CS 4248
Natural Language Processing

Professor NG Hwee Tou
Department of Computer Science
School of Computing
National University of Singapore
nght@comp.nus.edu.sg

Discourse

- Discourse: A coherent, structured group of sentences
  - Monologue: Communication flows in only one direction, from the speaker to the hearer
  - Dialogue: Each participant periodically takes turns being a speaker and hearer

Reference Resolution

Victoria Chen, Chief Financial Officer of Megabucks Banking Corp since 2004, saw her pay jump 20%, to 1.3 million, as the 37-year-old also became the Denver-based financial-services company's president. It has been ten years since she came to Megabucks from rival Lotsabucks.

Reference Resolution

- Reference: Use of linguistic expressions to denote an entity
- Reference resolution: Determining what entities are referred to by which linguistic expressions
- Referring expression: A natural language expression used to perform reference
- Referent: The entity referred to

Reference Resolution

- Two referring expressions *corefer* when they are used to refer to the same entity
- The first of the two coreferring expressions is the *antecedent* and the second the *anaphor*
- Anaphora: Reference to an entity that has been previously introduced into the discourse

Reference Resolution

- Coreference resolution
  - Finding referring expressions that refer to the same entity (finding expressions that corefer)
- Coreference chain: Set of coreferring expressions

Complexity of Coreference Resolution

- *John* went to the bar with Mike. *He* called for a glass of wine.
- *Mike* went to the bar with John. *He* called for a glass of wine.
- John has a BMW. He is attractive.
- John has a BMW. It is attractive.

Complexity of Coreference Resolution

- John parked his *car* in the garage after driving *it* around for hours.
- *John* telephoned Bill. *He* lost the laptop.
- John criticized *Bill*. *He* lost the laptop.

Neural Coreference Resolution

- Materials taken from “End-to-end neural coreference resolution”, Kenton Lee, Luheng He, Mike Lewis, Luke Zettlemoyer, EMNLP 2017 and associated presentation slides

Neural Coreference Resolution

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.

Cluster 1:

A fire

the blaze

Neural Coreference Resolution

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.

Cluster 2:

a Bangladeshi garment factory

the four-story building

Neural Coreference Resolution

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.

Cluster 3:

at least 37 people

the deceased

Neural Coreference Resolution

- Two subtasks:
  - Mention detection
    - 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.
  - Mention clustering
    - Cluster 1: A fire, the blaze
    - Cluster 2: a Bangladeshi garment factory, the four-story building
    - Cluster 3: at least 37 people, the deceased

Neural Coreference Resolution

- Previous work:
  - A pipelined approach: first mention detection, then mention clustering
  - Rely on a syntactic parser to detect the mentions (noun phrases) and head words, and to determine the features of the head words
  - Try all pairs of mentions to determine mention clusters
- Drawbacks:
  - Errors propagate from mention detection to mention clustering

Neural Coreference Resolution

- Contributions of this approach:
  - Joint mention detection and clustering
  - No pre-processing and not using a syntactic parser
  - A neural end-to-end approach

Neural Coreference Resolution

- A document D containing T words
- N = number of possible text spans in D
- $N = \binom{T+1}{2} = \frac{(T+1)T}{2}$
- $O(T^2)$ spans, $O(T^4)$ pairwise decisions
- Span i $(1 \leq i \leq N)$ has start index $\mathrm{START}(i)$ and end index $\mathrm{END}(i)$

Neural Coreference Resolution

- Consider all possible spans
- Learn to rank antecedent spans
- Aggressive pruning of search space

Neural Coreference Resolution

- Task: Assign to each span i an antecedent $y_i \in \mathcal{Y}(i)$
- $y_i = \epsilon$ (dummy antecedent) represents two possible scenarios:
  - Span i is not an entity mention; or
  - Span i is an entity mention but it is not coreferent with any previous span

Neural Coreference Resolution

Span representation $\boldsymbol{g_i}$ for span i

- Input sequence of word embeddings: $\boldsymbol{x_{1:T}}$
- Output vector at position t: $\text{biLSTM}(\boldsymbol{x_{1:T}}, t) = \boldsymbol{x_t^*} = \left[\text{LSTM}^f(\boldsymbol{x_{1:t}}); \text{LSTM}^b(\boldsymbol{x_{T:t}})\right]$
- Attention mechanism to represent head word of the span
- $\alpha_t = \text{MLP}(\boldsymbol{x_t^*}) \cdot \boldsymbol{w_\alpha^T}$
- $a_{i,t} = \frac{\exp(\alpha_t)}{\sum_{k=\text{START}(i)}^{\text{END}(i)} \exp(\alpha_k)}$
- $\hat{\boldsymbol{x}}_i = \sum_{t=\text{START}(i)}^{\text{END}(i)} a_{i,t} \cdot \boldsymbol{x}_t$
- $\phi(i)$ encodes the length of span i
- $\boldsymbol{g_i} = [\boldsymbol{x}^*_{\text{START}(i)}; \boldsymbol{x}^*_{\text{END}(i)}; \hat{\boldsymbol{x}}_i; \phi(i)]$

Neural Coreference Resolution

- Pairwise score $s(i,j)$ for a coreference link between span i and span j, where j is an antecedent of the anaphor i
- $s_m(i) = MLP(\boldsymbol{g_i}) \cdot \boldsymbol{w_m^T}$ (span i is a mention)
- $s_a(i,j) = MLP([\boldsymbol{g_i}; \boldsymbol{g_j}; \boldsymbol{g_i} \circ \boldsymbol{g_j}; \psi(i,j)]) \cdot \boldsymbol{w_a^T}$ (span j is an antecedent of span i), $\circ$ denotes element-wise multiplication of two vectors
- $\psi(i,j)$ encodes the distance between the two spans
- $s(i,j) = \begin{cases} 0 & j = \epsilon \\ s_m(i) + s_m(j) + s_a(i,j) & j \neq \epsilon \end{cases}$

Neural Coreference Resolution

- N spans in a document D
- $P(y_1, \ldots, y_N|D) = \prod_{i=1}^{N} P(y_i|D) = \prod_{i=1}^{N} \frac{\exp(s(i,y_i)))}{\sum_{y' \in \mathcal{Y}(i)} \exp(s(i,y')))}$
- y_i: antecedent for span i
- $\mathcal{Y}(i)$: Set of possible assignments for y_i

Neural Coreference Resolution

- Learning: minimize

$$-\log \prod_{i=1}^{N} \sum_{\hat{y} \in \mathcal{Y}(i) \cap GOLD(i)} P(\hat{y}|D)$$

- $GOLD(i)$: set of spans in the gold clustering containing span i

Neural Coreference Resolution

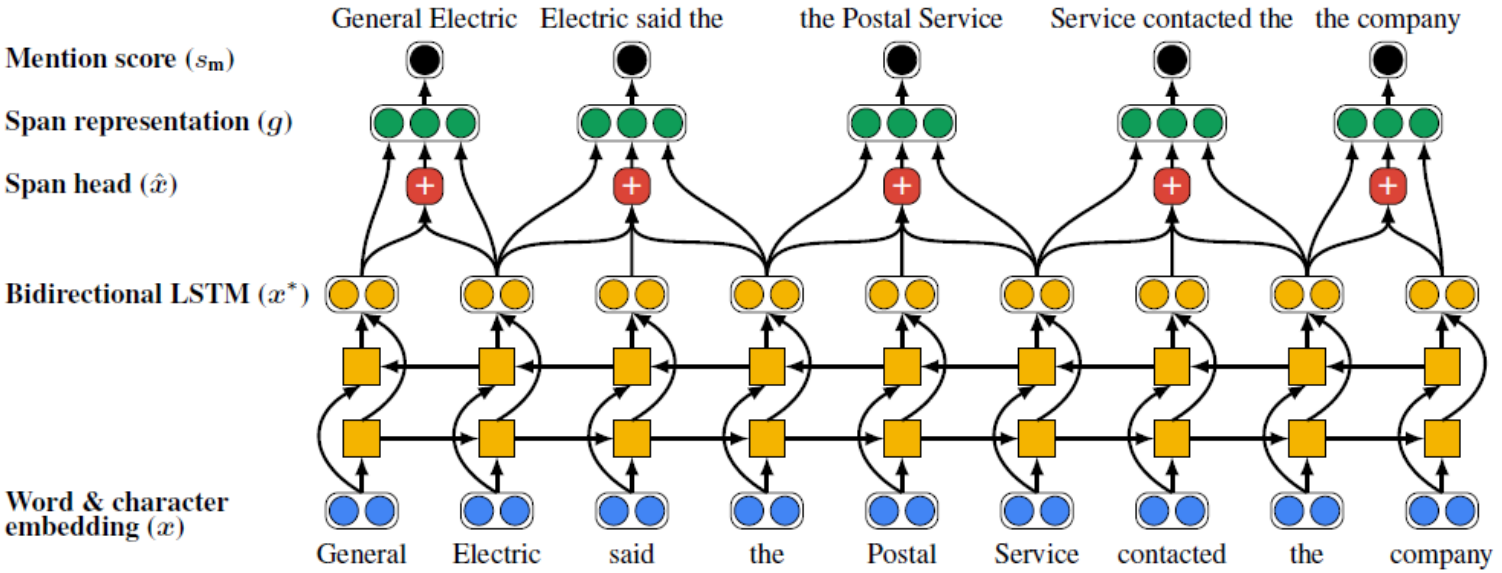


Diagram from “End-to-end neural coreference resolution”, Kenton Lee, Luheng He, Mike Lewis, Luke Zettlemoyer, EMNLP 2017

Neural Coreference Resolution

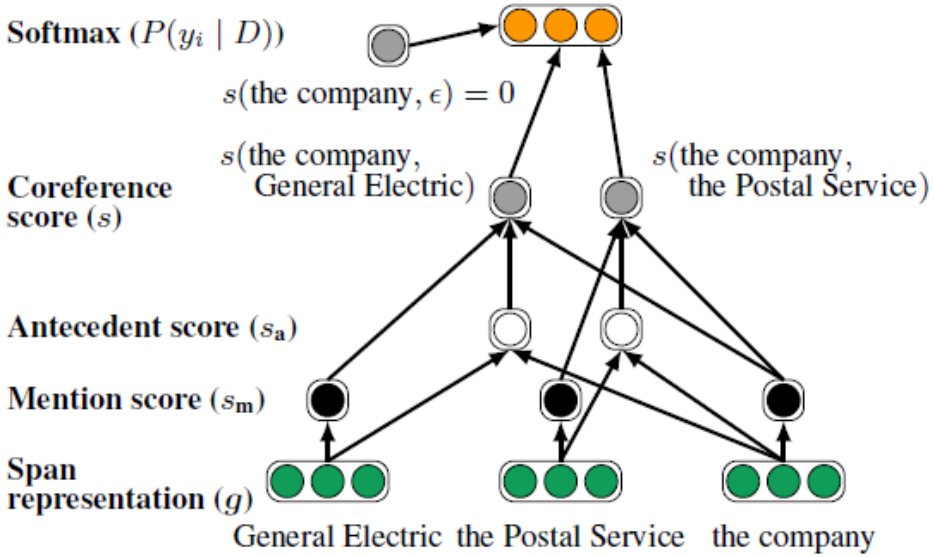


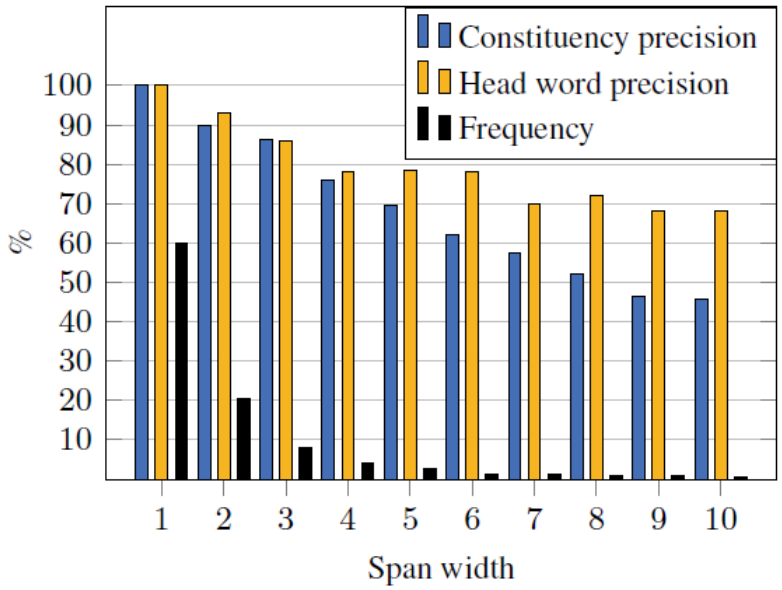
Diagram from “End-to-end neural coreference resolution”, Kenton Lee, Luheng He, Mike Lewis, Luke Zettlemoyer, EMNLP 2017

Neural Coreference Resolution

- Pruning: To maintain efficiency
  - Consider spans of up to L words ($L = 10$)
  - Keep up to λT spans ($\lambda = 0.4$) with the highest mention scores $s_m(i)$
  - Consider up to K antecedents ($K = 250$) for each span
  - Still maintain a high recall of 92.7% of gold mentions
- Achieves average F1 score of 68.8% for coreference resolution

Neural Coreference Resolution

- Spans identified correspond to syntactic constituents
- Attention mechanism effectively locates head words (the word with the highest attention weight $a_{i,t}$ in a span)



Neural Coreference Resolution

Coreference chains output by system:

1	(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.
	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).
2	We are looking for (a region of central Italy bordering the Adriatic Sea). (The area) is mostly mountainous and includes Mt. Corno, the highest peak of the Apennines. (It) also includes a lot of sheep, good clean-living, healthy sheep, and an Italian entrepreneur has an idea about how to make a little money of them.
3	(The flight attendants) have until 6:00 today to ratify labor concessions. (The pilots') union and ground crew did so yesterday.
4	(Prince Charles and his new wife Camilla) have jumped across the pond and are touring the United States making (their) first stop today in New York. It's Charles' first opportunity to showcase his new wife, but few Americans seem to care. Here's Jeanie Mowth. What a difference two decades make. (Charles and Diana) visited a JC Penney's on the prince's last official US tour. Twenty years later here's the prince with his new wife.
5	Also such location devices, (some ships) have smoke floats (they) can toss out so the man overboard will be able to use smoke signals as a way of trying to, let the rescuer locate (them).