

Lecture 6: Dependency Parsing

2018年7月8日 10:31

Some Terminologies

participial phrase	分词短语
prepositional phrase	介词短语
Adverbial	副词
Infinitive	不定式
Syntactic structure	句法结构
Lexical item	词法项
Conjunction	连词

Two views of linguistic structure

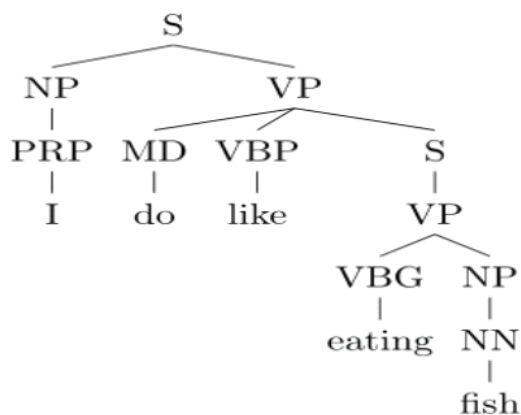
1. Context-free grammar
2. Dependency grammar

Context-free grammar CFG (1947)

- Attachments Ambiguity 挂靠歧义
 - Scientists [study whales] [from space]
 - 科学家在太空中（通过卫星）研究鲸鱼。
 - Scientists study [whales from space]
 - 科学家研究宇宙中的庞然大物。

Annotated Data: Treebanks

定义一套规则，如介词短语，不定式，动宾短语，进行机械地分解

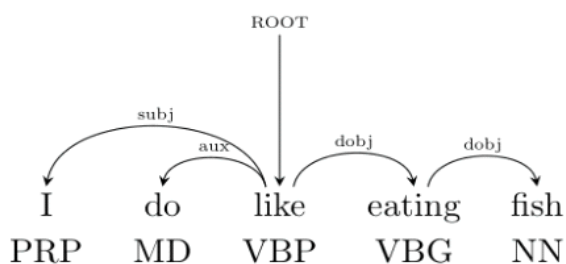


<http://blog.csdn.net/u014422406>

Dependency Grammar (1959)

In early 20 years, dependency grammar gains more and more popularity

- Single head
- Connected
- Asymmetry



<http://blog.csdn.net/u014422406>

中文与英文环境下的 Parsing 有何区别?

中文文本首先要进行分词,

分词的结果会影响 Parsing 效果.

[武汉市][长江大桥]

[武汉][市长][江大桥]

How do we choose features?

Bilexical affinities	双词汇亲和: [discussion-> issue]
Dependency distance	距离越近的词语越容易有依存关系
Intervening material:	中介词语: 依存关系几乎不会跨动词/标点
Valency of heads	词语配价: 一个词语最多有几个依赖者。

Dependency Parsing

- Constrains:
 - Only one word is dependent of Root
 - No cycle!
- Methods:
 - DP
 - Graph algorithms:
 - Minimum Spanning Tree
 - Constraint Satisfaction:
 - Eliminate edges that do not meet constraints
 - Transition-based Parsing and Deterministic Dependency Parsing:
 - Machine learning classifiers

Arc-standard transition-based parser

拥有三种操作 ("Left-Arc", "Right-Arc", "Shift") 的自动机

一个stack与一个buffer, 当stack仅剩ROOT, buffer为空时解析完成

stack	buffer	new dependency	transition
[ROOT, parsed, this]	[sentence, correctly]		SHIFT
[ROOT, parsed, this, sentence]	[correctly]		SHIFT
[ROOT, parsed, sentence]	[correctly]	sentence→this	LEFT-ARC
[ROOT, parsed]	[correctly]	parsed→sentence	RIGHT-ARC
[ROOT, parsed, correctly]	[]		SHIFT
[ROOT, parsed]	[]	parsed→correctly	RIGHT-ARC
[ROOT]	[]	ROOT→parsed	RIGHT-ARC

当 stack 中仅有 ROOT, 而 Buffer 非空
预测 Left/Right-Arc 造成停机, 如何解决?

Use classifier to predict operations at a specific condition

Fast and scalable: O(n) time

Close to the best parser.

- Categorical feature: very sparse, cost a lot of time to compute features
- NN: dense representation of features.