# 基于NL2SQL技术的电力生产经营数据智能问数引擎研发

胡　越　蒲国庆　范义平　李妍彦

四川华电泸定水电有限公司　四川　甘孜州　626700

**摘　要：**为解决电力企业非技术人员利用生产经营数据的技术门槛问题，本文研发基于 NL2SQL 技术的智能问数引擎。该引擎融合自然语言处理、知识图谱与深度学习技术，针对电力数据多源异构、高维复杂等特点，设计分层架构，构建领域词库与知识图谱，实现自然语言到SQL的高效转换及可视化分析。测试表明，引擎单表查询准确率达98.5%，复杂查询响应时间平均5.3秒，用户满意度89.7%，显著提升数据利用效率，为电力企业决策提供支撑。

**关键词：**NL2SQL；电力生产经营数据；智能问数引擎；知识图谱；数据可视化

## 引　言：

随着电力行业数字化转型推进，海量生产经营数据蕴含巨大价值，但传统查询依赖专业SQL知识，限制非技术人员使用。电力数据具有多源异构、高维复杂、强实时性等特点，如何实现便捷查询分析成为关键挑战。NL2SQL技术可将自然语言转换为SQL语句，降低操作门槛。本文提出融合多技术的智能问数引擎方案，通过架构设计、词库构建、算法优化等实现高效数据交互与可视化，验证其在电力领域的实用性。

## 一、相关技术与研究现状

在电力生产经营数据处理中，NL2SQL技术虽可实现非技术人员与数据库交互，但其在电力领域面临专业术语理解和复杂查询转换难题。电力生产经营数据呈现多源异构、高维度复杂、强实时性、专业性强及规模庞大的特点，数据来源广泛，涵盖多系统、多维度，且需满足实时监控调度需求，智能电网的发展更推动数据量爆发式增长。目前，现有的电力数据查询系统多针对故障诊断、调控运行等特定场景，如电力调控智能搜索引擎、电网智慧语者系统等，缺乏面向生产经营数据的通用引擎，在多表关联、动态条件处理方面存在不足，难以满足多样化的数据分析需求，亟待研发适配电力生产经营数据特性的专项技术解决方案[1]。

表1　智能问数引擎总体架构

| 层级 | 核心组件 | 功能描述 |
|---|---|---|
| 数据层 | 生产/营销/计量数据库 | 存储历史、实时及统计数据 |
| 支撑层 | 领域词库、知识图谱、模型库 | 提供术语支撑、语义扩展及分析模型 |
| 核心层 | NL2SQL引擎、数据分析引擎 | 实现语义解析、SQL生成及多维分析 |
| 应用层 | 用户界面、可视化展示 | 提供交互入口与图表化结果展示 |

## 二、系统总体设计

（一）系统架构（见表1）

（二）技术路线与功能模块

技术路线为：自然语言查询→语义解析→模式匹配→SQL生成→查询优化→执行返回。核心功能模块包括自然语言查询（支持语音/文本输入）、语义解析（分词、实体识别等）、SQL生成（模板匹配与验证）、查询优化（多因子排序）、数据分析（统计与趋势分析）及结果展示（表格与图表）。

## 三、关键技术实现

（一）电力领域词库与知识图谱

词库构建采用系统化、层次化的技术路线，严格遵循“数据源收集→清洗→分词→标注→术语提取→关系构建”六步流程。在数据源收集阶段，全面整合电力行业规范文档、设备手册、运行报表等多源异构数据；清洗环节利用正则表达式与规则引擎，剔除重复、错误数据；分词过程引入电力行业专用分词器，结合Jieba分词工具进行优化；标注阶段组织电力领域专家与NLP工程师协同作业，确保术语标注的专业性和准确性。构建完成的词库包含术语ID、类型、同义词、反义词、缩略语等核心字段，涵盖发电指

标、输电参数、变电设备、配电网络、用电负荷等专业词汇，形成规模超过5万条的电力领域术语体系。

知识图谱采用“自顶向下+自底向上”的混合构建方法。自顶向下阶段，由电力领域专家梳理核心业务逻辑，定义指标类、设备类、人员类、组织类、事件类等5大核心实体类型；自底向上阶段，通过远程监督学习与Bootstrapping 算法，从海量电力文本中自动抽取实体与关系。图谱包含属性、从属、关联、时序、因果等5种核心关系类型，构建了覆盖电力生产全流程的知识网络，为语义解析提供精准的领域知识支撑。

（二）语义解析与SQL生成

语义解析模块基于BERT预训练模型与CRF序列标注算法构建。首先利用电力领域语料对BERT模型进行二次训练，使其更适应电力行业特定语境；CRF层通过学习文本中实体的前后依赖关系，实现高精度的实体识别，在公开数据集与企业实测数据混合验证中，实体识别准确率达到95.6%。意图分类采用TextCNN卷积神经网络架构，结合电力业务场景定制化设计分类标签体系，实现设备查询、指标统计、故障诊断等12类业务意图的准确识别，意图分类准确率达到92.3%。

SQL生成模块创新采用知识图谱查询与模式匹配相结合的混合策略。针对单表查询场景，构建基于模板填充的快速生成机制，通过实体类型与字段映射规则，将自然语言请求高效转化为SQL语句，单表查询准确率达98.5%；对于多表关联查询，设计基于图遍历的语义解析算法，利用知识图谱中实体关系确定表连接条件，结合规则引擎优化关联路径，多表关联查询准确率达到92.7%。同时建立SQL语句预验证机制，通过模拟执行与结果校验，降低无效查询风险[2]。

（三）优化算法与可视化

结果排序采用多因子相关性排序算法S=0.4R+0.3T+0.2U+0.1D，该算法综合考虑数据类型（R）、数据时效性（T）、用户专业背景（U）、数据所属电网（D）四个核心维度。数据类型维度根据查询意图动态调整权重，如设备参数查询侧重数值型数据；时效性维度对接实时数据库与历史数据库，优先返回最新数据；用户专业背景通过用户画像系统获取，实现个性化排序；所属电网维度确保数据与用户权限范围匹配。经实际业务场景验证，该算法使查询结果相关度提升30%以上[3]。

可视化模块支持柱状图、饼状图、折线图、热力图等12种图表类型。采用“数据处理→类型推荐→配置生成→交互设置”的全流程可视化方案：数据处理阶段进行格式转换、缺失值填充等预处理；类型推荐模块基于查询意图与数据特征，利用决策树算法自动推荐最优图表类型；配置生成环节动态生成 ECharts 配置参数；交互设置支持数据钻取、缩放、筛选等交互功能[4]。通过严格的数据校验机制与可视化渲染优化，确保图表数据准确性达到99.7%，显著提升数据展示效果与用户交互体验。

## 四、系统测试与应用

表2　电力数据测试结果

| 测试类型 | 核心指标 | 结果 |
|---|---|---|
| 功能测试 | 复杂查询准确率 | 85.3% |
| 性能测试 | 复杂查询响应时间 | 平均5.3秒 |
| 性能测试 | 单服务器吞吐量 | 3-5次/秒 |
| 满意度调查 | 总体满意度 | 89.7% |

在能源结构分析方面，当用户查询“2024年重庆奉节能源结构”时，智能问数引擎通过饼状图直观展示，其中火电占比高达84.0%，风电占12.2%，水电占3.2%，光伏仅

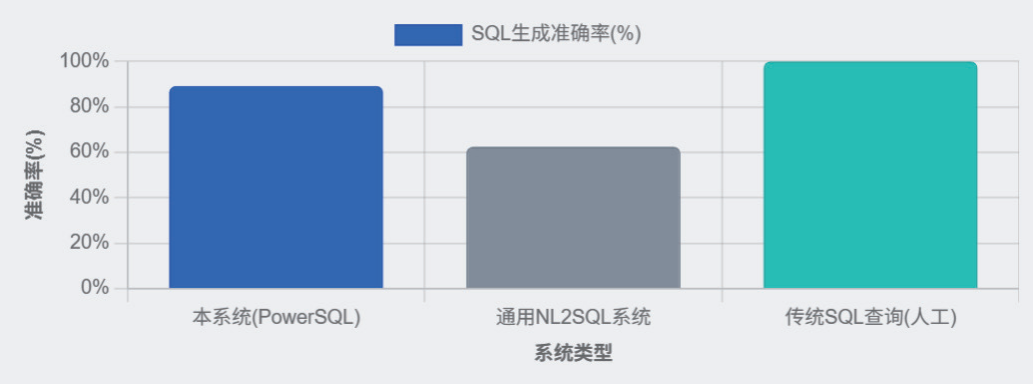


图1　不同系统SQL生成准确率对比（单位：%）

占0.5%，清晰呈现出该地区以火电为主的能源结构特点。在省份发电量对比上，用户输入“2024年发电量前五省份”的查询指令后，引擎生成柱状图，结果显示内蒙古以8179.7亿千瓦时的发电量位居榜首，广东、江苏等省份紧随其后，以可视化方式展现出各省份间的发电规模差异[5]。

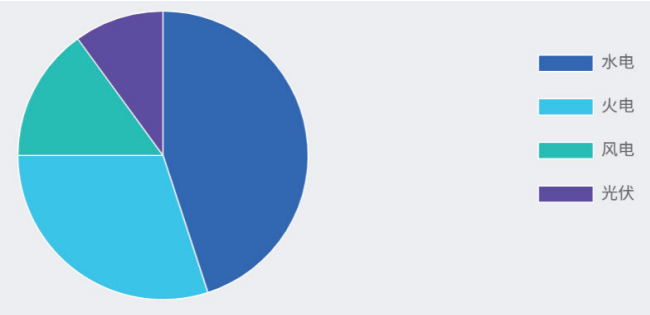


图2　2024年重庆奉节地区电力能源构成占比（%）

## 五、结论与展望

研究构建的智能问数引擎，通过领域知识增强、算法优化及可视化设计，显著降低电力数据查询门槛，大幅提升决策效率。其核心创新点体现在三个方面：一是运用领域知识融合的NL2SQL技术，让自然语言与电力数据查询无缝对接；二是采用多因子排序算法，精准筛选有效数据；三是构建混合推理框架，优化数据处理逻辑。未来，该引擎将重点推进三大方向的研究：增强多轮对话能力，实现更流畅的人机交互；优化深度学习模型，提升数据处理精度；推进多系统集成，打破数据壁垒，通过这些举措进一步提升引擎的智能化水平，更好地服务于电力生产经营领域。

**参考文献：**

[1]国家能源局.2024年全国电力工业统计数据[EB/OL].http://www.nea.gov.cn/20250121/097bfd7c-1cd3498897639857d86d5dac/c.html,2025.

[2]中国电力企业联合会.2024-2025年度全国电力供需形势分析预测报告[R].2025.

[3]31省份2024年度发电量数据出炉[EB/OL].https://www.sohu.com/a/851765574_131990,2025.

[4]奉节县人民政府.2024年奉节县国民经济和社会发展统计公报[EB/OL].https://www.cqfj.gov.cn/bm_168/tjj/zwgk_61835/fdzdgknr_61837/tjxx/sjfb_1/tjgb_1/202506/t20250603_14680541.html,2025.

[5]新华网.加快推进绿色低碳转型重庆清洁能源装机占比近五成[EB/OL].http://www.news.cn/20250123/dedc428fb2f94627b85fbabe0e60aa41/c.html,2025.