

Lecture 16: Coreference Resolution

Shubham Gupta

March 31, 2020

1 Introduction

- Identify all mentions that refer to the same real world entity
- When one word refers to two or more entities, it is called **split antecedent**. No system can deal with these words. Eg: A and B went out. *They* are retarded.
- Coref resolution helps in:
 - Full text understanding
 - Machine translation
 - Dialogue systems
- Steps:
 - Detect the mentions(easy)
 - Cluster the mentions(hard) aka coreference

2 Mention Detection

- Span of text referring to some entity
 - Pronouns: I, your, it, she him. Use POS tagger
 - Named entities: people places, Use NER
 - Noun phrases: a dog, cat stuck in tree. Use parser(constituency parser)
- Marking all pronouns, NE and NP over-generates mentions
- Solution: Train classifier to filter spurious mentions.
- Solution 2: Collect all mentions as "candidate mentions". Discard mentions that have not been marked as coreference with any other word.

3 Linguistics

- *Anaphora*: One term(anaphor) refers to another term(antecedent). Interpretation for anaphor dependent on interpretation of antecedent
- Obama said he would sign the bill.
- Obama: Antecedant. he: anaphor
- Not all anaphoric relations are coreferential
 - *Every Dancer* twisted *her* knee.
 - *No Dancer* twisted *her* knee.
- Hobbs Algorithm. Naive algorithm to do coreference. Baseline for coreference resolution.

3.1 Knowledge based Pronominal coreference

- "IT" can refer to different entities as follows:
 - 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.
- Hobbs algorithm fails for the above case.
- These kind of sentences are called **Winograd Schema**. Named after scientist Henry Winograd who found these sentences.
- Can be used as good alternative to Turing test.

4 Coreference Models: Mention Pair

- Take pairs of mentions and train binary classifier to classify if they are coreferent or non-coreferent.
- Train with cross entropy loss
 - Score pairs of words
 - Pick threshold. Add coreference links when the score is above threshold.
 - Take transitive closure to get the final clustering
- Not the best way to do coref. One bad relationship can lead to collapse.
- Many mentions only have one clear antecedent
- **Solution**: train the model to predict only one antecedent for each mention

5 Mention Ranking

- Assign each mention its highest scoring candidate antecedent acc to model
- Add NA mention to start of sentence to decline linking current mention to anything.

6 Neural Coref Model

- Standard feed-forward NN
- Input: word embeddings and categorical features
- Embeddings:
 - Previous two words, first word, last word, head word. etc
- Other features
 - Distance
 - Document genre
 - Speaker info

7 Current SOTA

- Mention ranking model
- Improvements
 - Use LSTM
 - Use attention
 - Do mention detection and coref end-to-end

8 Clustering Based

- Use agglomerative clustering
- Each mention will have it's own cluster
- Use model to score which cluster merges will be good
- Mention pairs $=_i$ Mention pair representations $=_i$ Cluster Pair Rep $=_i$ Score

9 Evaluation

- Metrics: MUC, CEaF, LEA, B-CUBED, BLANC
- Report average over a range of metrics