Resolving “You” in Multi-Party Dialog

Surabhi Gupta
surabhi@cs.stanford.edu

John Nickrasz
nickrasz@stanford.edu

Matthew Purver
mpurver@stanford.edu

Dan Jurafsky
jurafsky@stanford.edu

The Problem:
Imagine a decision: “OK, so you will send me that by next week.”
• Who does “you” refer to?
• Does it even refer to anyone?

Conclusions:
• Both problems are real: only 58% are referential, 37% to next speaker
• We can detect referentiality reasonably well (75% accuracy).
• We can beat our reference resolution baseline, but 47% accuracy still isn’t great. Video features (e.g. gaze directions) should help.

Classes of “you”

Generic
(references to nobody in particular):
B: Well, usually what you do is just wait until you think it’s stopped, and then you patch them up.
A: um if you can get that binding point also maybe with a nice example that would be helpful for John and me.
B: Oh yeah uh O K.

Reference Singular
(references to one interlocutor):
A: and um if you can get
B: Well, usually what you do is just wait until you think it’s stopped, and then you patch them up.
A: um if you can get that binding point also maybe with a nice example that would be helpful for John and me.
B: Oh yeah uh O K.

Reference Plural
(references to more than one interlocutor):
A: So y- so you guys will send to the rest of us um a version of um this, and - the - uh, description -
B: With sugge- yeah, suggested improvements and -

Data Stats
AMI: multi-modal dataset of 4-party meetings
Kappa of .84 over a subset of 108 utterances tagged by two authors
Experiment 1: Referentiality Detection: 952 utterances for training: 374 for test
Experiment 2: Reference Resolution: 291 utterances for training; 176 for test

Experiment 1: Referentiality Detection
Problem: disambiguation of generic versus referential uses of “you”
Previous work: (Gupta et al., ACL 2007) on Switchboard (two-party dialog). Baseline (dominant class) = 54.6%; Accuracy = 84.4% (lexical + part-of-speech, dialog act, oracle context features)
Baseline: Predicting the dominant class (referential).
Method: Binary classification using a CRF classifier (Finkel et al., 2005) to model sequence (rather than using oracle context features); utterance-based lexical/syntactic/pragmatic features (Set 1).

Features

Accuracy
Baseline 57.9%
Sent + POS + QM 63.0%
Sent + POS + Q_DA 70.6%
DA 71.9%
Sent + POS + QM + DA 75.1%

* Automatically extracted features achieve an accuracy above the baseline
* Accuracy for AMI is lower than the 84.4% achieved for two-person data, suggesting that referentiality in multi-party meetings is a harder task.

Experiment 2: Reference Resolution
Problem: detect which participant is being addressed, using speech transcripts alone
Baseline: Next Speaker: always predict the addressee to be the next (different) speaker
Previous Speaker: always predict the addressee to be the previous (different) speaker
Method: treat problem as a 4-way classification problem for each “you” utterance:
1 = “you”; refers to the next different speaker
2 = “you” refers to the second different speaker
3 = “you” refers to the third different speaker
4 = “you” refers to the whole group.
(reference to non-singleton subsets turns out to be very rare)

Extract features based on each possible pair of (a) the “you” utterance and (b) the next or previous utterance spoken by each candidate addressee; try to capture similarities and relations between (a) and (b) (Set 2).

Results:

Features
Accuracy
Baseline: Previous Speaker 23.0%
Baseline: Next Speaker 37.0%
Set 1 + Set 2 47.2%

Set 2: forward looking and backward looking features for an utterance
Biggest confusion was found to be between utterances being classified as 1 or 4 (i.e. the next speaker or the entire group).

Next Steps:
• Discourse structure (e.g. adjacency pairs) investigation
• Video features are likely to help performance (Stiefelhagen, 2002; Jovanovic et al., 2006)