

Annotation Guideline

标注指南

目录 CONTENTS



□ 项目介绍

□ 实体类别

□ 关系类别

□ 领域词典

□ 标注示例



PART ONE

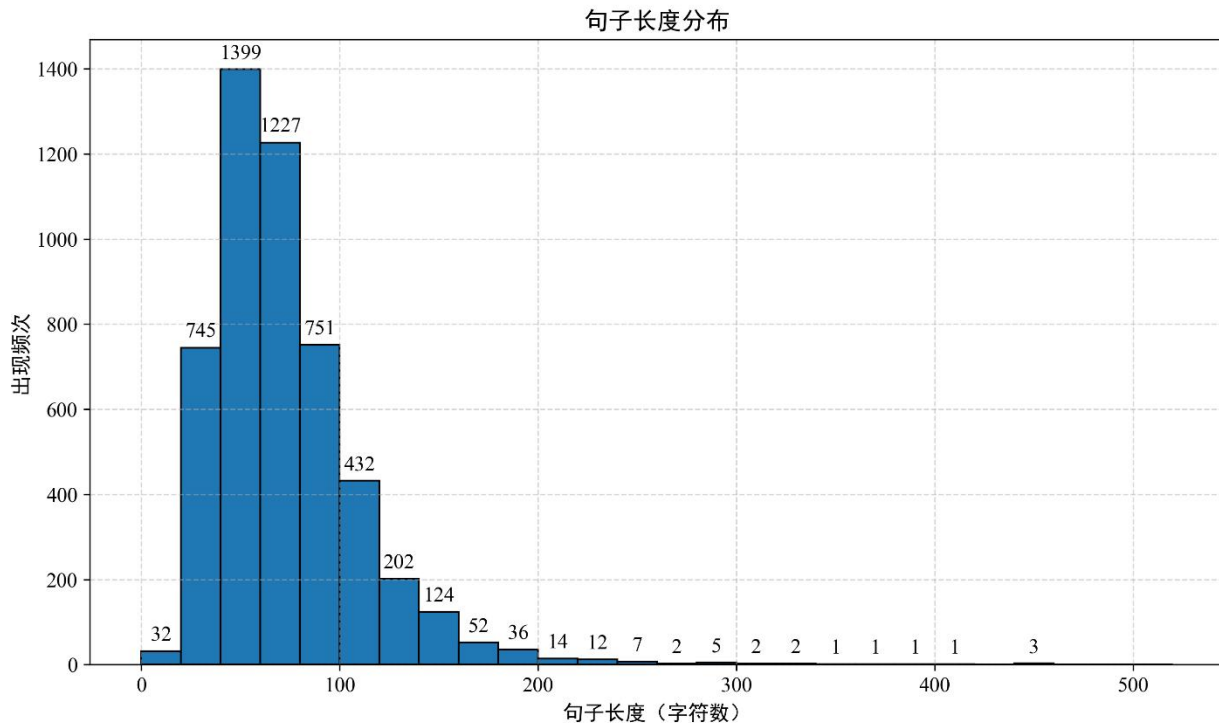
项目介绍

1.1 项目基本信息

- **项目名称：**民机装备供应链关系抽取项目。
- **项目目标：**建立民机供应链中文语料库标准数据集，以支撑下游关系抽取和民机知识图谱构建任务。
- **数据集大小：**5051个新闻句子。
- **项目意义：**以民机为代表的复杂装备具备供应链链条长、可见度低、系统性风险复杂等特征，弄清民机供应链网络对于民机供应链风险管理、数字化管理实现、可持续性和绩效提升有重要的现实意义。

1.1 项目基本信息

- **数据特征：**大部分句子是短句，字符数在100以内，便于人工和大模型标注。



1.2 项目方案设计

- 本项目结合人工和大模型进行标注。
- 人工标注方案（构建人工标准数据集）：3人进行标注，1人（项目负责人）进行仲裁，当3个标注者的标注者内一致性和标注者间一致性通过检验时有效。
- 大模型标注方案（构建LLM标准数据集）：用人工标准数据集微调Qwen 3 LLM，然后将多个Qwen 3 LLM构成集成器，保留高置信度样本扩充标准数据集。
- 根据句子长度衡量标注难度，选择2000个样本划分给人工进行标注，剩下的作为大模型标注的候选数据。

1.3 实体类别

1. 组织 (Organization)
2. 产品型号 (Product Model)
3. 产品/服务 (Product/Service)

1.4 关系类别

1. 供应 (Supply)
2. 生产产品/提供服务 (Produce_product/Provide_service)
3. 合作/拟合作 (Cooperate/Propose_to_cooperate)
4. 股东/母公司/并购/投资 (Hold_stake_in/Own/Acquire/Invest_in)
5. 共指 (Coreference)
6. 其他/无 (Others/None)

1.5 项目注意事项

- 1. 基于句内语义标注关系：**标注时仅依据本句语义判断两个实体是否存在供应链关系，不得参考上下句或外部知识。
- 2. 允许合理推理：**若句中语义明确表达行为（如“采购”、“交付”、“提供”、“签署订单”等），可结合句内语义判断关系。模糊性表达（如“出席”、“见证”、“探讨”）则不作关系标注。
- 3. 实体标注范围：**每个句子中出现的符合要求的实体全部标注。实体不能重叠或嵌套，每个字符仅属于一个实体。例如“空客a320”应拆为“空客”与“a320”；
- 4. 实体识别无误：**标注者默认为领域内专家，能准确识别组织、产品型号和产品/服务3类实体；
- 5. “其他/无”类比例限制：**对于确无供应链关系或不属于前面5类供应链关系的句子，关系类型标为“其他/无”。每位标注者“其他/无”类不超过全部标注数据的5%，超出将打回重标；
- 6. 标注句子范围说明：**所有句子已通过双轮人工筛选，大多数都含有供应链关系。企业主体多数为民机供应链企业，也可能包含其他行业的企业，均标注为组织实体。



PART TWO

实体类别

实体类别

2.1 组织 (Organization)

指民机供应链链条上的企业，既包括制造商、核心系统和设备供应商、材料和标准件供应商以及下游的航空运营单位，也包括参与技术研发的高校和科研机构以及提供适航认证等服务的政府单位，标注时对民机供应链外的企业进行保留，表现出包容性。

实体类别

2.2 产品型号 (Product Model)

指具体的产品名称，包括民机整机名称，如a320、b737MAX、c919等，也包括供应链企业生产的军机产品。此外，语料库中混入的其他产品，如和民机有密切联系的汽车领域的产品型号也被纳入该实体。该实体可通过追溯其制造商来实现新的关系三元组的推理，从而提高“组织-组织”关系的召回率。

实体类别

2.3 产品/服务 (Product/Service)

指具体的民机子系统或结构件及所有相关产品，如航电系统、机体、机头、飞机座椅、高温合金、原材料等；也包括维修、适航认证等领域相关的服务。



PART THREE

关系类别

关系类别

3.1 供应 (Supply)

“供应”关系指民机供应链组织与组织之间、组织与产品（包括产品型号和产品/服务两类实体）之间的业务流动关系。

关系类别

3.2 生产产品/提供服务 (Produce_product/Provide_service)

“生产产品/提供服务” 关系指供应链链条上组织对另一组织提供的产品或服务，可以是具体的和模糊的。

关系类别

3.3 合作/拟合作 (Cooperate/Propose_to_cooperate)

“合作/拟合作” 关系指组织与组织之间已建立或将要建立的合作关系，且无法明确供应链方向的情况均归入此类。

关系类别

3.4 股东/构成/并购/投资 (Hold_stake_in/Own/Acquire/Invest_in)

“股东（是……股东）/母公司（是……母公司）/并购/投资”关系是指两个组织之间，前者对后者存在资本或控制关系。具体来讲，在“A是B的股东”、“A是B的母公司”、“A并购了B的业务”或“A对B进行了投资”等表述中，A都对B有一定的资本控制权或决策影响力。

关系类别

3.5 共指 (Coreference)

“共指”关系用于识别文本中同一实体的不同表达方式，以实现实体消歧和连接。

关系类别

3.6 其他/无 (Others/None)

“其他/无”关系包含无法用前述5类供应链关系归纳的其他关系以及文本中不存在供应链关系这两种情形。但是在标注这一关系时要控制在总标注数量的5%以内。



PART FOUR

领域词典

4.1 组织 (Organization)

- **民机制造商：**中国商用飞机有限责任公司（或中国商飞或comac）、波音（boeing）、空中客车（或空客或airbus）、俄罗斯联合航空制造集团、庞巴迪等；
- **发动机制造商：**中国航发等；
- **民机运营商：**中国国际航空公司（或国航）、海航集团有限公司（或海航）、中国南方航空公司（或南航）、东航、泰国都市航空等；
- **民机租赁单位：**中银航空租赁、浦银租赁等租赁公司；
- **政府组织：**中国民航局（或民航局或caac）、欧洲航空安全局（easa）、美国联邦航空局（faa）等。

4.2 产品型号 (Product Model)

- 民机整机 (国内) : 中国商飞c系列, c909 (或909或arj21) 、 c919 (或919) 、 c929 (或929) 、 c939 (或939) 、 ma60 (或新舟60、 y-7-200a, 中航工业旗下公司研制) 、 ma600等;
- 民机整机 (国外) : a320、 a220、 b737、 b787 、 c系列 (庞巴迪公司生产) 、 erj190系列 (巴西航空工业产) , 同样可省略字母;
- 军机: 美国军机f22、 歼-11、 歼-15、 歼-16, 军机上下文往往有军字, 没有进行检索;
- 发动机: leap、 cj-1000 (系列发动机) 或长江-1000 (系列发动机) 等;
- 汽车: 如奔驰大g越野车这类具体的标注为型号, 如果句中只是单个的越野车则标为产品;

.....

4.3 产品/服务 (Product/Service)

- **飞机核心系统：**动力系统（发动机系统、辅助动力装置[apu]）、飞行控制系统、航电系统（导航系统、通信系统、监视系统、飞行仪表显示系统、任务系统）、机电系统、电源系统、液压系统、燃油系统、气动系统、环境控制系统、起落架系统、防冰除冰系统、机载信息系统、火警与监测系统；
- **飞机结构系统：**机体、机头、前中后机身、机翼、尾翼、起落架、舱门等；
- **飞机子系统或模块：**传感器、碳刹车盘等、液压泵、电源管理单元等；
- **飞机材料和标准件：**铝合金、钛合金、复合材料、标准紧固件、铆钉、电线线缆等；

.....

4.3 产品/服务 (Product/Service)

- **适航认证**，适航认证、适航取证等
- **XX解决方案**，如电传飞控解决方案、发动机解决方案等；
- **维修相关**：维护、维修、大修（或mro）等；

.....



PART FIVE

标注示例

标注示例

5.1 供应 (Supply)

在中国以外,唯一运营**中国商飞**飞机的航空公司是印度尼西亚的廉价航空公司 transnusa,该公司运营的是**arj21**支线飞机。

三元组

中国商飞	供应	transnusa
arj21	供应	transnusa
中国商飞	生产产品/提供服务	arj21

标注示例

5.2 生产产品/提供服务 (Produce_product/Provide_service)

赛峰电子与防务公司为其提供主要的系统和设备,如**驾驶舱控制系统**、**控制面板与调光控制**、**油门台组件**和**水平安定面配平作动器**。

三元组

赛峰电子与防务公司	生产产品/提供服务	驾驶舱控制系统
赛峰电子与防务公司	生产产品/提供服务	控制面板与调光控制
赛峰电子与防务公司	生产产品/提供服务	油门台组件
赛峰电子与防务公司	生产产品/提供服务	水平安定面配平作动器

标注示例

5.2 生产产品/提供服务 (Produce_product/Provide_service)

他们正强化产品性能和生产流程要求,努力获得**美国联邦航空局(faa)**、**欧洲航空安全局(easa)**适航认证。

三元组

美国联邦航空局	生产产品/提供服务	适航认证
美国联邦航空局	生产产品/提供服务	适航认证
美国联邦航空局	共指	faa
欧洲航空安全局	共指	easa

标注示例

5.3 合作/拟合作 (Cooperate/Propose_to_cooperate)

据**道达尔航空公司**控股合伙人保罗阿尔马达透露,该公司与**中国商飞**已进行了数月的谈判,并将于10月赴华商讨最多四架**c919**飞机的潜在订单。

三元组

道达尔航空公司	合作/拟合作	中国商飞
中国商飞	生产产品/提供服务	c919

标注示例

5.4 股东/母公司/并购/投资 (Hold_stake_in/Own/Acquire/Invest_in)

在收购前5个月,蚂蚁金服宣布完成B轮融资,中国投资有限责任公司和中国建设银行的子公司建信信托领投。

三元组

建信信托	股东/母公司/并购/投资	蚂蚁金服
中国投资有限责任公司	股东/母公司/并购/投资	蚂蚁金服
中国建设银行	股东/母公司/并购/投资	建信信托

标注示例

5.5 共指 (Coreference)

中国航展开始之前,中俄国际商用飞机有限责任公司(craic)(一家中俄合作企业)正在进行其主要供应商的甄选工作。

三元组

中俄国际商用飞机 有限责任公司	共指	craic
--------------------	----	-------

标注示例

5.6 其他/无 (Others/None)

这是因为,c919还要走出海外,要与**空客**、**波音**竞争,采用全球化供应链以分享利润,才有利于开拓市场。

三元组

c919	其他/无	空客
c919	其他/无	波音

标注示例

5.7 选择合适的标注工具

- 本项目使用Doccano进行实体和关系的联合标注

The screenshot displays the Doccano web interface for text annotation. The main window shows a sentence: "昂际航电作为中国商飞的一级供应商,为c919飞机提供核心航电系统。" (Angjiahangdian as a first-tier supplier of COMAC, providing core avionics for the C919 aircraft). The interface includes a left sidebar with navigation options: Home, Dataset, Labels, Relations, Members, Comments, Guideline, Metrics, and Settings. A top toolbar contains icons for various annotation actions. The sentence is annotated with the following labels:

- 组织** (Organization): 昂际航电 (Angjiahangdian)
- 组织** (Organization): 中国商飞 (COMAC)
- 产品/服务** (Product/Service): c919
- 产品/服务** (Product/Service): 核心航电系统 (Core Avionics System)
- 关系** (Relation): 供应 (Supply) - connecting 昂际航电 and 中国商飞
- 关系** (Relation): 生产产品/提供服务 (Produce product/Provide service) - connecting 昂际航电 and 核心航电系统

On the right side, there is a **Progress** section showing 907 of 5051 items, with a 1% completion bar. Below it is the **Label Types** section, which lists various relation types with corresponding colored buttons: 供应 (a), 生产产品/提供服务 (b), 股东/母公司/并购/投资 (d), 共指 (e), 其他/无 (f), and 合作/拟合作 (c). The 'Relation' toggle is currently turned on.