Lecture 16: Coreference Resolution

Shubham Gupta

March 30, 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