

# 中文语义依存分析评测大纲

## 一、评测对象

本次评测的对象是中文句子的词语之间的语义及句法依存关系分析技术。

## 二、评测内容

本次评测主要评测依存关系分析系统对中文句子进行依存分析的准确性。

## 三、评测方法

### 1. 评测方式

本次评测为离线评测，参赛单位自行处理数据，生成相应结果后提交。训练语料将提前 1 个月公布。

### 2. 评测步骤

- 1) 评测单位预先给出若干训练数据（已标注依存分析结果）；
- 2) 评测单位给出测试数据（未标注依存分析结果）；
- 3) 参赛单位运行被测系统，得出测试结果；
- 4) 参赛单位提交测试结果；
- 5) 评测单位标注答案，运行自动评测程序，统计评测结果。

### 3. 评测标准

#### 1) 量化指标

本次评测采用三个指标对被测试系统进行评测，分别为：

- i. 依存标注准确率（Labeled Attachment Score, LAS）
- ii. 依存准确率（Unlabeled Attachment Score, UAS）
- iii. 标注准确率（Labeled Accuracy, LA）

令整个测试预料包含的词数为 $N$ ，任意词语的依存用三元组 $\langle \text{word}, \text{headword}, \text{relation} \rangle$ 进行表示，其中  $\text{word}$  为词本身， $\text{word}$  以关系  $\text{relation}$  依存于  $\text{headword}$ 。在被测试系统的输出中，令所有  $\text{headword}$  正确的词语的数目为 $N_a$ ，所有  $\text{relation}$  正确的词语的数目为 $N_l$ ，所有  $\text{headword}$  和  $\text{relation}$  都正确的词为 $N_{al}$ ，那么，测试指标的计算方法如下：

$$LAS = \frac{N_{al}}{N}$$

$$UAS = \frac{N_a}{N}$$

$$LA = \frac{N_l}{N}$$

三个指标的优先级分别为 $LAS > UAS > LA$ 。

### 4. 数据格式

测试数据采用 CONLL<sup>1</sup>格式，数据文件采用 UTF-8 编码。测试数据总共提供 6 列值，分别为：词语 ID、词语、词性（粗粒度）、词性（细粒度）、被依存词 ID 和依存关系，即 CONLL 格式中的第 1、2、4、5、7 和 8 列。其他列留空（用“\_”填充）。列与列之间用“\t”分隔，行与行之间用“\n”分隔，句子与句子之间用空行分隔。被测试系统可以自行对其他列进行补充，**但不允许修改分词结果**（清华大学 SDN 语料的词性为自动标注结果，参赛者可对词性进行重新标注）。

输出数据同样采用 CONLL 格式，其中第 1、2 列应当与测试数据严格一致，分析结果填充到第 7、8 列。其他列可留空（用“\_”填充），也可以保留自行补充的数据。

---

<sup>1</sup> <http://ilk.uvt.nl/conll/#dataformat>

测试数据文件命名为 test.conll，输出数据文件统一命名为 teamID\_X\_THU.conll 或 teamID\_X\_HIT.conll，其中 teamID 为参赛队伍 ID，X 为数字，代表当前参赛队伍的第 X 组结果。THU 和 HIT 分别表示在清华大学语料库和哈工大语料库上的测试结果（详见第四节：评测数据集）。

数据样例如下：

句子 1（来源：清华大学 SDN 语料）

1	你们	_	r	rr	_	4	限定
2	的	_	u	ude1	_	1	“的”字依存
3	拼搏	_	v	v	_	4	限定
4	精神	_	n	n	_	9	关系主体
5	和	_	c	cc	_	4	连接依存
6	无私	_	v	vl	_	7	方式
7	奉献	_	v	vl	_	8	限定
8	精神	_	n	n	_	4	连接依存
9	值得	_	v	v	_	0	核心成分
10	大家	_	r	rr	_	11	施事
11	学习	_	v	v	_	9	内容

句子 2（来源：哈尔滨工业大学语料库）

1	钱其琛	_	NR	NR	_	2	施事
2	谈	_	VV	VV	_	0	根
3	香港	_	NR	NR	_	4	领有者
4	前景	_	NN	NN	_	7	并列
5	和	_	CC	CC	_	7	助词-方位词-连词
6	台湾	_	NR	NR	_	7	领有者
7	问题	_	NN	NN	_	2	内容

#### 四、评测数据集

本次评测分别在两个数据集上进行，其一为清华大学语义依存网络（Semantic Dependency Network, SDN），其二为哈尔滨工业大学语义依存语料库。这两个语料库都遵循 Robinson 提出的依存关系的四大公理<sup>[1]</sup>：

- i. 一个句子中只有一个词不依存于任意词，该词称为核心词；
- ii. 除核心词外的任意词都直接依存于其本身之外的某一词；
- iii. 任何一个词都不能依存于两个和两个以上的词；
- iv. 如果词 A 直接依存于词 B，而词 C 在句子中位于 A 和 B 之间，那么，C 或者直接依存于 A，或者直接依存于 B，或者直接依存于 A 和 B 之间的某一词。

1. 清华大学语义依存网络（Semantic Dependency Network, SDN）

1) 依存关系集合

SDN 的依存关系集合总共包含 70 个依存关系，其中语义依存关系 59 个，特殊句法依存关系 9 个，特殊依存关系 2 个（见附表 1）。

2) 数据规模

训练集 20,000 句，开发集 2,000 句，测试集 3,000 句。

2. 哈尔滨工业大学语义依存语料

1) 依存关系集合

语义关系共分为四大类，主语义关系 20 种，辅助语义关系 35 种，定语语义关系 78 种（语料中出现 56 种），句法语义关系 16 种（见附表 2）。

2) 数据规模

训练集 8,301 句，开发集 534 句，测试集 1,233 句

## 五、评测要求

参评单位应当采用自动的方法，针对中文句子进行依存分析。参评系统应当预先训练模型、调整好所有参数，运行过程中不得有人工干预。本次评测不限制使用各种自然语言处理工具及语义资源。**特别地，参评系统可以融合两个数据集的训练和开发数据进行模型的训练和调整，然后分别在两个测试集上进行测试。每个参评单位在每个语料库上最多提供三组结果。**

附表 1：清华大学语料依存关系集合

主语义关系 (16)	施事 领有者 受事 结果	关系主体 内容 整体	经验者 类指 部分	描写体 占有物 代价	存现体 目标 触及部件
辅助语义关系 (24)	时间 后延时段 处所 终状态 相伴体 根据 比较内容	进程时间 原处所 参照体 来源 比较量	起始时间 通过处所 目的 工具	终止时间 终处所 原因 手段	时距 原状态 方向 材料
定、状语语义关系 (10)	限定 评论 频率	数量 方式	描述 程度	同位语 范围	动量
连动词及从句语义关系 (9)	结果事件 条件	接续 并列	伴随 递进	事件过程 让步	除了
特殊句法结构(9)	“的”字依存 语气依存	...是...的依存 关联词依存	连接依存 趋向动词依存	方位词依存 介词依存	时态语态依
特殊关系 (2)	核心成分	依存失败			

附表 2：哈工大语料依存关系集合

主语义关系 (20)	施事，经验者，致事，领有者，存现体，整体，关系主体，类指，内容，占有物，受事，部分，损益者，参照体，相伴体，依据，原因，代价，范围，关于
辅助语义关系 (35)	时段，终止时间，起始时间，时间点，时间状语终处所，原处所，通过处所，终状态，原状态，方向，距离，状态，处所，伴随，接续，泛指频率，工具，材料，手段，角度，动量，顺序数，顺序量，否定，程度，情态，强调，方式，体，插入语，助词-方位词-连词，介词，标点符号，根
定语语义关系 (78)	类别，成员，事域，名量，数字，指量，指示，宿主，时间短语修饰语，地点短语修饰语，机构短语修饰语，属性，限定，材料，内容，顺序数，顺序量，未知， 反施事，反受事，反领属者，…，反关于，（与主语义关系对应） 隐施事，隐受事，隐领属者，…，隐关于，（同上） 间接施事，间接受事，间接领属者，…，间接关于（同上）
句法语义关系 (16)	原因，让步，假设，并列，选择，递进，除外，顺承，目的，措施，割舍，选取，总括，分述，关于，结果

[1] Robinson, Jane J.. "Dependency structures and transformation rules." Language 46 (1970): 259-285.