已完成工作：

-文本按子句分句

-每个子句已进行分词和依存句法分析，得到类似 parsed

-每个子句中识别出了实体和候选实体对，包含实体类型与预定义关系（未完成）

需完成：

-在每个子句内，对每一对候选实体对，查找它们在该子句的依存路径

* 提取典型依存路径

定义：依存路径是指，在依存句法树中，连接两个实体（词元）之间的路径关系（包含路径上的依存关系和词性）

研究方法 ——[nsubj]→ 提出 ——[dobj]→ 模型

可形成三元组抽取模式：(研究方法, 提出, 模型)

依存路径提取路线：

1. 确定实体对

从句子中找出所有可能的实体对组合（实体1，实体2）

1. 依存图中找“最短路径”

·在 HanLP 的依存分析结果中构建依存图（树）

·使用 BFS / DFS 寻找 实体1 到 实体2 的最短依存路径

·保留路径中每个节点的：词、词性、依存关系、方向

1. 路径格式化

·实体1 -[nsubj]→ 动词（谓词） -[dobj]→ 实体2

或[

{"word": "研究方法", "dep": "nsubj", "upos": "NN"},

{"word": "提出", "dep": "root", "upos": "VV"},

{"word": "模型", "dep": "dobj", "upos": "NN"}

]

1. 统计并归纳典型路径模式

·统计最频繁出现的依存路径结构

·提取结构上类似的抽象模式

如：方法 –[nsubj]→ 动词 –[dobj]→ 对象

-----------------------------------------------------实验环节----------------------------------------------------------

以《大数据下数模联动的随机退化设备剩余寿命预测技术》为例

* 实体识别：

| **实体** | **实体类型** |
| --- | --- |
| 随机退化设备剩余寿命预测 | Problem |
| 机理模型与数据混合驱动的预测技术 | Method |
| 基于机器学习的预测技术 | Method |
| 统计数据驱动的预测技术 | Method |
| 机器学习和统计数据驱动相结合的预测技术 | Method |
| 多源传感监测大数据 | Sensor / Dataset |
| 航空发动机多源监测数据 | Dataset / System |
| 数模联动预测思路 | Model |
| 深度学习与退化建模交互联动 | Model |
| 不确定性量化问题 | Problem |
| 剩余寿命预测结果 | Finding |
| 非理想大数据 | Dataset |
| 性能指标 | Metric |

* 实体对筛选：

| **主语实体** | **主语类型** | **宾语实体** | **宾语类型** | **关系** |
| --- | --- | --- | --- | --- |
| 基于机器学习的预测技术 | Method | 随机退化设备剩余寿命预测 | Problem | 解决 (addresses) |
| 统计数据驱动的预测技术 | Method | 不确定性量化问题 | Problem | 解决 (addresses) |
| 深度学习与退化建模交互联动 | Model | 航空发动机 | System | 应用于 (is\_applied\_to) | |
| 深度学习与退化建模交互联动 | Model | 不确定性量化问题 | Problem | 预测/诊断 (predicts) | |
| 数模联动预测思路 | Model | 剩余寿命预测结果 | Finding | 得出 (achieves) | |
| 剩余寿命预测结果 | Finding | 研究展望 | Future Work | 建议 (suggests) |
| 性能指标 | Metric | 数模联动预测思路 | Model | 评估 (is\_evaluated\_by)\*（反转） |
| 特征 | Feature | 多源传感监测大数据 | Sensor | 来自 (is\_from) |

* 核心函数

| **函数名** | **作用** |
| --- | --- |
| find\_entity\_token\_span | 将实体文本对齐到分词序列，得到 token ID 范围 |
| build\_dependency\_graph | 构建依存图（有向图 + id 映射） |
| find\_shortest\_dependency\_path | 在依存图上查找实体对之间的最短路径 |
| process\_entity\_pair\_in\_sentence | 以上函数的整合：输入实体对 + 分词，返回依存路径 |

--------------------------------------------------------目前存在问题-----------------------------------------------------

1. 每个子句的实体识别与实体对的确定（当前是LLM+人工）

格式如：

 {

    "subject": "基于机器学习的预测技术",

    "subject\_type": "Method",

    "relation": "addresses",

    "object": "随机退化设备剩余寿命预测",

    "object\_type": "Problem"

  },

1. 实体的id对齐，函数效果不算特别好，有些id无法进行对齐
2. 依存路径结果如何理解，是否真的有用，需要可视化嘛？

提高篇数，使用LLM提取典型规则，查看其思考方式

评价指标

------------------------------------------------------结果展示--------------------------------------------------------------

对齐失败: subject=特征, object=多源传感监测大数据, tokens=['面向', '大', '数据', '背景', '下', '随机', '退化', '设备', '剩余', '寿命', '(', 'Remaining', 'useful', 'life', ',', 'RUL', ')', '预测', '的', '现实', '需求', ',', '结合', '随机', '退化', '设备', '监测', '大', '数据', '特点', '及', '剩余', '寿命', '预测', '不确定性', '量化', '这', '一', '核心', '问题', ',', '深入', '分析', '了', '机理', '模型', '与', '数据', '混合', '驱 动', '的', '剩余', '寿命', '预测', '技术', '、', '基于', '机器', '学习', '的', '剩余', '寿命', '预测', '技术', '、', '统计', '数据', '驱动', '的', '剩余', '寿命', '预测', '技术', '以及', '机器', '学习', '和', '统计', '数据', '驱动', '相', '结合', '的', '剩余', '寿命', '预测', '技术', '的', '基本', '研究', '思想', '和', '发展', '动态', ',', '剖析', '了', '当前', '研究', '存在', '的', '局限性', '和', '共性', '难题', '.', '针对', '存在', '的', '局限性', '和', '共性', '难题', ',', '以', '多源', '传感', '监测', '大', '数据', '下', '剩余', '寿命', '预测', '问题', '为', '例', ',', '提出', '了', '一', '种', '数模', '联动', '的', '大', '数据', '下', '随机', '退化', '设备', '剩余', '寿命', '预测', '解决', '思路', ',', ' 并', '通过', '航空', '发动机', '多源', '监测', '数据', '初步', '验证', '了', '该', '思路', '的', '可行性', '和', '有效性', '.', '最后', ',', '借鉴', '数模', '联动', '思路', ',', '综合', '考虑', '机器', '学习', '方法', '和', '统计', '数据', '驱动', '方法', '的', '优势', ',', '紧紧', '扭住', '大', '数据', '背景', '下', '随机', '退化', '设备', '剩余', '寿命', '预测', '不确定性', '量化', '问题', ',', '提出', '了', '大', '数据', '背 景', '下', '深度', '学习', '与', '随机', '退化', '建模', '交互', '联动', '、', '监测', '大', '数据', '与', '剩余', '寿命', '及', '其', '预测', '不确定性', '映射', '机制', '、', '非', '理想', '大', '数据', '下', '的', '剩余', '寿命', '预测', '等', '亟待', '解决', '的', '关键', '科学', '问题', '.']

[

{

"subject": "基于机器学习的预测技术",

"object": "随机退化设备剩余寿命预测",

"relation": "addresses",

"path": [

[

"预测",

"assmod"

],

[

"需求",

"dobj"

],

[

"面向",

"root"

]

]

},

{

"subject": "统计数据驱动的预测技术",

"object": "不确定性量化问题",

"relation": "addresses",

"path": [

[

"数据",

"nn"

],

[

"背景",

"lobj"

],

[

"下",

"dep"

],

[

"预测",

"assmod"

],

[

"需求",

"dobj"

],

[

"面向",

"root"

]

]

},

{

"subject": "深度学习与退化建模交互联动",

"object": "航空发动机",

"relation": "is\_applied\_to",

"path": [

[

"退化",

"amod"

],

[

"设备",

"nn"

],

[

"寿命",

"nn"

],

[

"预测",

"assmod"

],

[

"需求",

"dobj"

],

[

"面向",

"root"

]

]

},

{

"subject": "深度学习与退化建模交互联动",

"object": "不确定性量化问题",

"relation": "predicts",

"path": [

[

"退化",

"amod"

],

[

"设备",

"nn"

],

[

"寿命",

"nn"

],

[

"预测",

"assmod"

],

[

"需求",

"dobj"

],

[

"面向",

"root"

]

]

},

{

"subject": "数模联动预测思路",

"object": "剩余寿命预测结果",

"relation": "achieves",

"path": [

[

"预测",

"assmod"

],

[

"寿命",

"nn"

],

[

"剩余",

"amod"

]

]

},

{

"subject": "剩余寿命预测结果",

"object": "研究展望",

"relation": "suggests",

"path": [

[

"剩余",

"amod"

],

[

"寿命",

"nn"

],

[

"预测",

"assmod"

],

[

"需求",

"dobj"

],

[

"面向",

"root"

],

[

"分析",

"conj"

],

[

"动态",

"dobj"

],

[

"研究",

"nn"

]

]

},

{

"subject": "特征",

"object": "多源传感监测大数据",

"relation": "is\_from",

"path": null,

"note": "实体对无法与token对齐"

}

]