

“自然语言处理导论”课程讲义

# 共指消解

# Coreference Resolution

孙栩

信息科学技术学院

[xusun@pku.edu.cn](mailto:xusun@pku.edu.cn)

## □ 什么是共指消解?

## □ 表述检测

## □ 共指消解方法

### □ 规则方法

### □ 学习方法

- 表述对模型
- 表述排序模型

### □ 深度学习方法

- 表述排序
- 表述聚类排序
- 记忆网络(零指消解)

## □ 什么是共指消解?



## □ 表述检测

## □ 共指消解方法

### □ 规则方法

### □ 学习方法

- 表述对模型

- 表述排序模型

### □ 深度学习方法

- 表述排序

- 表述聚类排序

- 记忆网络(零指消解)

# 什么是共指消解?

## □ 共指消解 (Coreference Resolution)

- 为文本中的**表述**确定其在真实世界中所指向的**实体**的过程
- **表述(mention)**: 又称**指称语**
- **实体(entity)**: 抽象概念, 在文本中的具体体现是各种**表述**

## □ 表述的例子

例1: 张三对人很热情, 大家都叫他 张哥。张哥是一名医生, 他工作非常认真负责, 同时, 他也是一个好父亲。

例2: 美利坚合众国总统 巴拉克·奥巴马将于11月15日至18日对中国进行国事访问。

例3: 两会闭幕后, 今年下半年将召开中国共产党第十八届全国代表大会。目前, 十八大的各项准备工作正有条不紊地进行。

# 什么是共指消解?

## □ 表述的分类

- 普通名词短语
- 专有名词
- 代词

例1：张三对人很热情，大家都叫他张哥。张哥是一名医生，他工作非常认真负责，同时，他也是一个好父亲。

例2：美利坚合众国总统 巴拉克·奥巴马将于11月15日至18日对中国进行国事访问。

例3：两会闭幕后，今年下半年将召开中国共产党第十八届全国代表大会。目前，十八大的各项准备工作正有条不紊地进行。

# 什么是共指消解?

## □ 指代类型的语言学观点

- 缩略语：例3
- 别名：例1中的“张哥”和“张三”
- 同位语：例2
- 零形指代：[警方]表示他们自杀的可能性很高，不过 $\phi_1$ 也不排除 $\phi_2$ 有他杀的可能。 $\phi_1$ 指代警方， $\phi_2$ 虚指这个死亡事件，不可消解。

例1：张三对人很热情，大家都叫他张哥。张哥是一名医生，他工作非常认真负责，同时，他也是一个好父亲。

例2：美利坚合众国总统 巴拉克·奥巴马将于11月15日至18日对中国进行国事访问。

例3：两会闭幕后，今年下半年将召开中国共产党第十八届全国代表大会。目前，十八大的各项准备工作正有条不紊地进行。

# 共指消解的难点

## □ 共指消解很困难

- 识别所有的表述
- 可能有多个实体
- 不是所有的表述均有共指

## □ 例子：所有的表述

Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.

# 共指消解的难点

## □ 共指消解很困难

- 识别所有的表述
- 可能有多个实体
- 不是所有的表述均有共指

## □ 例子：多个实体



Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.



# 共指消解的难点

## □ 共指消解很困难

- 识别所有的表述
- 可能有多个实体
- 不是所有的表述均有共指

## □ 例子：多个实体



Barack Obama nominated **Hillary Rodham Clinton** as his **secretary of state** on Monday. He chose **her** because **she** had foreign affairs experience as a former **First Lady**.

# 共指消解的难点

## □ 共指消解很困难

- 识别所有的表述
- 可能有多个实体
- 不是所有的表述均有共指

## □ 例子：无共指表述

Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.

## □ 需要推理和世界知识

She poured water from the pitcher into the cup until it was full

She poured water from the pitcher into the cup until it was empty

The trophy would not fit in the suitcase because it was too big

The trophy would not fit in the suitcase because it was too small

## □ 这类句式被称为Winograd Schema

- Named after Terry Winograd, a professor of computer science at Stanford University

# Winograd Schema Challenge (WSC)

## □ Winograd Schema Challenge (WSC)

- 选择题：让模型选择上述句式中介词真正所指
- 被设计为图灵测试的改进版
  - 图灵测试需要人参与，而人很容易被欺骗
  - 共指消解，尤其是上述句式的消解，需要世界知识和基本推理
- 2014年起，Nuance Communications设立25,000美元奖金
  - 已进行两届，分别在IJCAI-16和AAAI-18
  - 目前没有胜出者

## □ 共指指的是两个表述指向同一实体

- Barack Obama traveled to ... Obama
- Obama 与 Barack Obama 是共指但不是回指

## □ 另一种指代是回指(anaphora)

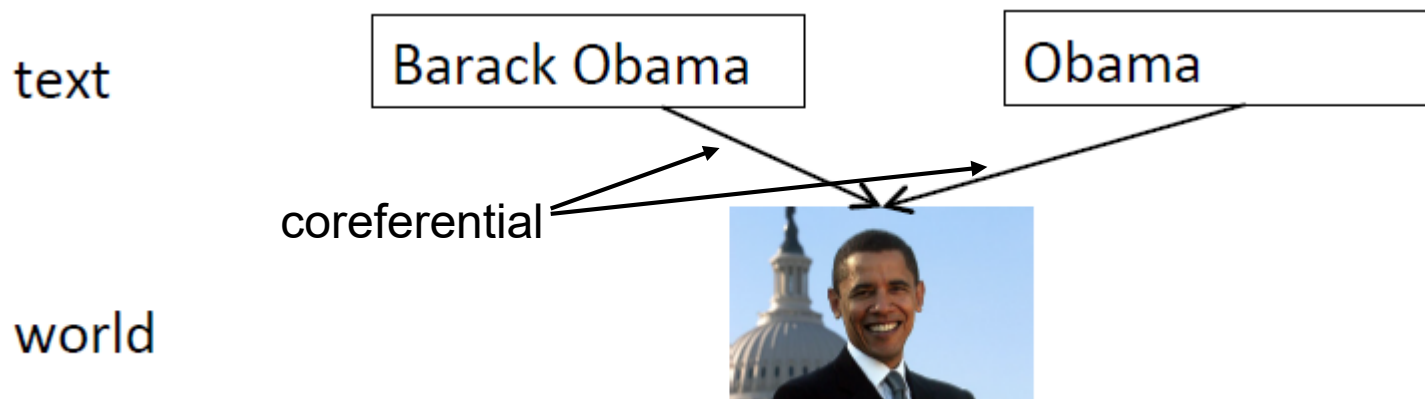
- 一个词语（回指成分, anaphora）指向另一个词语（先行语, antecedent），且回指成分的含义由先行语的含义确定
- Barack Obama said he would sign the bill.  
          antecedent          anaphor
- 确定回指成分及其先行语的过程被称为指代消解

## □ 广义上，共指消解和指代消解可以大致视为一个任务

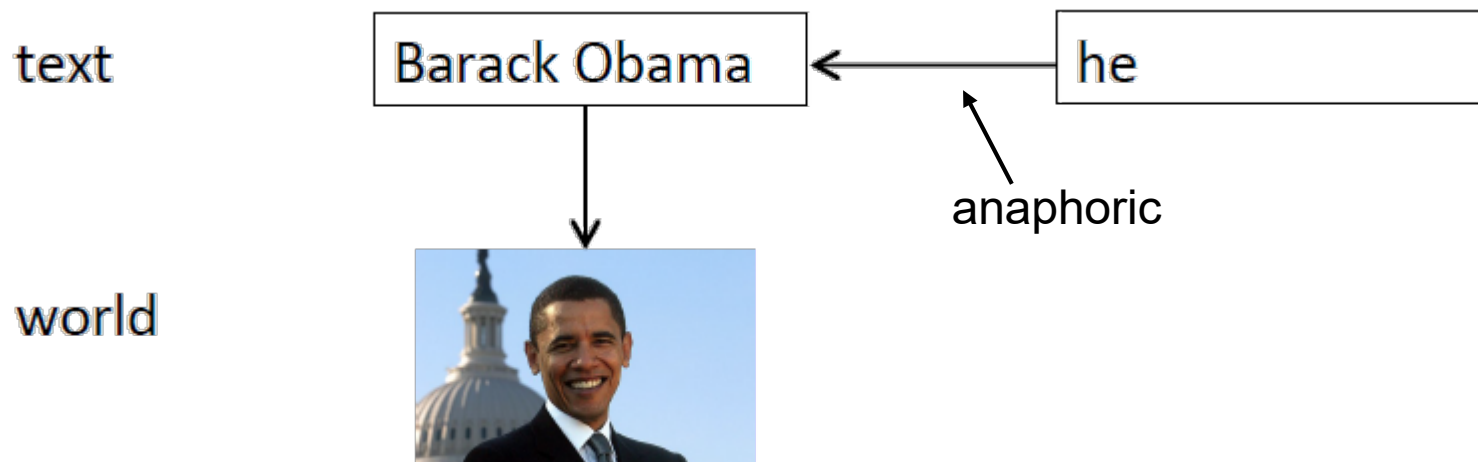
- 方法上有共性，但任务的目标、对象、评价均不同

# 共指消解与指代消解

## □ 共指:



## □ 回指:



## □ 不是所有的回指关系都是共指

### □ 例子

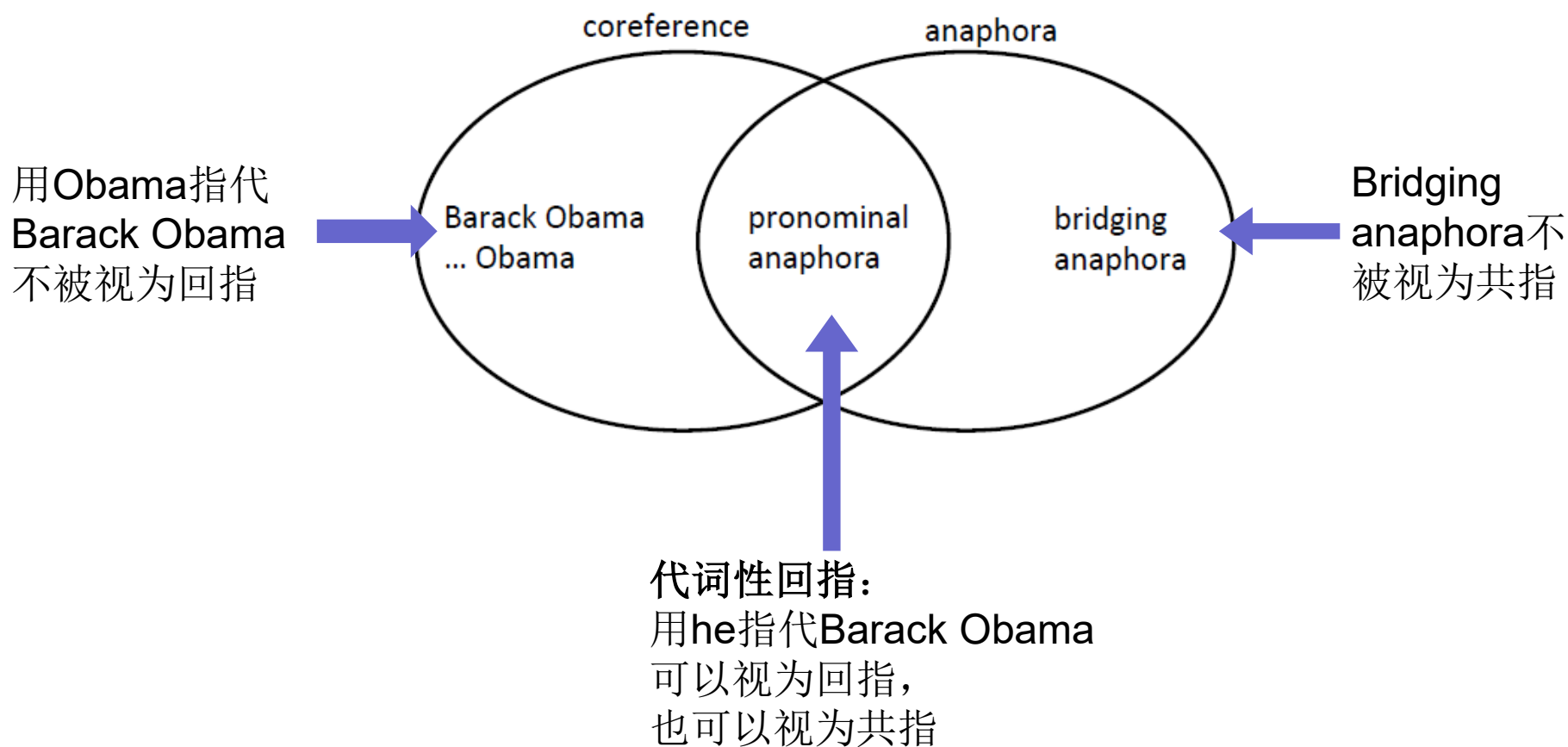
- We went to see a concert last night. The tickets were really expensive
- a concert与the tickets并不指称同一实体，但the tickets的含义由a concert从某种程度上确定
- 一般称之为bridging anaphora，这类指代与人如何理解自然语言密切相关

## □ 指代的另一类：外指

### □ 外指(exophora)：代词所指不在句子中，例子：

- This garden hose is better than that one.
- Jerry is standing over there.

## □ 共指和回指的区别





## □ 深入的文本理解

- 信息抽取，问答系统，摘要等
- He was born in 1961.
- 对如信息抽取，识别共指的名词短语有利于挖掘出更多实体和属性信息
- 对于问答系统和摘要，对于共指成分有选择性的用代词替换，可以增强文字的可读性

## □ 机器翻译

- 不同语言对于 gender, number, dropped pronouns有不同的特征
- 英文中的they，根据共指实体确定姓名，以正确的翻译为他们、她们 或 它们
- 对于中文中的零指，正确的翻译为对应的英文

## □ 机器翻译

- 如西班牙语译英语，google translate搞不清指代关系

English Spanish French Spanish - detected ▼



English Spanish Arabic ▼

Translate

a Alicia le gusta Juan porque es inteligente



44/5000



Alicia likes Juan because he's smart



Suggest an edit

English Spanish French Spanish - detected ▼



English Spanish Arabic ▼

Translate

a Juan le gusta Alicia porque es inteligente



44/5000



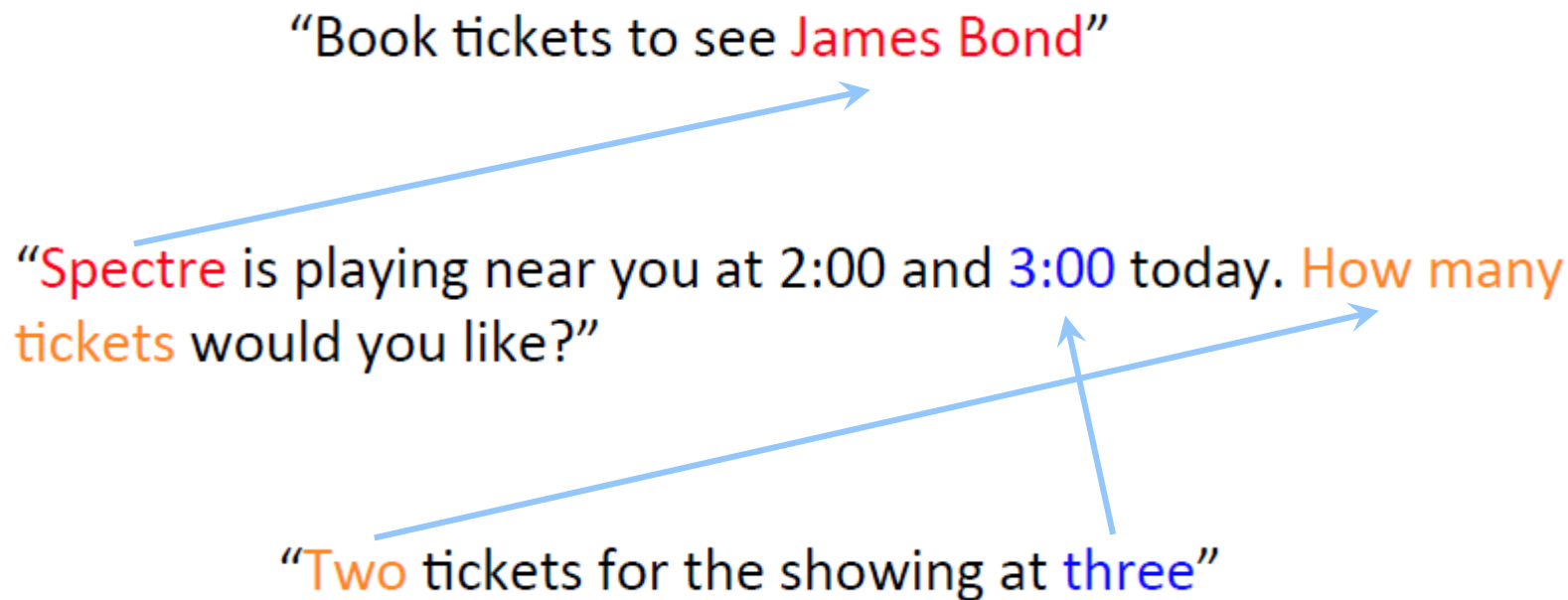
Juan likes Alicia because he's smart



Suggest an edit

## □ 对话系统

- 明确不同表述的对应实体，是对话有效进行的基础



- 共指消解十分重要，最近30年国际评测不断

## □ MUC (Message Understanding Conference)

- 美国国防高级研究计划委员会(DARPA)资助, 1987-1998 共7届
- 1995年第6届与1998年第7届中出现了共指消解相关任务
- 评测语料库均为英文

## □ ACE (Automatic Content Extraction)

- 美国国家标准与技术研究院(NIST)组织, 持续到2008年
- 主要任务之一是实体检测与跟踪, 将**篇章**中出现的各种表述指向对应的实体
  - 包含大量单表述实体(无共指关系的表述, singleton entity)
- 2003年起, 开始提供中文语料的共指消解任务

## □ TAC (Text Analysis Conference)

- 2009年起替代ACE, 共指消解任务过渡为实体链接任务

## ▣ SemEval (Semantic Evaluation)

- ▣ 2010年增加了多语言共指消解任务，采用OntoNotes 2.0数据集
- ▣ 不包含单表述实体，仅标注那些发生共指关系的表述

## ▣ CoNLL (Conference on Natural Language Learning)

- ▣ 2011年举办了英文的共指消解评测，采用OntoNotes 4.0版本
- ▣ 除了不包含单表述实体外，对于同位语以及谓语主格均不看作为共指关系
- ▣ 额外标注了事件名词与动词的共指关系（事件共指）
- ▣ 2012年举办了多语言共指消解评测，采用OntoNotes 5.0数据集
- ▣ 包含英文、中文以及阿拉伯文语料
- ▣ 研究不同语言的共指消解研究的差别与共性

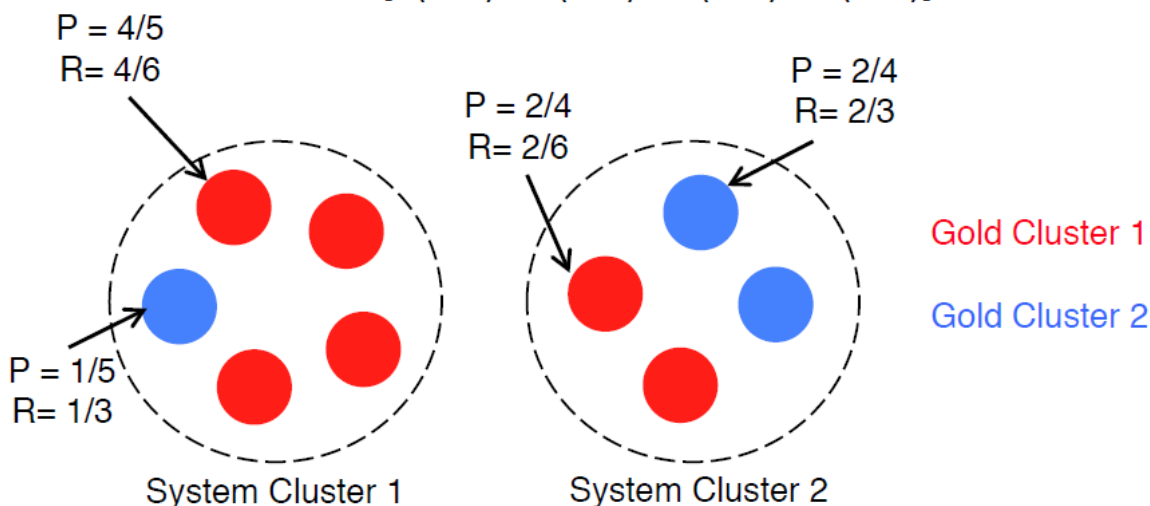
## □ 多种评价指标，如MUC, CEAF, LEA, B-CUBED, BLANC等

- 通常汇报一些指标的平均值

## □ 以B-CUBED 为例

- 对每个表述计算准确率与召回率
- 然后对每个表述的准确率和召回率分别平均

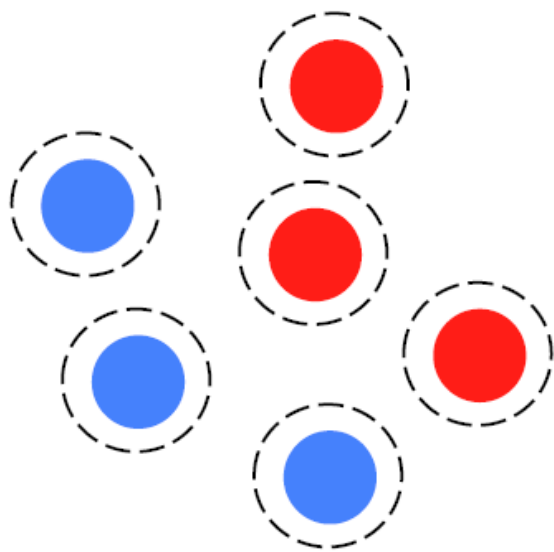
$$P = [4(4/5) + 1(1/5) + 2(2/4) + 2(2/4)] / 9 = 0.6$$



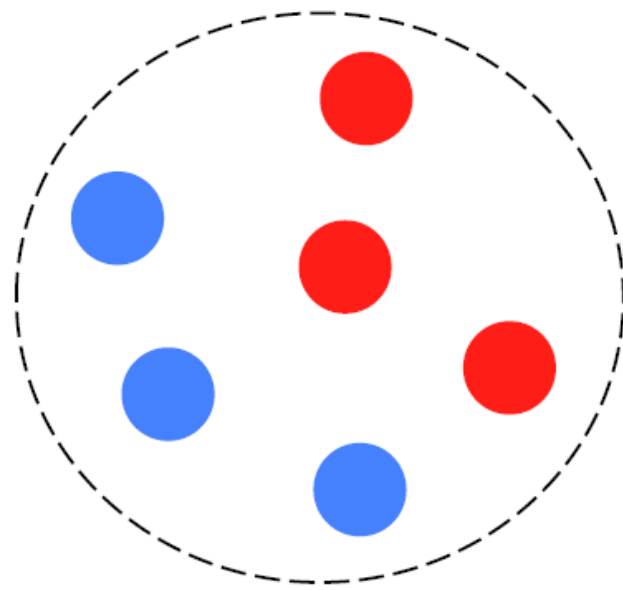
## ■ B-CUBED

- 每个表述单独一类：100% Precision
- 所有表述分为一类：100% Recall

100% Precision, 33% Recall



50% Precision, 100% Recall,



# 共指消解的两步骤解决方法

## □ 1. 检测表述 (较为简单)

- “[I] voted for [Nader] because [he] was most aligned with [[my] values]” , [she] said
- 表述是可以嵌套的!

## □ 2. 对表述聚类 (较为困难)

- “[I] voted for *[Nader]* because *[he]* was most aligned with [[my] values]” , **[she]** said
- 类1: I, my, she
- 类2: Nader, he
- 类3: my values



## □ 什么是共指消解?

## □ 表述检测

## □ 共指消解方法

### □ 规则方法

### □ 学习方法

- 表述对模型
- 表述排序模型

### □ 深度学习方法

- 表述排序
- 表述聚类排序
- 记忆网络(零指消解)

## □ 表述：一段指向某一实体的文本

- 可以分为3类
- 检测：使用现有的NLP系统

## □ 代词：词性标注

- I, your, it, she, him等

## □ 命名实体：命名实体识别

- 人名、地名等

## □ 名词短语：chunking或短语结构句法分析

- a dog, the big fluffy cat stuck in the tree

## □ 如此简单吗？

- 将所有的代词、命名实体、名词短语标记为表述有过度泛化的嫌疑

## □ 以下文本是否算表述？（实现中不算，理论中是灰色地带）

- It is sunny 中的 it
- Every student
- No student
- The best donut in the world
- 100 miles

## □ 与自然语言有关的定义总是存在灰色地带的

- Pick a convention and go with it
- 对于计算语言研究者来说 选定标准 继续做就好

## □ 什么是共指消解?

## □ 表述检测

## □ 共指消解方法

### □ 规则方法

### □ 学习方法

- 表述对模型
- 表述排序模型

### □ 深度学习方法

- 表述排序
- 表述聚类排序
- 记忆网络(零指消解)



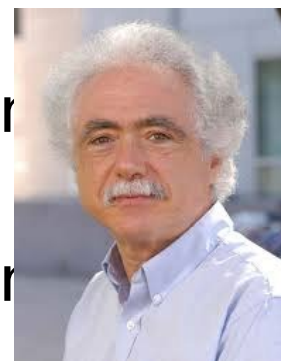
## □ Hobbs' 算法/树查询算法(1978)

- 经典
- 针对**代词**的回指消解，在句法树上运作两种算法
  - 朴素Hobbs算法，完全基于句法知识
  - 另一种Hobbs算法，既考虑句法知识又考虑语义知识
- 算法根据英文语言上的直觉编写，可以拿到80%准确率



## □ Winograd等认为指代消解的完全解决必须有一个外部知识库(1972)

- [The city council] refused [the women] a permit because [they] feared violence.
- [The city council] refused [the women] a permit because [they] advocated violence.
- 但效果不如Hobbs算法



## ■ A Multi-Pass Sieve System (多遍过滤方法)

- CoNLL 2011 Winner, Stanford
- 基于准确率由高到低构建了多个筛子(Sieve)
  - 每个筛子都是一个规则系统
  - 用于为文本中出现的不同表述确定可能的先行语
- 每层过滤基于之前得到的共指结果
- 每层过滤可以利用之前筛子得到的表述的属性信息
- 方法简单，但超过一众机器学习方法

### Ordered sieves

---

1. **Mention Detection Sieve**
2. **Discourse Processing Sieve**
3. **Exact String Match Sieve**
4. **Relaxed String Match Sieve**
5. **Precise Constructs Sieve** (e.g., appositives)
- 6-8. **Strict Head Matching Sieves A-C**
9. **Proper Head Word Match Sieve**
10. **Alias Sieve**
11. **Relaxed Head Matching Sieve**
12. **Lexical Chain Sieve**
13. **Pronouns Sieve**

## □ 什么是共指消解?

## □ 表述检测

## □ 共指消解方法

### □ 规则方法

### □ 学习方法

- 表述对模型
- 表述排序模型



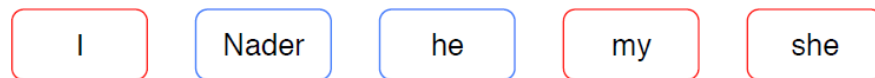
### □ 深度学习方法

- 表述排序
- 表述聚类排序
- 记忆网络(零指消解)

# 1. 表述对模型

## □ 以下面句子为例

- “I voted for **Nader** because **he** was most aligned with **my** values” , **she** said.



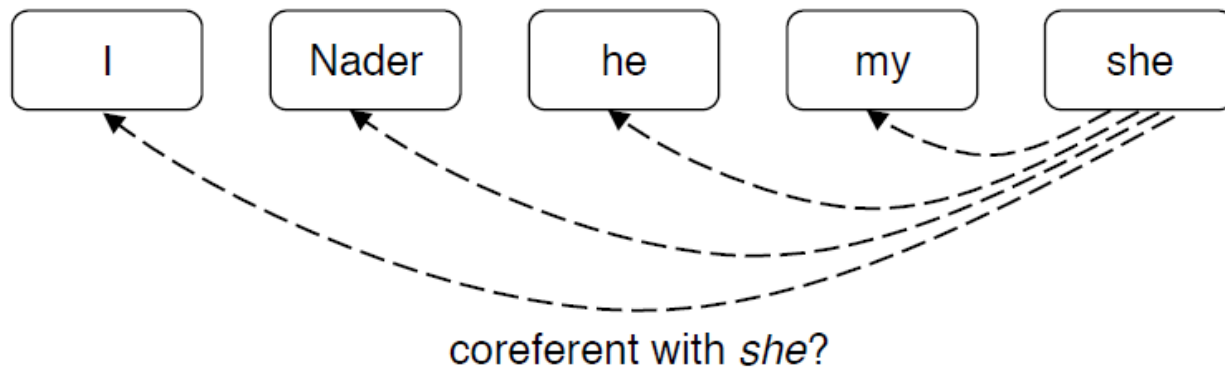
Coreference Cluster 1

Coreference Cluster 2

## □ 表述对

- 训练二元分类器计算每一对表述的共指概率 $p(m_i, m_j)$
- 例如，对于最后出现的she，计算所有之前出现的回指候选

*“I voted for Nader because he was most aligned with my values,” she said.*





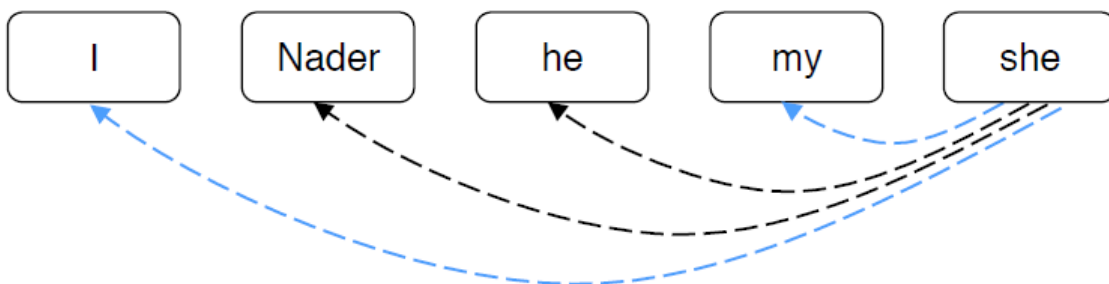
# 1. 表述对模型

## □ 表述对模型的训练

- 训练二元分类器计算每一对表述的共指概率 $p(m_i, m_j)$
- 例如，对于最后出现的she，计算所有之前出现的回指候选

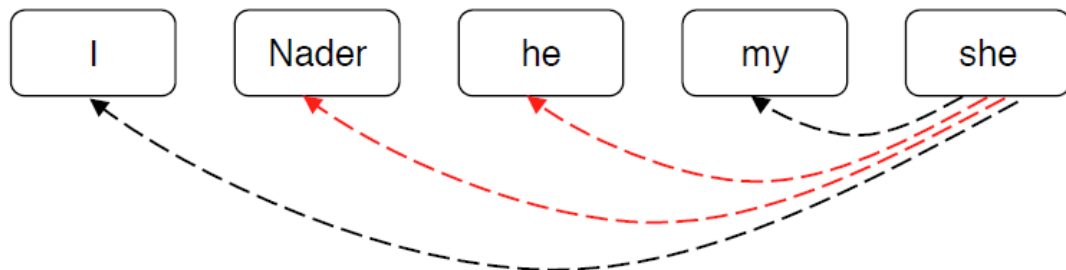
- 正例概率为1

*"I voted for Nader because he was most aligned with my values," she said.*



- 负例概率为0

*"I voted for Nader because he was most aligned with my values," she said.*

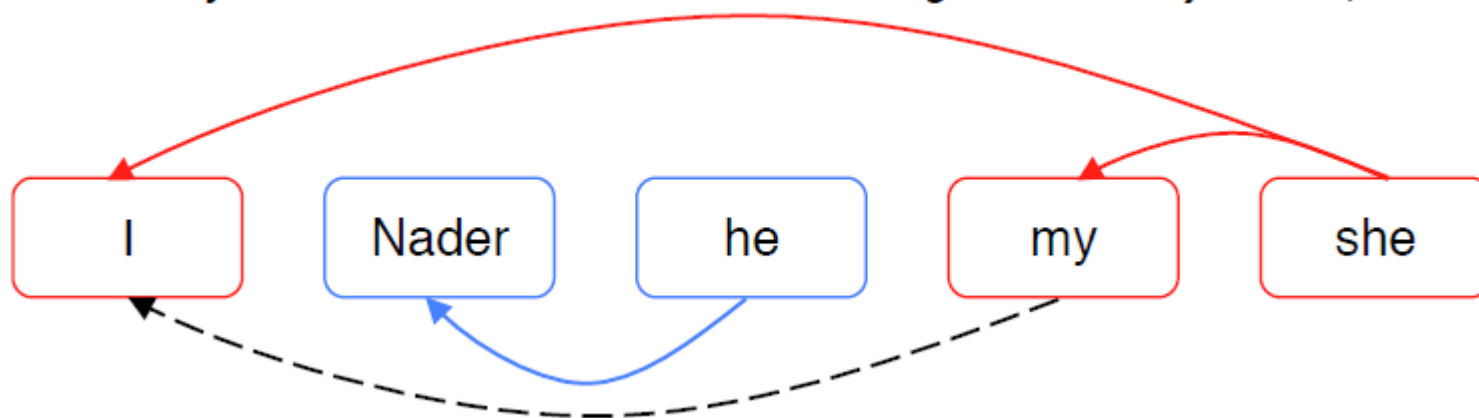


# 1. 表述对模型

## □ 表述对模型在测试时的推理过程

- 共指消解是聚类的任务
- 获得每个表述对评分后，根据设定的**阈值**添加共指关系
- 根据共指关系取并集，获得聚类结果
  - 下图中的虚线，模型并未预测出来
  - 但可以通过取并集获得

*"I voted for Nader because he was most aligned with my values," she said.*

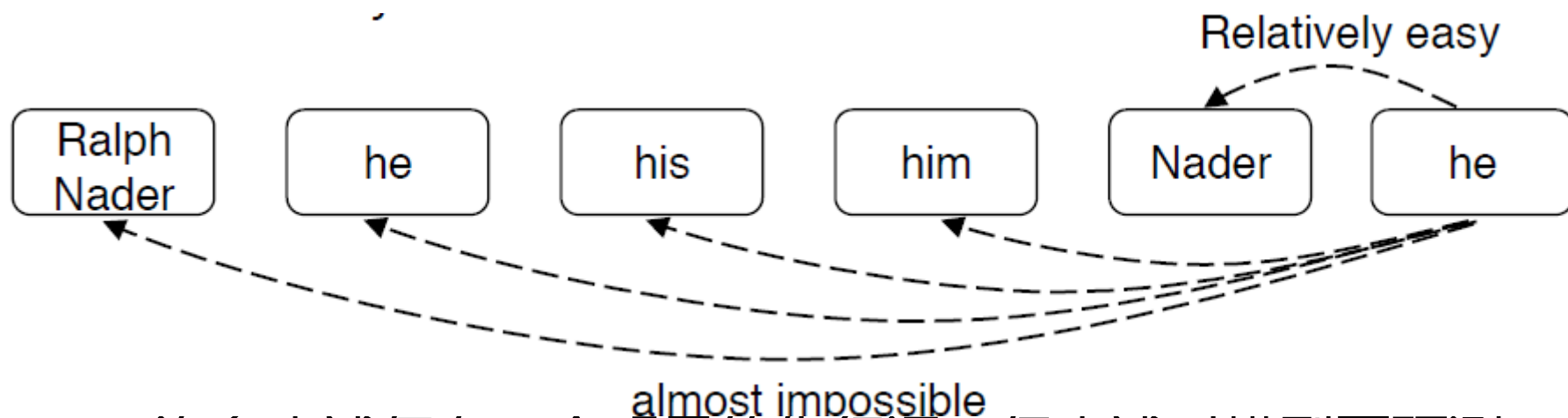


# 1. 表述对模型

## □ 表述对模型的不足

### □ 假设给定的文本非常长

- **Ralph Nader ... he ... his ... him ... <paragraphs> .. voted for Nader because he ...**



- 许多表述仅有一个明显的先行词，但表述对模型要预测一个表述所有的先行词
- 解决方案：模型对每个表述仅预测一个先行词
  - 语言学上更合理

## □ 什么是共指消解?

## □ 表述检测

## □ 共指消解方法

### □ 规则方法

### □ 学习方法

- 表述对模型

- 表述排序模型

### □ 深度学习方法

- 表述排序

- 表述聚类排序

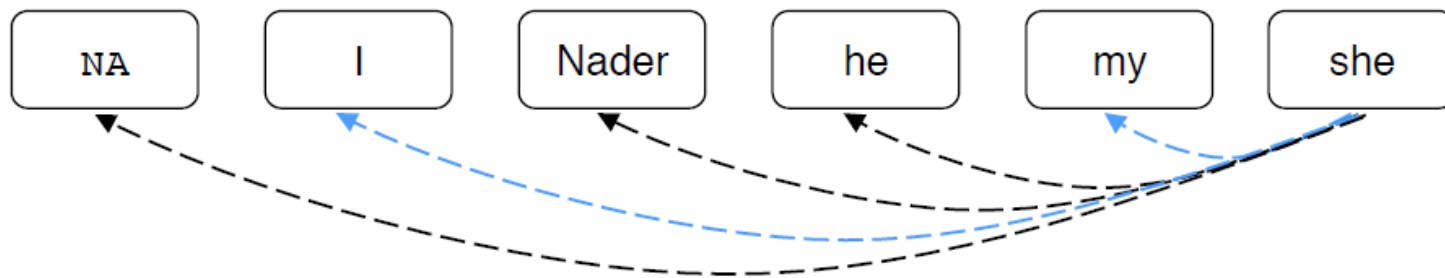
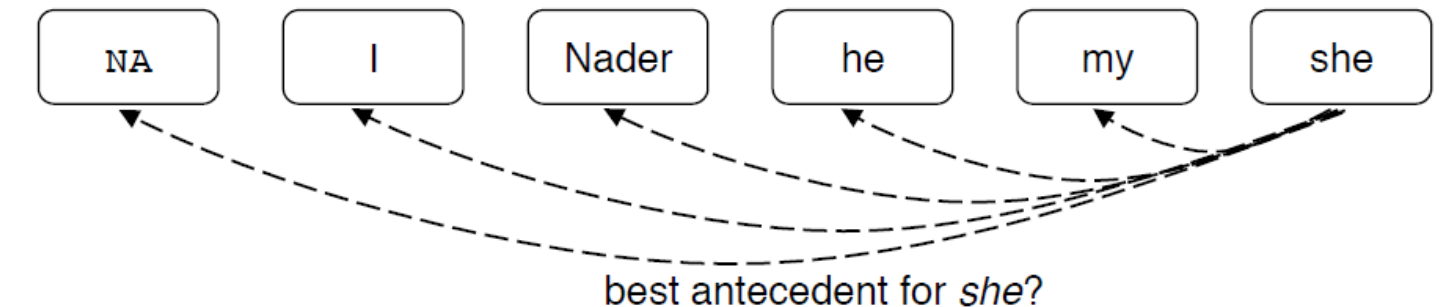
- 记忆网络(零指消解)



## 2. 表述排序模型

### □ 表述排序模型

- 对于一个表述，仅选出最可能的一个先行词
  - 额外的“NA表述”以允许模型不连接当前表述



Positive examples: model has to assign a high probability to either one (but not necessarily both)

## 2. 表述排序模型

### □ 表述排序模型

- 对于一个表述，仅选出最可能的一个先行词
  - 额外的“NA表述”以允许模型不连接当前表述



only add highest scoring  
coreference link

$$p(\text{NA}, \text{she}) = 0.1$$

$$p(\text{I}, \text{she}) = 0.5$$

$$p(\text{Nader}, \text{she}) = 0.1$$

$$p(\text{he}, \text{she}) = 0.1$$

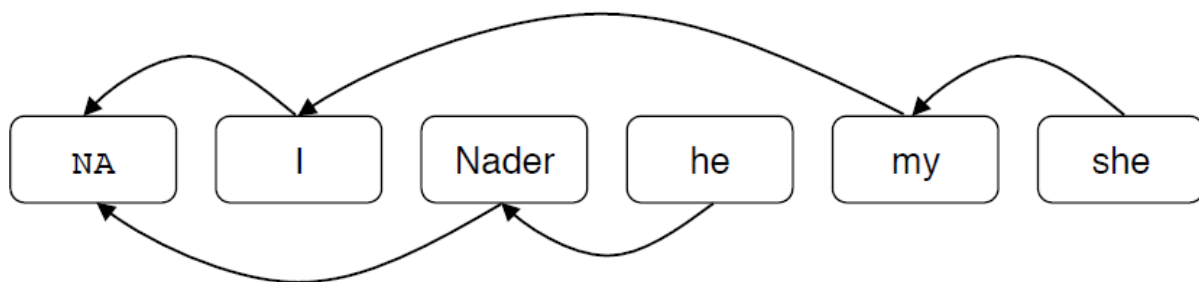
$$p(\text{my}, \text{she}) = 0.2$$

} Apply a softmax over the scores for  
candidate antecedents so  
probabilities sum to 1

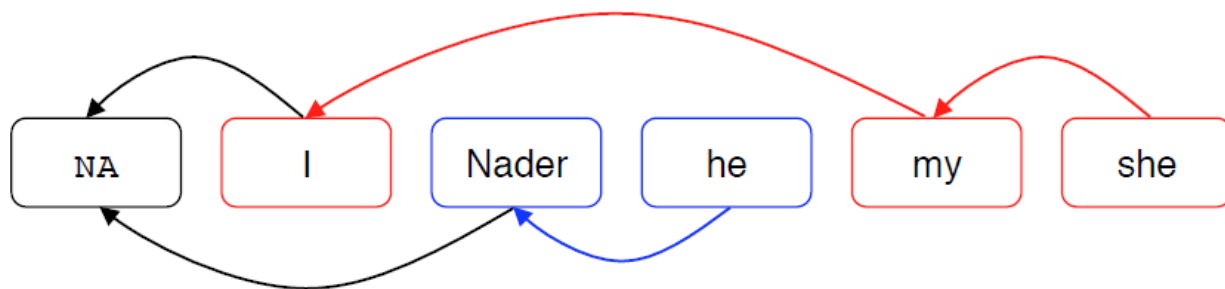
## 2. 表述排序模型

### □ 表述排序模型在测试时的推理过程

- 与表述对模型基本一致，除了每个表述只连接了一个先行词



- 聚类结果：同样采用取并集的方法



- **表述对模型和表述排序模型均需要计算分数/概率**

- **两种方法**

- 非神经网络基于统计的分类器
- 简单神经网络

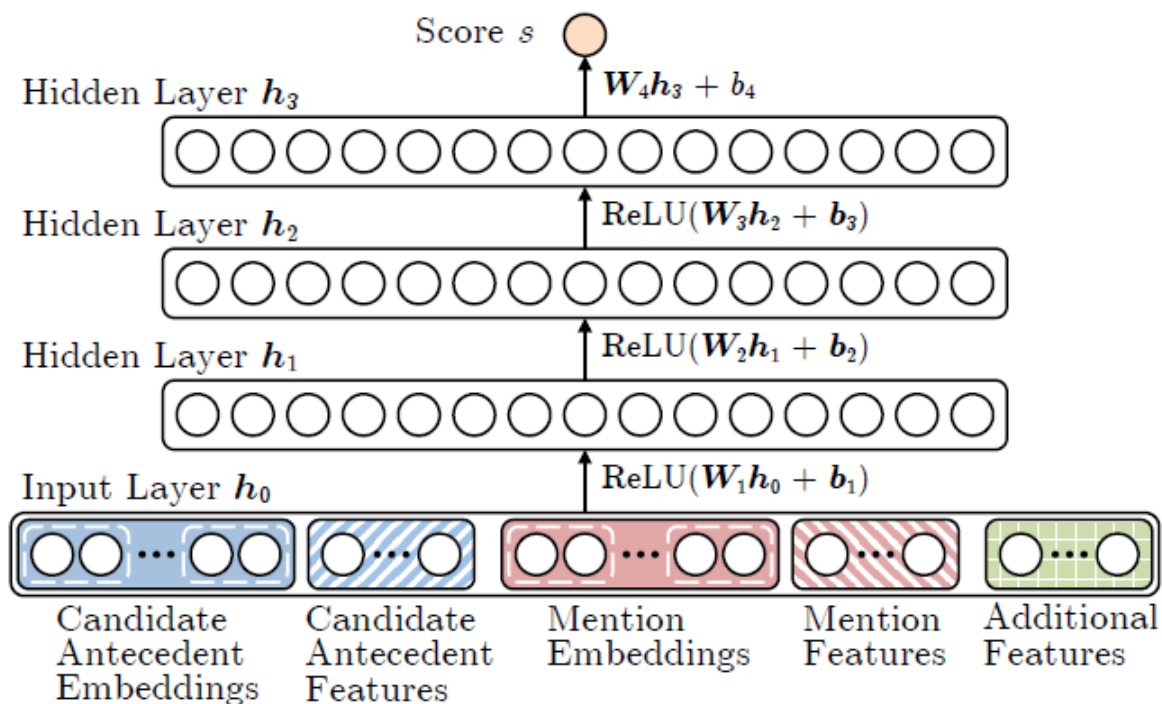


## □ 需要设计有效的特征集

- Person/Number/Gender agreement
  - Jack gave **Mary** a gift. **She** was excited.
- Semantic compatibility
  - ... **the mining conglomerate** ... **the company** ...
- Certain syntactic constraints
  - John bought **him** a new car. [him can not be John]
- More recently mentioned entities preferred for referenced
  - **John** went to a movie. **Jack** went as well. **He** was not busy.
- Grammatical Role: Prefer entities in the subject position
  - **John** went to a movie with **Jack**. **He** was not busy.
- Parallelism:
  - **John** went with **Jack** to a movie. **Joe** went with **him** to a bar.
- ...

## □ 使用标准前馈神经网络

- 输入：词向量和一些分类特征
- 词向量：每个表述之前两个词，内第一个词、最后一个词、头词
  - 头词即表述中最重要的词，用parser可以得到
- 分类特征



## ▣ OntoNotes: 3000人工标注文档

- ▣ 中英数据
- ▣ F1 score

Model	English	Chinese	
Lee et al. (2010)	~55	~50	Rule-based system, used to be state-of-the-art!
Chen & Ng (2012) [CoNLL 2012 Chinese winner]	54.5	57.6	
Fernandes (2012) [CoNLL 2012 English winner]	60.7	51.6	Non-neural machine learning models
Wiseman et al. (2015)	63.3	—	
Clark & Manning (2016)	65.4	63.7	Neural clustering model
Lee et al. (2017)	67.2	--	
			End-to-end neural mention ranker

## ▣ OntoNotes: 3000人工标注文档

- ▣ 神经网络对于那些没有字符串匹配的名词短语和命名实体有很大标注

方法	普通名词短语	命名实体识别
深度学习方法	18.90	68.7
非深度学习方法	10.71	66.1

— 这些模型在自然语言处理任务中取得了显著的成功  
Example Wins

Anaphor	Antecedent
the country's leftist rebels	the guerillas
the company	the New York firm
216 sailors from the ``USS cole''	the crew
the gun	the rifle

## □ 什么是共指消解?

## □ 表述检测

## □ 共指消解方法

### □ 规则方法

### □ 学习方法

- 表述对模型
- 表述排序模型

### □ 深度学习方法

- 表述排序
- 表述聚类排序
- 记忆网络(零指消解)



## □ 介绍三篇论文

## □ 共指消解

### □ End-to-end Neural Coreference Resolution

□ UW EMNLP2017, citation 17

### □ Improving Coreference Resolution by Learning Entity-Level Distributed Representations

□ Stanford ACL 2016, citation 42

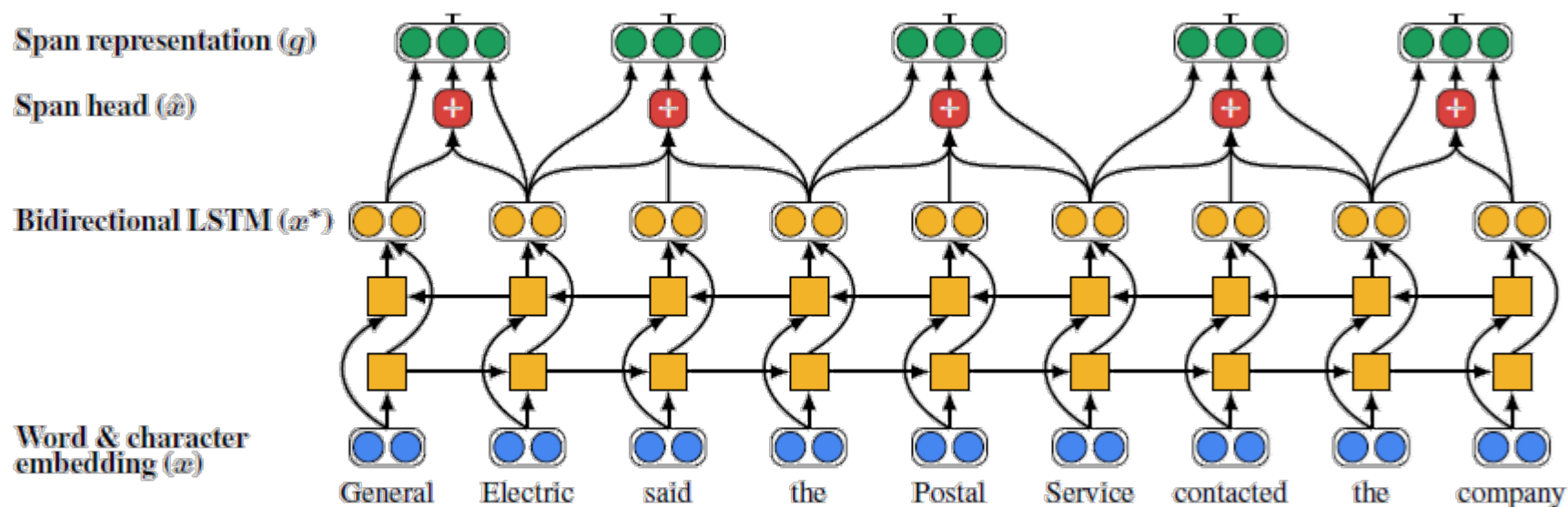
## □ 零指消解

### □ Chinese Zero Pronoun Resolution with Deep Memory Network

□ HIT EMNLP2017, citation 0

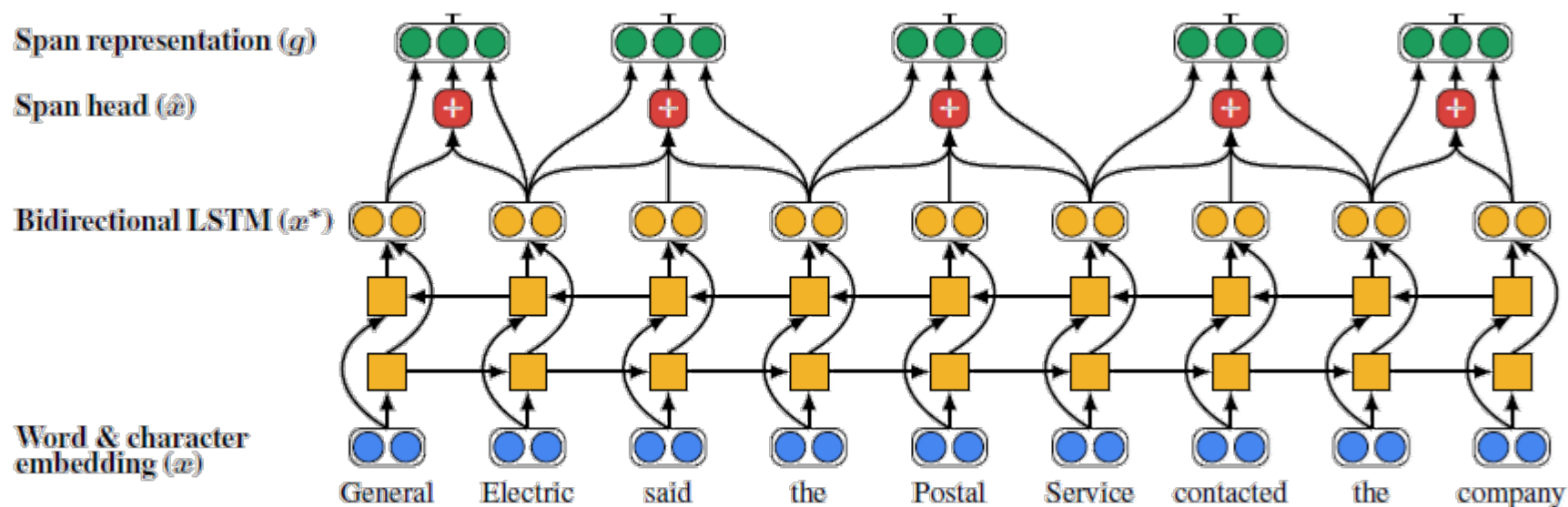
- **当前的SOTA**
- **Lee et al. EMNLP 2017**
- **属于表述排序模型**
- **对基于简单神经网络模型进行了一系列改进**
  - LSTM + attention
  - 端到端的表述检测和共指消解
    - 省去单独的表述检测步骤
    - 将所有一定长度内的文本片段(span)都视为表述候选

- 使用BLSTM获得每个span的表示
- 输入：word embedding matrix + char-level CNN

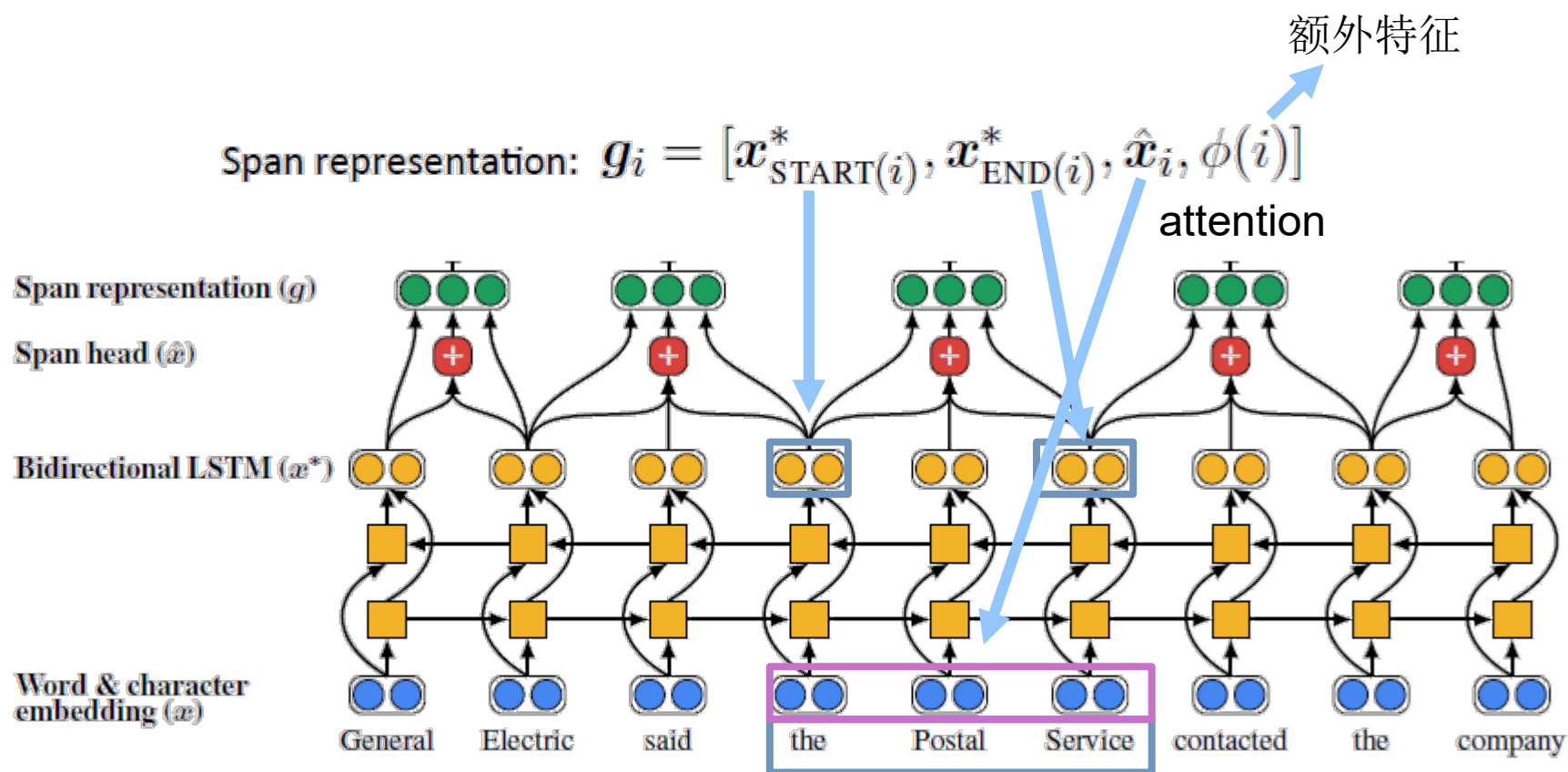




- 使用BLSTM获得每个span的表示
- Span: general, general electric, general electric said...



## □ 例如，对于the postal service



## □ 对于span的attention表示是基于BLSTM前的表示

- Query: 隐向量
- Key: 可训练权值矩阵
- Value: 输入向量

### Attention scores

$$\alpha_t = \mathbf{w}_\alpha \cdot \text{FFNN}_\alpha(\mathbf{x}_t^*)$$

dot product of weight  
vector and transformed  
hidden state

### Attention distribution

$$a_{i,t} = \frac{\exp(\alpha_t)}{\sum_{k=\text{START}(i)}^{\text{END}(i)} \exp(\alpha_k)}$$

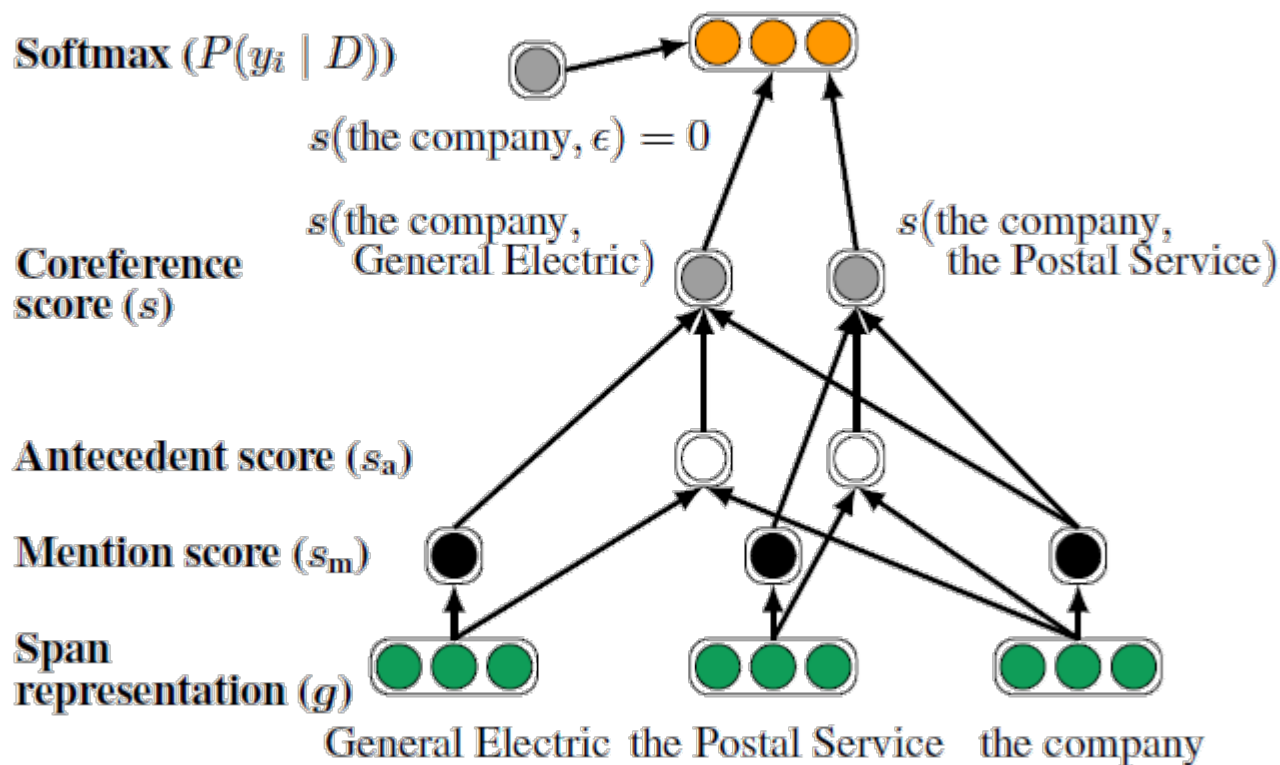
just a softmax over attention  
scores for the span

### Final representation

$$\hat{\mathbf{x}}_i = \sum_{t=\text{START}(i)}^{\text{END}(i)} a_{i,t} \cdot \mathbf{x}_t$$

Attention-weighted sum  
of word embeddings

## □ 某对span的得分



## □ 某对span的得分

$$s(i, j) = s_m(i) + s_m(j) + s_a(i, j)$$

Are spans  $i$  and  $j$  coreferent mentions?      Is  $i$  a mention?      Is  $j$  a mention?      Do they look coreferent?

## □ 评价函数利用span的向量表示

$$s_m(i) = w_m \cdot \text{FFNN}_m(g_i)$$

$$s_a(i, j) = w_a \cdot \text{FFNN}_a([g_i, g_j, g_i \circ g_j, \phi(i, j)])$$

include multiplicative interactions between the representations      again, we have some extra features

## □ 考虑所有的span对是不现实的

- $O(T^2)$ 对
- $O(T^4)$ 计算代价
- 需要剪枝，只考虑那些可能是表述的span

## □ 注意力可以学习在一个表述中哪些词是重要的

- Attention learns which words are important in a mention (a bit like head words)

(A **fire** in a Bangladeshi garment factory) has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee (the **blaze**) in the four-story building.

## □ 考虑所有的span对是不现实的

- $O(T^2)$ 对
- $O(T^4)$ 计算代价
- 需要剪枝，只考虑那些可能是表述的span

## □ 注意力可以学习在一个表述中哪些词是重要的

- Attention learns which words are important in a mention (a bit like head words)

(A **fire** in a Bangladeshi garment factory) has left at least 37 people dead and 100 hospitalized. Most of the deceased were killed in the crush as workers tried to flee (the **blaze**) in the four-story building.

## □ 有显著提升

	MUC			B <sup>3</sup>			CEAF <sub><math>\phi_4</math></sub>			Avg. F1
	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	
Our model (ensemble)	<b>81.2</b>	<b>73.6</b>	<b>77.2</b>	<b>72.3</b>	<b>61.7</b>	<b>66.6</b>	<b>65.2</b>	<b>60.2</b>	<b>62.6</b>	<b>68.8</b>
Our model (single)	78.4	73.4	75.8	68.6	61.8	65.0	62.7	59.0	60.8	67.2
Clark and Manning (2016a)	79.2	70.4	74.6	69.9	58.0	63.4	63.5	55.5	59.2	65.7
Clark and Manning (2016b)	79.9	69.3	74.2	71.0	56.5	63.0	63.8	54.3	58.7	65.3
Wiseman et al. (2016)	77.5	69.8	73.4	66.8	57.0	61.5	62.1	53.9	57.7	64.2
Wiseman et al. (2015)	76.2	69.3	72.6	66.2	55.8	60.5	59.4	54.9	57.1	63.4
Clark and Manning (2015)	76.1	69.4	72.6	65.6	56.0	60.4	59.4	53.0	56.0	63.0
Martschat and Strube (2015)	76.7	68.1	72.2	66.1	54.2	59.6	59.5	52.3	55.7	62.5
Durrett and Klein (2014)	72.6	69.9	71.2	61.2	56.4	58.7	56.2	54.2	55.2	61.7
Björkelund and Kuhn (2014)	74.3	67.5	70.7	62.7	55.0	58.6	59.4	52.3	55.6	61.6
Durrett and Klein (2013)	72.9	65.9	69.2	63.6	52.5	57.5	54.3	54.4	54.3	60.3



## □ How important is the feature or the mention?

	Avg. F1	$\Delta$
Our model (ensemble)	69.0	+1.3
Our model (single)	67.7	
– distance and width features	63.9	-3.8
– GloVe embeddings	65.3	-2.4
– speaker and genre metadata	66.3	-1.4
– head-finding attention	66.4	-1.3
– character CNN	66.8	-0.9
– Turian embeddings	66.9	-0.8

Table 2: Comparisons of our single model on the development data. The 5-model ensemble provides a 1.3 F1 improvement. The head-finding attention, features, and all word representations contribute significantly to the full model.

	Avg. F1	$\Delta$
Our model (joint mention scoring)	67.7	
w/ rule-based mentions	66.7	-1.0
w/ oracle mentions	85.2	+17.5

Table 3: Comparisons of various mention proposal methods with our model on the development data. The rule-based mentions are derived from the mention detector from [Raghunathan et al. \(2010\)](#), resulting in a 1 F1 drop in performance. The oracle mentions are from the labeled clusters and improve our model by over 17.5 F1.

## □ 优势

- Embedding: 基于词义的表述匹配
- Attention: 关注并列连词、头词

## □ 不足

- Embedding: 过度使用embedding中的相似度, 出现 false positive 的连接
  - The flight attendants 与 the pilots
- 意料之中, 模型无法解决需要世界知识的共指消解
  - Let the rescuer locate them
  - Them -> some ships
  - Gold: them -> the men overboard

## □ 什么是共指消解?

## □ 表述检测

## □ 共指消解方法

### □ 规则方法

### □ 学习方法

- 表述对模型
- 表述排序模型

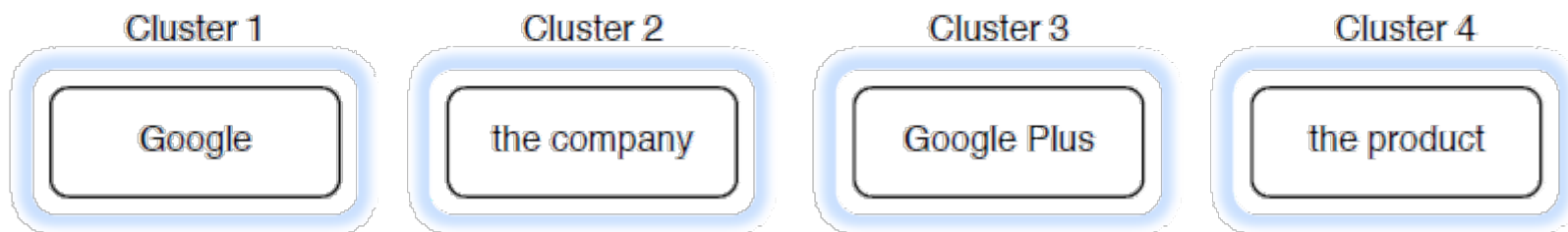
### □ 深度学习方法

- 表述排序
- 表述聚类排序
- 记忆网络(零指消解)



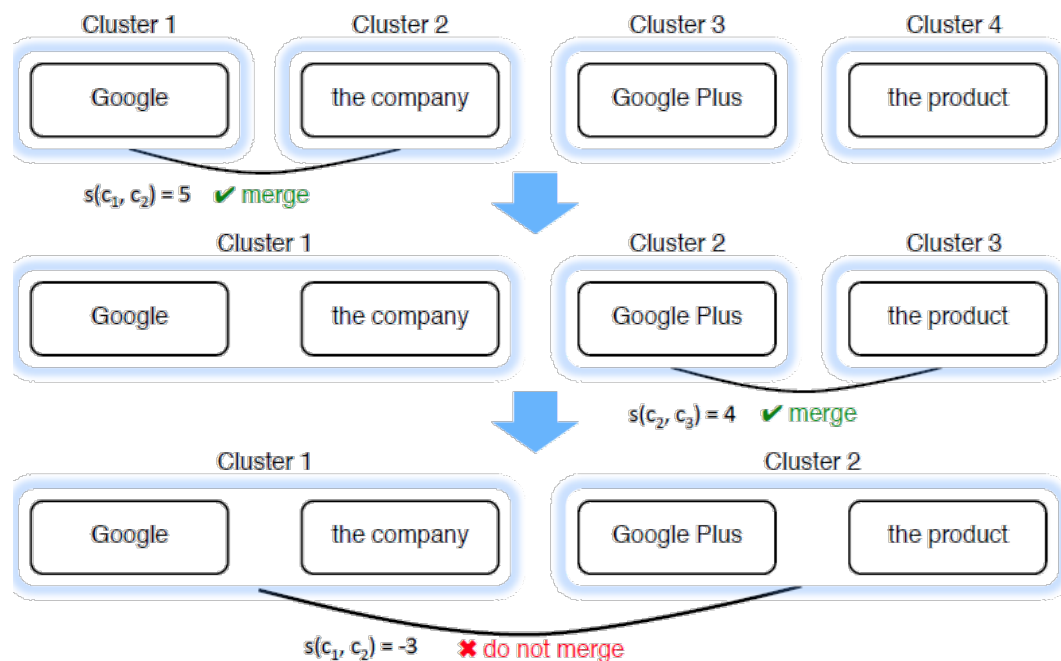
- 共指消解是个聚类任务，使用聚类方法非常自然
- 特别的，一般使用agglomerative clustering（凝聚聚类/丛集聚类）
- 首先，每个表述独自一类(cluster)

Google recently ... the company announced Google Plus ... the product features ...



- 然后，每次合并一对
  - 使用模型打分

- 首先，每个表述独自一类(cluster)
- 然后，每次合并一对
  - 使用模型打分

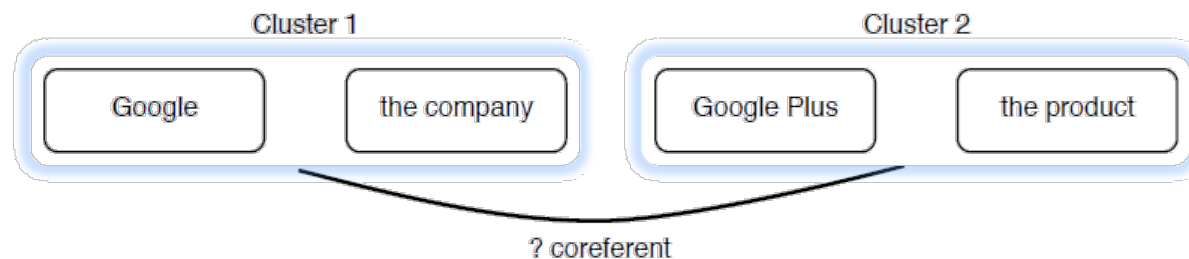
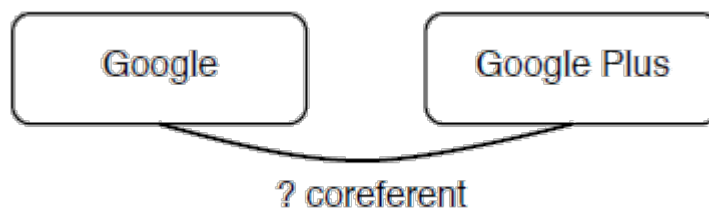


## □ 类对评分相对于表述对评分更简单

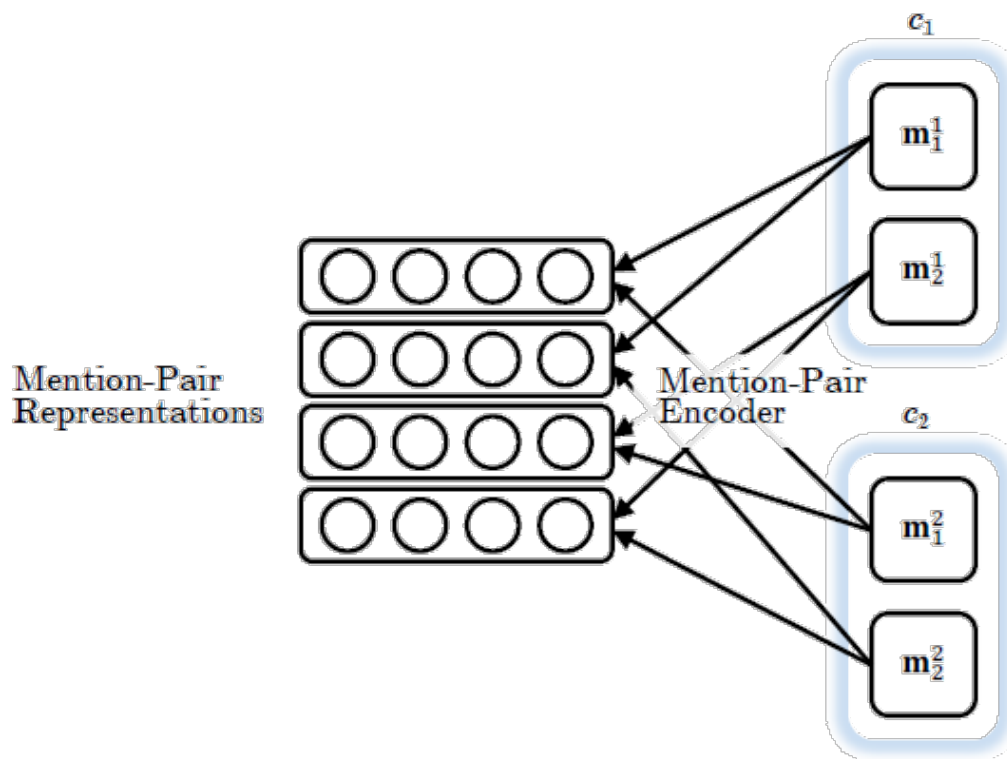
- 类内信息更充分

## □ 表述对评分

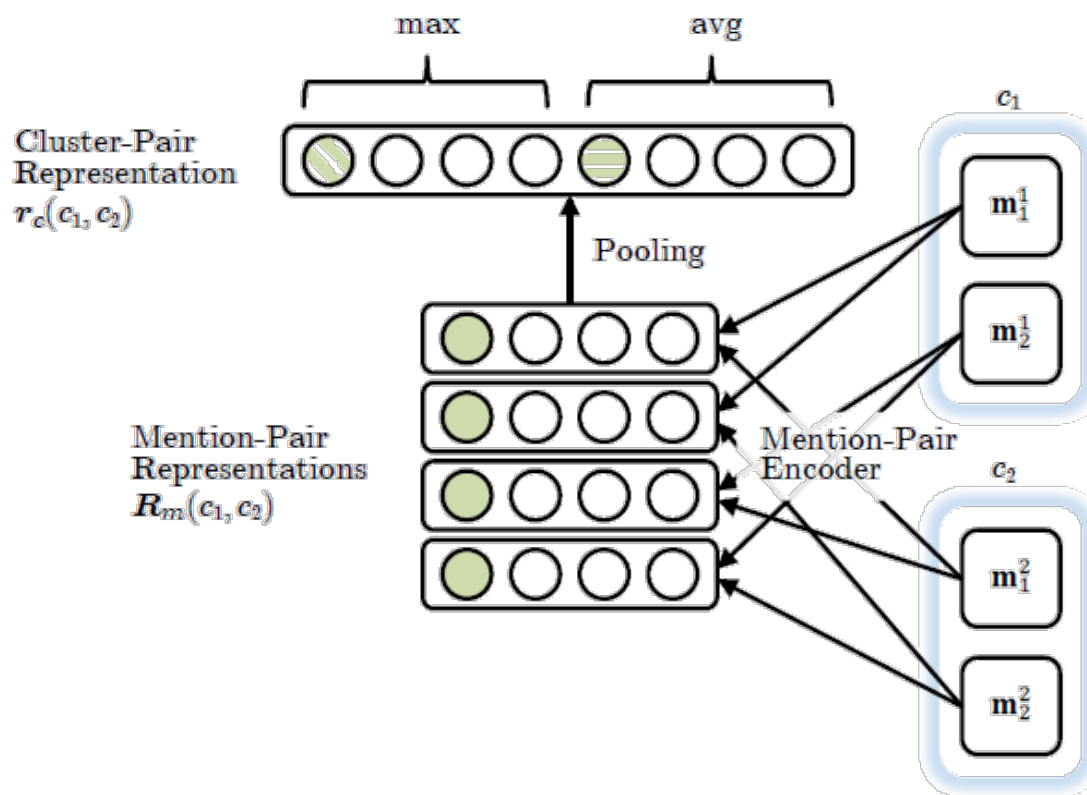
## □ 类对评分



- 先用前馈神经网络获得待合并类对中每对表述的表示

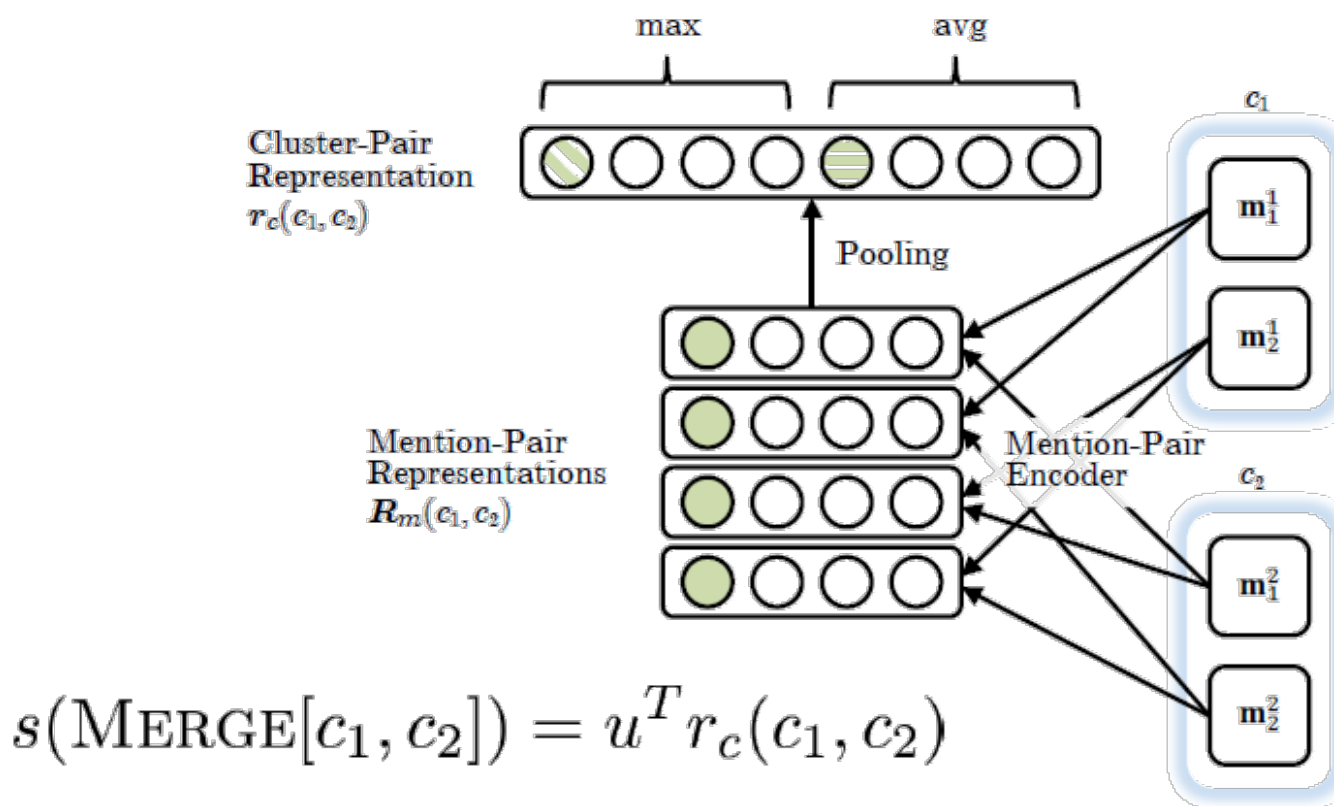


## □ 再用池化操作获得类对的表示





## □ 得分通过每维线性加和得到



## □ 训练

- 每一步依赖于之前的结果
- 可以使用普通的有监督学习方法：oracle序列
- 但测试时很可能出现不一致的情况
  
- 使用强化学习的方法训练模型
- 回报函数：合并导致的评价指标的变化幅度

## □ 仅添加类表示的表述排序模型

Model	English $F_1$	Chinese $F_1$
Full Model	65.52	64.41
– MENTION	−1.27	−0.74
– GENRE	−0.25	−2.91
– DISTANCE	−2.42	−2.41
– SPEAKER	−1.26	−0.93
– MATCHING	−2.07	−3.44

Table 1: CoNLL  $F_1$  scores of the mention-ranking model on the dev sets without mention, document genre, distance, speaker, and string matching hand-engineered features.

## ▣ 表述排序模型和聚类模型

All-Pairs	Top-Pairs	English $F_1$	Chinese $F_1$
Yes	Yes	65.52	64.41
Yes	No	-0.36	-0.24
No	Yes	-0.54	-0.33
No	No	-3.58	-5.43

Table 2: CoNLL  $F_1$  scores of the mention-ranking model on the dev sets with different pretraining methods.

Model	English $F_1$	Chinese $F_1$
Full Model	66.01	64.86
– PRETRAINING	-5.01	-6.85
– EASY-FIRST	-0.15	-0.12
– L2S	-0.32	-0.25

Table 3: CoNLL  $F_1$  scores of the cluster-ranking model on the dev sets with various ablations.

## □ 比较

	MUC			B <sup>3</sup>			CEAF <sub>ϕ<sub>4</sub></sub>			Avg. F <sub>1</sub>
	Prec.	Rec.	F <sub>1</sub>	Prec.	Rec.	F <sub>1</sub>	Prec.	Rec.	F <sub>1</sub>	
CoNLL 2012 English Test Data										
Clark and Manning (2015)	76.12	69.38	72.59	65.64	56.01	60.44	59.44	52.98	56.02	63.02
Peng et al. (2015)	–	–	72.22	–	–	60.50	–	–	56.37	63.03
Wiseman et al. (2015)	76.23	69.31	72.60	66.07	55.83	60.52	59.41	54.88	57.05	63.39
Wiseman et al. (2016)	77.49	69.75	73.42	66.83	56.95	61.50	62.14	53.85	57.70	64.21
NN Mention Ranker	79.77	69.10	74.05	69.68	56.37	62.32	63.02	53.59	57.92	64.76
NN Cluster Ranker	78.93	69.75	<b>74.06</b>	70.08	56.98	<b>62.86</b>	62.48	55.82	<b>58.96</b>	<b>65.29</b>
CoNLL 2012 Chinese Test Data										
Chen & Ng (2012)	64.69	59.92	62.21	60.26	51.76	55.69	51.61	58.84	54.99	57.63
Björkelund & Kuhn (2014)	69.39	62.57	65.80	61.64	53.87	57.49	59.33	54.65	56.89	60.06
NN Mention Ranker	72.53	65.72	68.96	65.49	56.87	60.88	61.93	57.11	59.42	63.09
NN Cluster Ranker	73.85	65.42	<b>69.38</b>	67.53	56.41	<b>61.47</b>	62.84	57.62	<b>60.12</b>	<b>63.66</b>

Table 5: Comparison with the current state-of-the-art approaches on the CoNLL 2012 test sets. NN Mention Ranker and NN Cluster Ranker are contributions of this work.

## □ 什么是共指消解?

## □ 表述检测

## □ 共指消解方法

### □ 规则方法

### □ 学习方法

- 表述对模型

- 表述排序模型

### □ 深度学习方法

- 表述排序

- 表述聚类排序

- 记忆网络(零指消解)



- ❑ **Convese, 2006**
- ❑ **Yeh and Chen, 2007**
- ❑ **Zhao & Ng 2007**
- ❑ **Kong & Zhou 2010**
- ❑ **Chen & Ng 2013**
- ❑ **Chen & Ng 2014**
- ❑ **Chen & Ng 2015**
- ❑ **Chen & Ng 2016**
- ❑ **Yin et al 2017**

- **当前SOTA**
- **Yin et al., EMNLP 2017, 刘挺老师组**
- **基于memory network**
  - 候选先行词作为记忆
  - 零指代作为查询
  - 注意力权值表达候选先行词作为零指先行词的概率



## □ 零指代

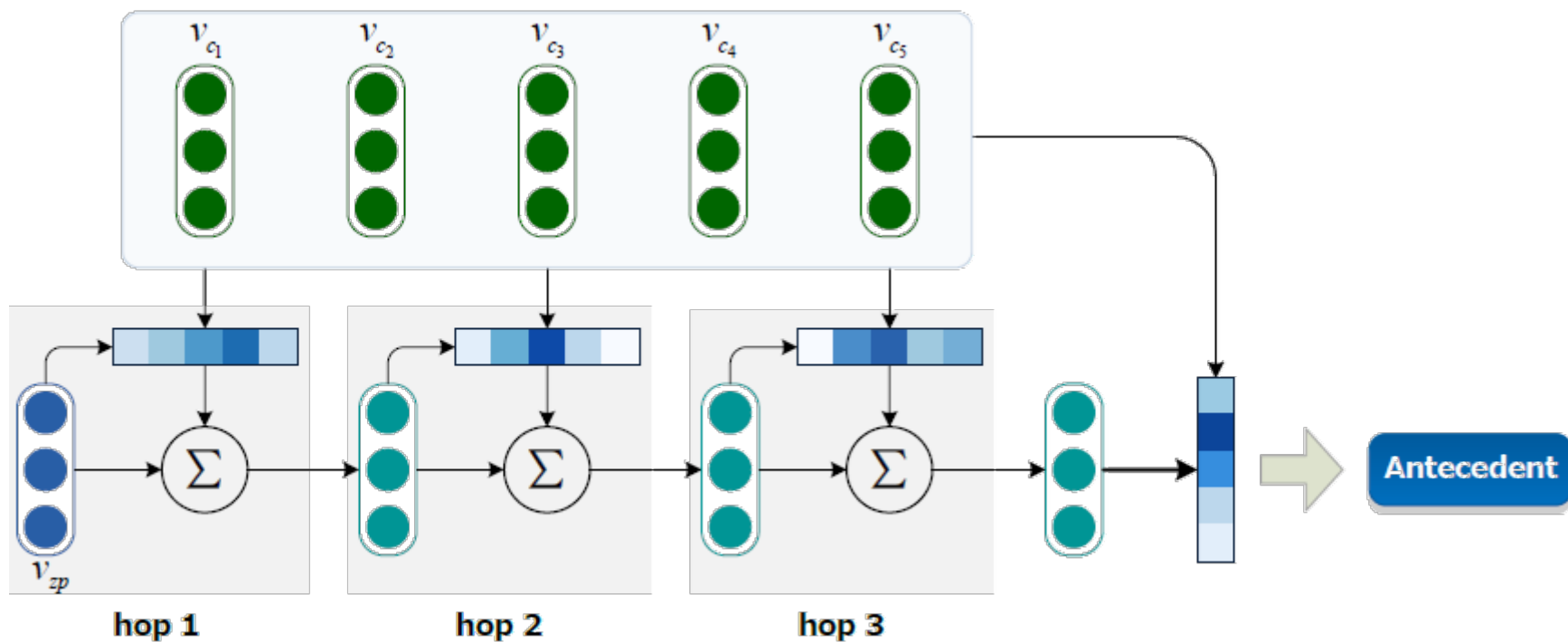
- 中文中非常常见
- 在GB理论中为空语类中的pro类型
- CTB中标出了小pro，有时标出了co-index
- OntoNotes标出了零指及其所指

[警方] 表示 他们 自杀 的 可能性 很高， 不过  $\phi_1$  也不排除  $\phi_2$  有 他杀 的 可能。

## □ 可消解的零指代一般为零形回指(anaphoric zero pronoun, AZP)

## 记忆网络

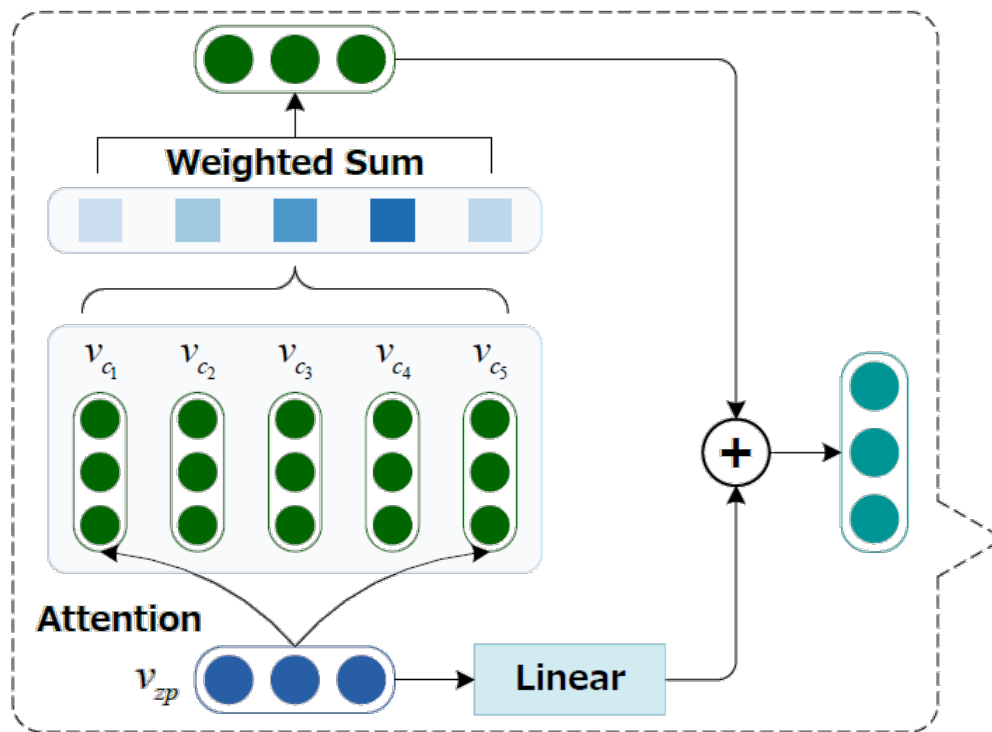
- 候选先行词作为记忆，零指作为查询



## 记忆网络每一跳

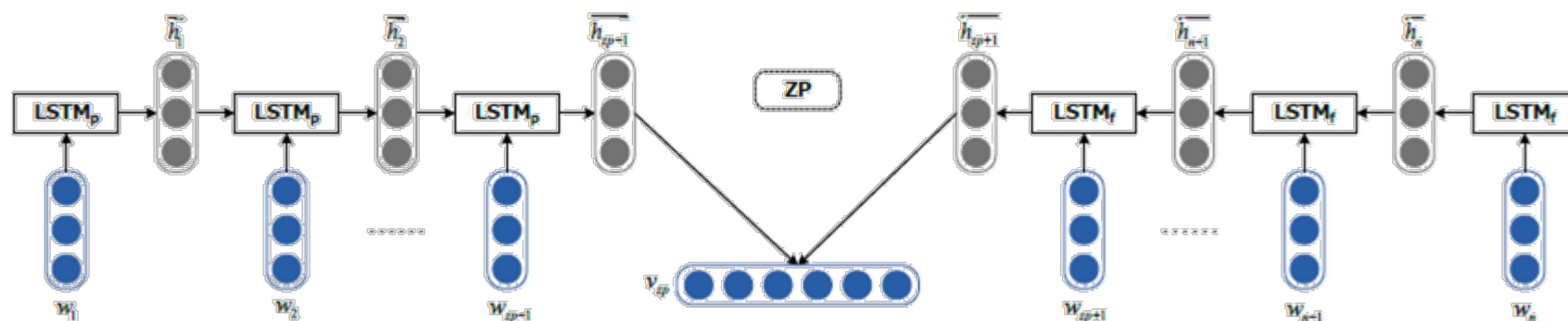
- 注意力机制
- Query: 零指+候选+特征
- Key: 权值矩阵
- Value: 候选

## 共6跳



## □ 零指表示

### □ 零指中心LSTM



## □ 如何检测是否是零指?

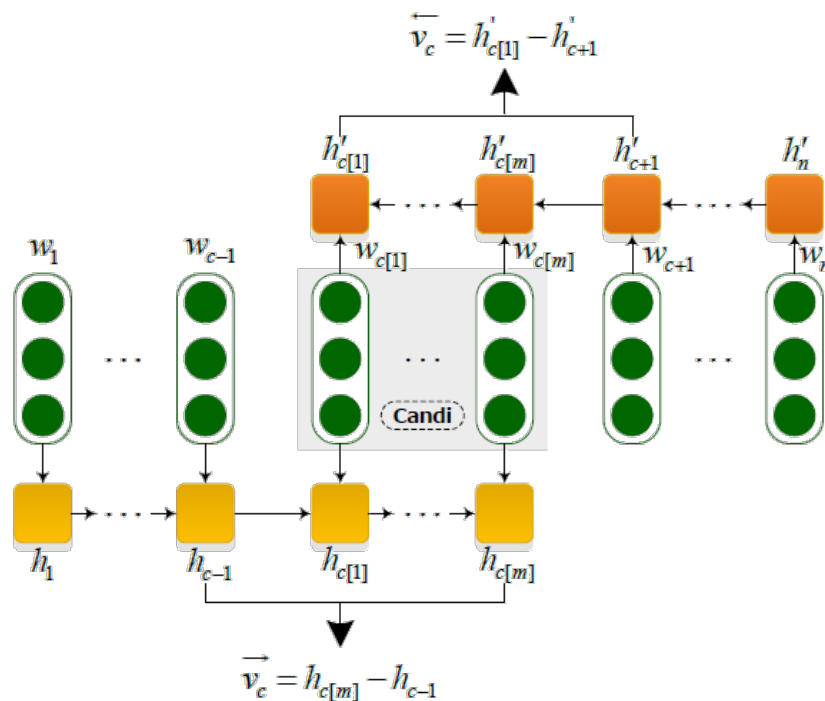
### □ 简单的基于特征的SVM分类器

## □ 候选先行词表示

- 最早来自Pei et al

## □ 如何获得候选先行词

- 规则筛



## ■ SOTA by a huge margin

	Setting 1			Setting 2			Setting 3		
	Gold Parse + Gold AZP			Gold Parse + System AZP			System Parse + System AZP		
	R	P	F	R	P	F	R	P	F
Zhao and Ng (2007)	41.5	41.5	41.5	22.4	24.4	23.3	12.7	14.2	13.4
Chen and Ng (2015)	50.0	50.4	50.2	35.7	26.2	30.3	19.6	15.5	17.3
Chen and Ng (2016)	51.8	52.5	52.2	<b>39.6</b>	27.0	32.1	21.9	15.8	18.4
ZPMN (1)	53.0	53.3	53.1	37.9	30.0	33.4	27.8	17.4	21.4
ZPMN (2)	53.7	54.0	53.9	38.8	30.6	34.0	28.1	18.2	22.1
ZPMN (3)	53.9	54.2	54.1	38.6	30.4	34.2	28.2	17.7	21.7
ZPMN (4)	54.4	54.7	54.5	39.0	30.7	34.3	<b>29.3</b>	<b>18.5</b>	<b>22.7</b>
ZPMN (5)	54.1	54.4	54.3	38.8	30.6	34.2	28.6	17.8	22.0
ZPMN (6)	<b>54.8</b>	<b>55.1</b>	<b>54.9</b>	39.4	<b>31.1</b>	<b>34.8</b>	28.9	18.2	22.3

## ■ 使用LSTM建模词在上下文中的含义是否有效？

- ZPContextFree: 使用governing verb和preceding word表示AZP
- AntContentAvg: 候选先行词使用词向量平均表示
- AntContentHead: 候选先行词使用头词向量表示

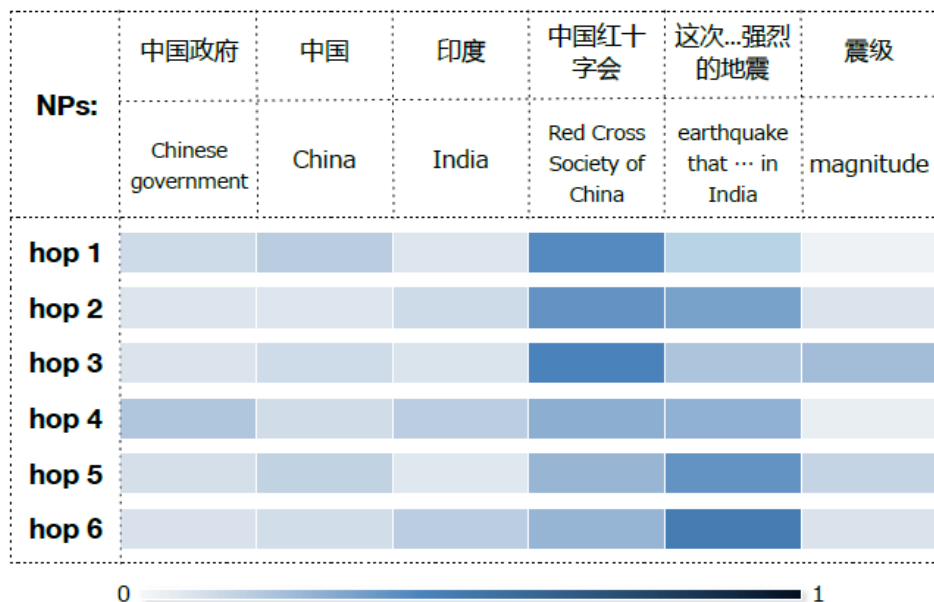
	R	P	F
ZPContextFree	53.5	53.8	53.6
AntContentAvg	52.6	52.9	52.7
AntContentHead	53.8	54.1	53.9
ZPMN (hop 6)	54.8	55.1	<b>54.9</b>

## □ 注意力可视化

## □ 多跳模拟了推理过程

这次近50年来印度发生的最强烈地震震级强， $\phi$ 波及范围广，印度邻国如尼泊尔也受到了影响。

The earthquake that is the strongest one occurs in India within recent 50 years has a high-magnitude,  $\phi$  influences a large range of areas, and the neighbouring country of India like Nepal is also affected.





## □ 共指消解

- 有用
- 有挑战性
- 有趣

## □ 多种类型的系统

- 效果逐渐提升
  - 很大程度上归功于神经模型
- 但整体上，结果仍需提高

**THANKS!**