Solving Hard Coreference Problems

Cognitive Computation Group, University of Illinois

Problem Description

Standard Coreference Instances

[John] adores [Mary] because [she] is pretty.
[Prof. Roth] is satisfied with his [students] because [they] work hard.

Problems with Existing Coref Systems

Rely heavily on gender / plurality information

Hard Coreference Instances

[A bird] perched on the [limb] and [it] bent.
[A bird] perched on the [limb] and [it] sang.

Goal

A better overall coreference system

- Improve on solving hard coreference problems
- Maintain the state-of-art performance on standard coreference problems

Predicate Schemas

Type 1 Schema

$$pred_m(m, a) \qquad m\text{: Sub } a\text{: Obj}$$

[The bee] landed on [the flower] because [it] had pollen.

S(*have*(m=*[the flower]*, a=*[pollen]*)) > S(*have*(m=*[the bee]*, a=*[pollen]*))

Type 2 Schema

$$pred_m(m, a) | \widehat{pred}_m(m, \hat{a}), cn$$

[Jim] was afraid of [Robert] because [he] gets scared around new people.

S(*be afraid of* (m=*, a=*) | *get scared around* (m=*, a=*), *because*) >
S(*be afraid of* (a=*, m=*) | *get scared around* (m=*, a=*), *because*)

Schema Variations

Type 1	$\mathcal{S}(pred_m(m,a))$ $\mathcal{S}(pred_m(a,m))$ $\mathcal{S}(pred_m(m,*))$ $\mathcal{S}(pred_m(*,m))$
Type 2	$\mathcal{S}(pred_m(m,a) \vert \widehat{pred}_m(m,\hat{a}), cn)$ $\mathcal{S}(pred_m(a,m) \vert \widehat{pred}_m(m,\hat{a}), cn)$ $\mathcal{S}(pred_m(m,a) \vert \widehat{pred}_m(\hat{a},m), cn)$ $\mathcal{S}(pred_m(a,m) \vert \widehat{pred}_m(\hat{a},m), cn)$ $\mathcal{S}(pred_m(m,*) \vert \widehat{pred}_m(m,*), cn)$ ⋮

Example Beyond Above Schemas

[Lakshman] asked [Vivan] to get him some ice cream because [he] was hot.

Utilizing Knowledge

Knowledge as Features

$$w_{u,v} = \mathbf{w}^T\phi(u,v) + \tilde{\mathbf{w}}^T\mathbf{s}(u,v)$$

Pairwise Mention Scoring Function

Scoring Function for Predicate Schemas

- Noise in Knowledge
- Inexplicit Textual Inference

Knowledge as Constraints

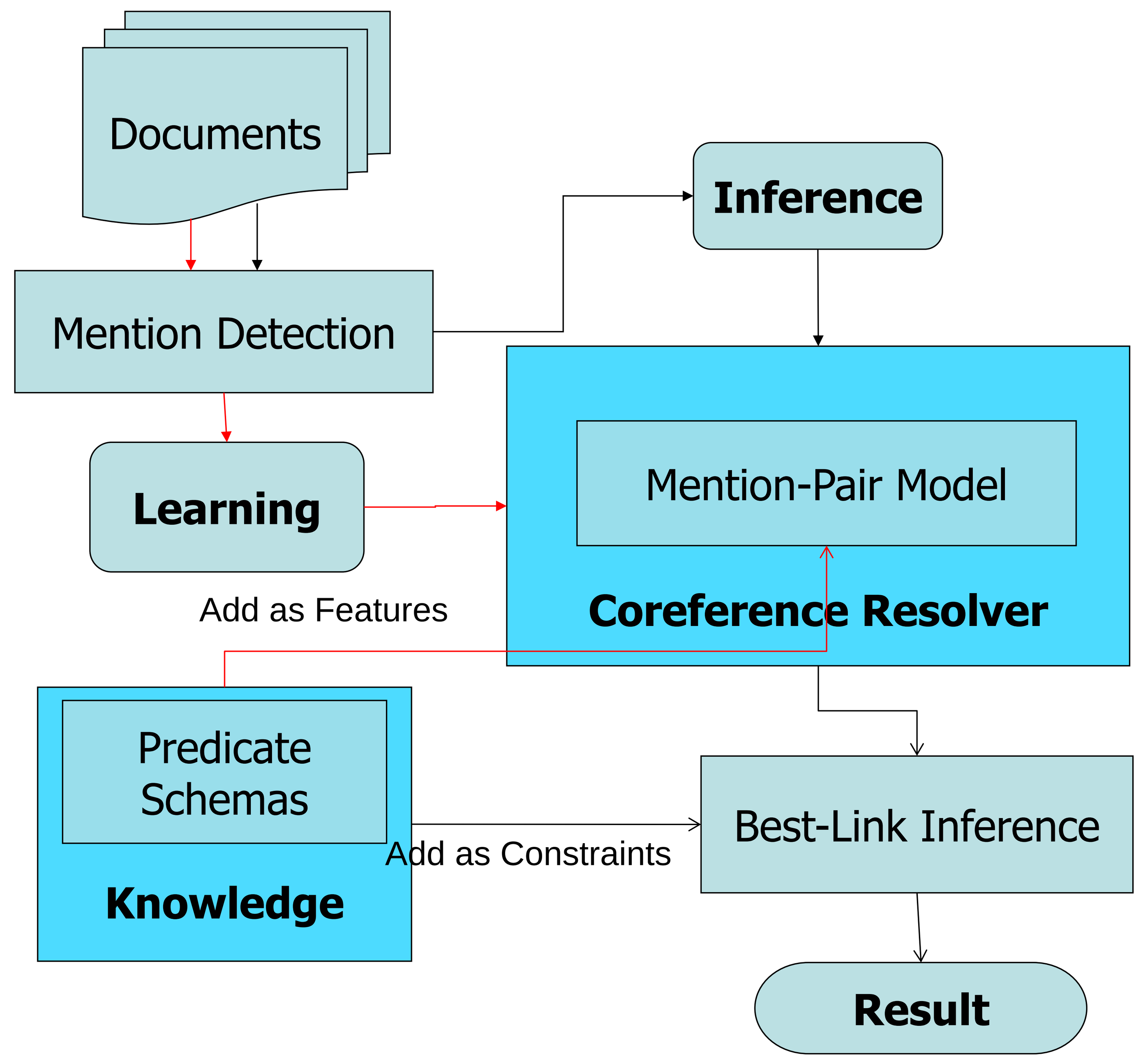
Generating Constraints

$$\begin{cases} \text{if } s_i(u,v) \geq \alpha_i s_i(w,v) \Rightarrow y_{u,v} \geq y_{w,v}, \\ \text{if } s_i(u,v) \geq s_i(w,v) + \beta_i \Rightarrow y_{u,v} \geq y_{w,v} \end{cases}$$

ILP inference (Best-Link)

$$\arg\max_y \sum_{u,v} w_{uv} y_{uv}$$
$$\text{s.t. } \sum_{u<v} y_{uv} \leq 1, \forall v$$
$$y_{uv} \in \{0,1\}, \forall u, v$$
$$\text{if } s_{uv} \geq t + s_{um} \text{ then } y_{uv} \geq y_{um}$$
$$\text{if } s_{uv} \geq t' \cdot s_{um} \text{ then } y_{uv} \geq y_{um}$$

Learning and Inference Framework



Knowledge Acquisition

Gigaword co-occurrences

Extract triples from Gigaword

$$\mathcal{S}_{giga} = [\mathcal{S}^{(1)}_{giga}\ \mathcal{S}^{(2)}_{giga}]$$

$$\mathcal{S}^{(1)}_{giga}(u,v) \equiv \mathcal{S}(pred_v(m = u, a = a_v))$$
$$\mathcal{S}^{(2)}_{giga}(u,v) \equiv \mathcal{S}(pred_u(m = u, a = a_u) | \widehat{pred}_v(m = u, a = a_v), cn).$$

Wikipedia Disambiguated Co-occurrences

- Extract **disambiguated** noun, verbs and entities, etc. in Wikipedia (Illinois-Wikifier)
- Collect co-occurrence statistics: 1) immediately after 2) immediately before 3) before 4) after

$$\mathcal{S}(u,v)_{wiki} \equiv \mathcal{S}(pred_v(m = u, a = a_v)).$$

Web Search Statistics

Generate queries to get the score

$$1)\ \text{“}u\ a_v\text{”} \quad 2)\ \text{“}u\ pred_v\text{”} \quad 3)\ \text{“}u\ pred_v\ a_v\text{”} \quad 4)\ \text{“}a_v\ u\text{”}$$

$$\mathcal{S}_{web}(u,v) \equiv \mathcal{S}(pred_v(m = u, a = a_v))$$

Polarity Information

- Initialize $Po(pred_u)$ and $Po(pred_v)$ (Wilson et al, 2005) $\mathcal{S}_{pol}(u,v) \equiv \mathcal{S}(pred_v(m = u, a = a_v))$
- Negate polarity when mention role is *object*
- If there is a polarity reversing connective (such as “but”) preceding the predicate, reverse the polarity.

$$s_{pol}(u,v) = \begin{bmatrix} \mathbf{1}\{Po(p_u) = + \text{ AND } Po(p_v) = +\} \text{ OR } \mathbf{1}\{Po(p_u) = - \text{ AND } Po(p_v) = -\} \\ \mathbf{1}\{Po(p_u) = + \text{ AND } Po(p_v) = +\} \\ \mathbf{1}\{Po(p_u) = - \text{ AND } Po(p_v) = -\} \end{bmatrix}$$

Results

Datasets:

- Winograd (Rahman&Ng, 2012)
- Winocoref: Winograd with more mentions
- Standard Coref: ACE-2004, Ontonotes

Metrics:

- **Accuracy:** for “Winograd” (binary classification)
- MUC, BCUB, CEAF (general coref)
- **AntePre:**
  - k pronouns in sentence
  - Each pronoun has n_k antecedents
  - m correct binary decisions

$$AntePre = \frac{m}{\sum_{i=1}^{k} n_k}$$

Evaluations:

- Hard Coreference Problems

Dataset	Metric	Illinois	Rahman and Ng (2012)
Winograd	Precision	51.48	73.05
WinoCoref	AntePre	68.37	—

Dataset	KnowFeat	KnowCons	KnowComb
Winograd	71.81	74.93	**76.41**
WinoCoref	88.48	88.95	**89.32**

- Standard Coreference Problems

System	MUC	BCUB	CEAFe	AVG
ACE				
IlliCons	**78.17**	81.64	**78.45**	**79.42**
KnowComb	77.51	**81.97**	77.44	78.97
OntoNotes				
IlliCons	84.10	**78.30**	**68.74**	**77.05**
KnowComb	**84.33**	78.02	67.95	76.76

- Ablation Study
  - We categorized instances in Winograd data

Category	Cat1	Cat2	Cat3
Size	317	1060	509
Portion	16.8%	56.2%	27.0%

  - Evaluated on each category:

Schema	AntePre(Test)	AntePre(Train)
Type 1	76.67	86.79
Type 2	79.55	88.86
Type 1 (Cat1)	90.26	93.64
Type 2 (Cat2)	83.38	92.49

COGNITIVE COMPUTATION GROUP UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Haoruo Peng, Daniel Khashabi and Dan Roth