

# 第21章 计算语言学

## 21.1.1

### 一指代消解 (coreference resolution) | 线性话语分割 (linear segmentation)

① 信息抽取,

② 摘要,

③ 会话代理,

↓ 无监督的, 基于词汇内聚性.

- Text Tiling 算法, 包含:

① 分词 (tokenization)

② 词汇分值确定 (lexical score determination)

③ 边界识别 (boundary identification)

- 评价方法:

4) ① 分界前后的词语向量,

② 余弦度量  $\text{sim}_{\cos}(\vec{b}, \vec{a}) = \frac{\vec{b} \cdot \vec{a}}{|\vec{b}| |\vec{a}|}$

③ 深度分数 (depth score)

## 21.1.2 有监督的话语分割

① 二元分类器 (SVM / 决策树)

② 序列分类器 (HMM / CRF)

(x) 准确率, 召回率, F值

(v)  $P_k$  最新改进的 Windows Diff 算法

## 21.2 连贯关系理论: 修辞结构理论.

### 21.4 指代现象, → 21.5 代词指代消解所使用的特征 (人称, 性, 数的一致)

### 21.6 指代消解的3种算法:

① hobbs 算法 (通过端点回指)

② 中心理论 (centering theory) 中心理论话语模型