53] Nakano R, Hilton J, Balaji S, Wu J, Ouyang L, Kim C, Hesse C, Jain S, Kosaraju V, Saunders W, Jiang X, Cobbe K, Eloundou T, Krueger G, Button K, Knight M, Chess B, Schulman J. WebGPT: Browser-assisted question-answering with human feedback. *arXiv:2112.09332*, 2022.

- [54] Zhang KC, Zhang HZ, Li G, Li J, Li Z, Jin Z. ToolCoder: Teach code generation models to use API search tools. arXiv:2305.04032, 2023.
- [55] Ahn M, Brohan A, Brown N, *et al.* Do as I can, not as I say: Grounding language in robotic affordances. arXiv:2204.01691, 2022.
- [56] Boiko DA, MacKnight R, Gomes G. Emergent autonomous scientific research capabilities of large language models. arXiv:2304.05332, 2023.
- [57] Huang WL, Abbeel P, Pathak D, Mordatch I. Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. In: Proc. of the 39th Int'l Conf. on Machine Learning. Baltimore: PMLR, 2022. 9118–9147.

附中文参考文献

- [4] 米家. <http://home.mi.com/>
- [22] 上海艾瑞市场咨询有限公司. 2023 年中国智能家居 (AioH) 发展白皮书. 2023 艾瑞咨询 4 月研究报告会论文集. 北京: 艾瑞咨询集团, 2023. 408–448.
- [23] 艾瑞咨询集团. 千人千面、无感智能: 中国全屋智能行业研究白皮书. 艾瑞咨询系列研究报告 (2022 年第 3 期). 北京: 艾瑞咨询集团, 2022. 1302–1346.
- [33] 王博, 张昱, 耿佳宁, 李向阳. SSRules: 让智能家居自动化规则更易于编写和检查. 软件学报, 2021, 32(12): 3728–3750. <http://www.jos.org.cn/1000-9825/6098.htm> [doi: 10.13328/j.cnki.jos.006098]
- [34] 边寒, 陈小红, 金芝, 张民. 基于环境建模的物联网系统 TAP 规则生成方法. 软件学报, 2021, 32(4): 934–952. <http://www.jos.org.cn/1000-9825/6224.htm> [doi: 10.13328/j.cnki.jos.006224]

作者简介

刘力玮, 博士生, 主要研究领域为智能化软件工程, 物联网.

陈伟, 博士, 研究员, 博士生导师, CCF 专业会员, 主要研究领域为智能化软件工程, 人机物融合计算, 云原生系统与服务计算.

魏峻, 博士, 研究员, 博士生导师, CCF 高级会员, 主要研究领域为智能化软件工程, 分布式计算.

朱家鑫, 博士, 副研究员, CCF 专业会员, 主要研究领域为软件工程, 基于人工智能的软件开发方法, 开源软件.

王伟, 博士, 研究员, 博士生导师, CCF 高级会员, 主要研究领域为分布式系统软件.

吴国全, 博士, 研究员, 博士生导师, CCF 高级会员, 主要研究领域为软件工程, 软件测试与维护, 面向服务的计算.

汪涛, 博士, 主要研究领域为 Code Agent, LLM 推理优化.