November 25, 2003

Chapter 19

Dialogue Systems

Overview

- Overview of dialogue systems

- Properties of dialogue

- Dialogue acts

- Interpreting dialogue acts

- (Discourse structure)

- An overly simple dialogue manager (FSA)

- A more adequate dialogue manager

- Demo: WITAS

What does a dialogue system do?

- (Speech recognition)

- Syntactic and semantic analysis

- Dialogue management

- $\Rightarrow$ content to express

- Generation

- (Speech synthesis)

What is dialogue management?

- Interpreting user's intent (including implicatures)

- Deciding when to speak

- Maintaining a list of things to express

- Deciding what kind of utterance is required/which item from the queue to actually express

- Deciding how to express it (which semantic form to send to generation)

Properties of dialogue

- Utterances (v. sentences)

- Turn-taking

- Common ground

- Implicature

Turn-taking

- People rarely overlap in speech, and rarely leave more than a few hundred milliseconds between turns.

- How do we figure out who gets to talk when?

- Detect transition-relevant places (typically utterance boundaries), then:

Turn-taking

- Turn-taking Rule (Sacks et al 1974)

  - If during this turn the current speaker has selected A as the next speaker, then A must speak next.

  - If the current speaker does not select the next speaker, any other speaker may take the next turn.

  - If no one else takes the next turn, the current speaker may take the next turn.

- Select the next speaker via adjacency pairs: QUESTION-ANSWER, GREETING-GREETING, COMPLIMENT-DOWNPLAYER, REQUEST-GRANT

Detecting utterance boundaries

- Cue words: *well, so, and*

- *N*-gram word or POS sequences

- Prosody: intonational phrases stand in some (complex) relation to utterances

Grounding

- Conversation can be seen as a process of developing and maintaining common ground: a set of beliefs mutually held and known to be held by both speakers.

- (cf. discourse model for potential pronoun referents)

- In order for a speaker's utterance to make it into the common ground, the hearer has to acknowledge it:
  - Continued attention
  - Relevant next contribution
  - Acknowledgment (nod, *uh huh*, *that's great, …*)
  - Demonstration
  - Display

Grounding

- Alternatively, hearers can contest something, including the presuppositions of an utterance:

  Have you stopped smoking?

- Or, hearers can request a repair or clarification.

- Does monologue involve common ground?

- How would you expect a computer to acknowledge an utterance you made?

Conversational Implicature

- A particular class of licensed inferences, licensed by Gricean maxims and the assumption that participants are being cooperative.

- Maxims:

  - Maxim of Quantity: Be exactly as informative as is required.

    I have one child. (and only one)

  - Maxim of Quality: Try to make your contribution one that is true.

    The world is flat. (and I believe it to be flat)

Conversational Implicature

- Maxims, cont:

  - Maxim of Relevance: Be relevant.
    Kim always turned in neat homework. (but that's the only good thing I can say about Kim)

  - Maxim of Manner: Be perspicuous.
    Kim emitted a sequence of sounds which approximated the lyrics and tune of "Happy Birthday". (but you couldn't really say it was singing)

- Does implicatures apply in monologue as well?

Dialogue acts, or: How to do things with words

- Locutionary v. illocutionary v. perlocutionary acts

- Illocutionary acts:

  - Assertives (*suggesting, putting forward, swearing, boasting, concluding*)

  - Directives (*asking, ordering, requesting, inviting, advising, begging*)

  - Commissives (*promising, planning, vowing, betting, opposing*)

  - Expressives (*thanking, apologizing, welcoming, deploring*)

  - Declarations (performatives: *I dub thee... You're fired*)

(Searle 1975)

Dialogue acts, DAMSL codes

- Forward-looking functions:

  - Statement

  - Info-Request: check

  - Influence-on-Addressee: Open-Option, Action-Directive

  - Influence-on-Speaker: Offer, Commit

  - Conventional: Opening, Closing, Thanking

Dialogue acts, DAMSL codes

- Backward-looking functions:

  - Agreement: Accept, Accept-Part, Maybe, Reject-Part, Reject, Hold

  - Answer

  - Understanding: Signal-Non-Understanding, Signal-Understanding: Acknowledge, Repeat-Rephrase, Completion

Recognizing Dialogue Acts

- Dialogue acts are not unambiguously signaled by syntactic form

- "It's hot in here."

- "Can you pass the salt?"

- $\rightarrow$ indirect speech acts

- **idiom** vs. **inferential** approaches

Belief, Desire, Intention: BDI

- A model of inference to generate or interpret indirect speech acts.

- Formal definitions/representations of belief and desire:
  - B(S,P)
  - KNOW(S,P) ≡ P ∧ B(S,P)
  - KNOWIF(S,P) ≡ KNOW(S,P) ∨ KNOW(S,¬P)
  - W(S,P)

Belief, Desire, Intention: BDI

- Action schemas (axiomatizations of actions and planning):

- A set of parameters with constraints about the type of each variable, plus"
  - Preconditions
  - Effects
  - Body

Speech Act Schema

INFORM(S,H,P)

    Constraints: Speakers(S) ∧ Hearer(H) ∧ Proposition(P)

    Precondition: KNOW(S,P) ∧ W(S, INFORM(S,H,P))

    Effect: KNOW(H,P)

    Body: B(H,W(S,KNOW(H,P)))

Plan Inference Rules

- Action-Effect Rule:

  $\forall S, H \text{Effect}(X, Y) \wedge B(H, W(S, X))$

  $\Rightarrow \text{(plausible) } B(H, W(S, Y))$

- Precondition-Action Rule:

  $\forall S, H \text{Precondition}(Y, X) \wedge B(H, W(S, X))$

  $\Rightarrow \text{(plausible) } B(H, W(S, Y))$

- Body-Action Rule: $\forall S, H \text{Body}(Y, X) \wedge B(H, W(S, X))$

  $\Rightarrow \text{(plausible) } B(H, W(S, Y))$

Plan Inference Rules

- Know-Desire Rule:

  $\forall S, HB(H, W(S, KNOWIF(S, P)))$
  $\Rightarrow$ (plausible) $B(H, W(S, P))$

- Extended Inference Rule: If

  $B(H, W(S, X) \Rightarrow$ (plausible)$B(H, W(S, Y))$

  is a PI rule, then the following is also a PI rule:

  $B(H, W(S, B(H, W(S, X)))) \Rightarrow$
  (plausible)$B(H, W(S, B(H, W(S, Y))))$

BDI

- Use plan inference rules to reason backwards (cf. discussion of text coherence)

- Again the search problem arises: when to apply which rule?

- Powerful yet expensive approach

Cue-based dialogue act interpretation

- Based on the idiom intuition of indirect speech acts

- Combines multiple sources of information:

  - N-gram language models, one for each speech act, trained on distinct subcorpora

  - Prosodic information (again applied probabilistically)

  - N-gram discourse structure models

System-Initiative Dialogue Managers

- Can be modeled by FSAs (think flow charts)

- The system controls what comes next

- Examples?

- Drawbacks?

A more adequate model: TRAINS-93

while conversation is not finished
    if user has completed a turn
    then interpret user's utterance
    if system has obligations
    then address obligations
    else if system has turn
    then if system has intended conversation acts
        then call generator to produce NL utterances
        else if some material is ungrounded
        then address grounding situation
        else if high-level goals are unsatisfied
        then address goals
        else release turn or attempt to end conversation
    else if no one has turn
    then take turn
    else if long pause
    then take turn

Overview

- Overview of dialogue systems

- Properties of dialogue

- Dialogue acts

- Interpreting dialogue acts

- (Discourse structure)

- An overly simple dialogue manager (FSA)

- A more adequate dialogue manager

- Demo: WITAS

Demo: WITAS

- http://www-csli.stanford.edu/semlab/witas/