

Reference and Anaphora in Dialog

CEARA CHEWNING

Overview

- Summary of the main topic for this group of papers
- Why it's interesting
- Overview of the primary paper
- Overview of the (more interesting) secondary paper
- Overview of the (less interesting) secondary paper
- Ways to apply this to CompLing/ SDSs
- Questions

So, what are we talking about here?

- A dialog generates a history that informs the interpretation and generation of future utterances
- There are different criteria that determine what history is most salient to the current utterance
- We want to understand what effect past utterances and actions have on the current situation

Okay, now why do I care?

- This is a tool to help us with our old friend- Ambiguity
- Dialog history is a great resource we can use to disambiguate utterances that the user produces
- We can also use history to produce an utterance that is more likely to be understood and accepted by the user

The Primary Paper – Interpretation using dialog history (DeVault and Stone, 2009)

- This paper explored a method of keeping track of different threads of possible interpretations of an utterance
- The goal was to be able to use these different possible threads and requests for clarification to produce more labelled data to train the system to better interpret future utterances
- This is proposed as an alternative to using annotated data- the user knows which interpretation of their utterance is correct- why not ask them?

Primary Paper (DeVault and Stone, 2009), cont.

- The experiment had users interact with the system to identify a specific candidate object to add to a sequence of shapes. Need to:
 - Identify object by color, shape, and whether it is solid
 - Tell director that they've selected the object
 - Decide whether to move on to the next object
- System needs to infer tacit actions, and come up with a set of possible interpretations of the utterance, and their probabilities
- System then asks question to try to resolve the ambiguity, and either culls unlikely threads, kills incorrect threads, or accepts correct threads

Primary Paper (DeVault and Stone, 2009), cont

- The system then harvests training data from the ultimate fate of all these possible threads of interpretations, and trains a MaxEnt model with features from both the observed features of the utterance, and assumptions from the possible interpretation.
- After testing with held out dialogs, the learned model correctly resolves 82% of ambiguous utterances vs the baseline of 21%. Cool!

Wait, what about _____ ? Problems in DeVault and Stone

- Small sample size: only 318 ambiguous events to train on.
- Interesting way to test it- for each subject they train on all other subject's utterances, and test with that one's utterances.
- Not very clear about what constituted an observation "o". Early in the paper "o" is defined as an utterance, but in feature table they say "if o is an utterance...". What's the other option?

Secondary- Conceptual Pacts (Brennan and Clark, 1996)

- This paper focused on human- human interactions, and more on generation than interpretation.
- Main question: What are the rule governing what words we use to refer to an object?
- Assumptions- no true synonyms, word choice implies a particular concept of the object in question

What is this? (Brennan and Clark, 1996)



Which one? (Brennan and Clark, 1996)



Which one? (Brennan and Clark, 1996)



Conceptual Pacts (Brennan and Clark, 1996)

- “Make your contribution as informative as required... Do not make your contribution more informative than required.” Not quite.
- Recency, Frequency of Use, Provisionality, Partner Specificity also play a role in what terms a speaker uses
- Interlocutors cooperatively negotiate a term to agree on – conceptual pact
- Experiment 1: test recency and frequency of use
- Experiment 2: test provisionality
- Experiment 3: test partner specificity

Wait, what about _____? Problems in Brennan and Clark

- Inconsistent annotation and other aspects of the experiments. Why?
- Do the pictures in trial A represent a stereotypical example of the general class?
- How do they determine what a general term is?

Other Secondary Paper – Bossy Robots (Giuliani et al, 2010)

- A robot tells you how to build something- compare one with static references to one with adaptive references
- Stores a history of referring expressions from itself and the human. Looks at past references to an object to determine how to refer to it in this utterance- deictic, indefinite, definite, pronoun, etc.
- Users rate the robot on intelligence, task ease and success, feeling of user, and conversation quality
- Only real difference in conversation quality- better with adaptive references

What can I do with this?

- Use history as in the secondary papers to better establish a connection with the user by choosing references that they are likely to expect and understand.
- Use history to disambiguate utterance interpretations

Questions

- How would you expand this to more open tasks?
- Is the effect of making the user more comfortable worth the engineering effort? Is sounding human always the goal?
- More?