Multiparty Dialogue

Stefan Behr

Multiparty Dialogue (MPD)

- Extension of the two-party case
- Involves >2 participants
- Each participant can be either human or a computational agent
- Makes sense to consider >1 computational agent
- Probably simplest to initially consider case with only one computational agent
- Organization of issues largely derived from Traum 2004

Affected Issues

- Design considerations for high-level aspects of spoken dialogue systems can be greatly affected by moving to the multiparty case
  - Participant roles
    - Determining who plays what part
    - Acting and reacting accordingly
  - Interaction management
    - Detecting types of interaction between different participants, participant roles
    - Arbitration of interaction
  - Grounding and obligations
    - Adding to common ground between participants
    - Determining obligations to act or respond

Participant Roles

- Dialogue consists of conversations
- Different important types of roles
  - Local roles, centered on single conversation
    - Shift during conversation
    - e.g., speaker, hearer
  - Non-local roles tied to activity subsuming conversation
  - Constant across convo, maybe not dialogue
    - e.g., Mission Rehearsal Exercise (MRE)
      - Simulation of military missions
      - Conversation roles include Army Lieutenant, Medic, Sgt., etc.
  - Global roles transcending dialogue session
    - Constant across dialogue
    - e.g., professor, student, etc.

Conversational Roles

- Change during conversation
- Two-party case: speaker, hearer
- Multi-party case more complicated
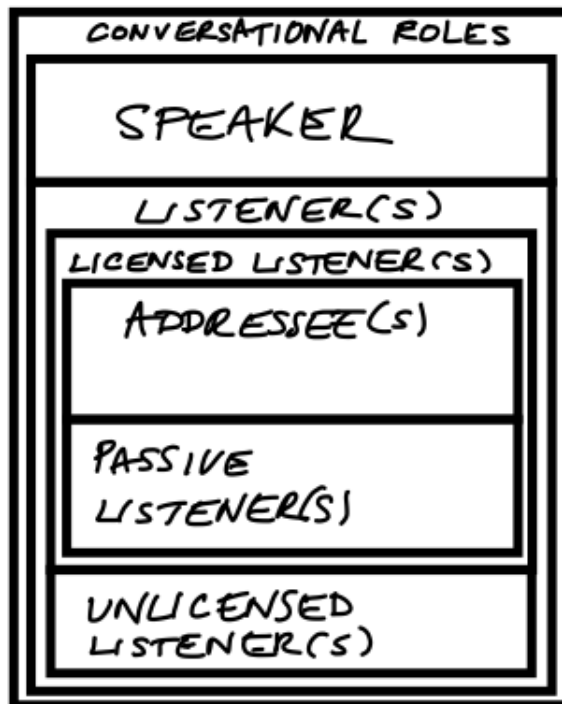
Conversational Roles

- Additional issue of whether listener is in-context or not
  - Generally not a consideration in two-party dialogue since two participants assumed from outset
  - Participants can jump in to MPD without having heard prior utterances

Non-local Conversational Roles

- Roles which are constant across conversation
  - Not necessarily constant across dialogue
  - Example roles
    - Active participant
      - Actively alternates between speaker and listener roles
    - Passive participant
      - May listen and occasionally contribute
      - Generally not central to conversation
      - e.g., peanut gallery?
    - Non-participating listener
      - e.g., eavesdropper
      - Might still need to understand utterances to plan actions

Determining Roles

- Several main issues in MPD concerning determining roles
  - Speaker Identification
  - Addressee Recognition
  - Determining other participant roles

Speaker Recognition

- Easy in two-party: speech not from you must be from other
- Possibly non-trivial in MPD
  - Agent-agent communication: directly in channel
  - Given single audio stream
    - Classify speakers by acoustic features
    - Classify speakers by stylistic features
    - Self-identification (assuming speaker integrity)
  - Multi-modal information (audio, vision, etc.)
    - Stereo microphone to localize speech output
    - Computer vision to see lip movement, gestures
    - Cues from one speaker to another to start speaking

Addressee Recognition

- Two-party: trivial, like speaker identification
- MPD, as usual, complex
  - Distinguish between *hearers* and *addressees*
  - Hearers
    - Fairly simple to do identify
    - Computed by properties such as audio output volume, background noise, assumed perceptual abilities, distance from agents
  - Addressees
    - Indicate via vocative
      - e.g., "Gina, ...", "Professor, ...", etc.
    - Utterance content if it makes addressee clear
      - e.g., "Give me yesterday's statistics." to sole statistician

Addressee Recognition

- Utterance context
  - Assume default that previous speaker is addressee of current speaker
- Multi-modal information
  - Speaker gaze information
  - Speaker body orientation information
  - Gaze and body orientation of other participants at speaker
  - Hearer presence in video frame
  - Not enough alone, but can support addressee hypotheses
- Simple addressee recognition algorithm from Traum 2004, used in Mission Rehearsal Exercise (MRE)
  - [next slide]

Addressee Recognition

1. **If** utterance specifies addressee (e.g., a vocative or utterance of just a name when not expecting a short answer or clarification of type person)
   then Addressee = specified addressee
2. **else if** speaker of current utterance is the same as the speaker of the immediately previous utterance
   then Addressee = previous addressee
3. **else if** previous speaker is different from current speaker
   then Addressee = previous speaker
4. **else if** unique other conversational participant
   then Addressee = participant
5. **else** Addressee unknown

Fig. 1. MRE Agent Speech Addressee Identification Algorithm

Other Participant Roles

- Consider tasks in conversations
  - i.e., answering questions, performing physical action
- Two-party
  - Agents are typically task performers or eliciters
- MPD requires more complex tasking model
  - Some agents perform primitive tasks
    - Primitive task: atomic unit of complex task?
  - Others responsible for executing complex task
    - Delegate primitive tasks to others
    - Alternatively, execute all primitive subtasks
  - Yet others can authorize task execution
  - Another feature is desire for task execution
    - Agents can also be guards *against* execution

Other Participant Roles

- Hodgepodge of other complex relationships between dialogue participants
  - Relative length and content of turns
  - Right to assign turns
  - Right to set/end/change conversation topic
  - e.g., Shopkeeper/buyer, teacher/student dialogue
  - Courtroom dialogue
    - Judge (nearly all rights)
    - Clerk (mostly passive)
    - Prosecutor/defense counsel (limited rights)
    - Witness (limited rights)
- Persistent social roles
  - Status (superior, equal), closeness (friend, family)

Interaction Management

- Turn management
  - *When* are things said, by whom?
- Channel management
  - *Where* are things said?
- Thread/conversation management
  - *What* things are said when?
- Initiative management
  - *Who* determines topic agenda?
- Attention management
  - Are agents involved?

Turn Management

- When to speak?
- When to stop speaking?
- Turn-taking can be modelled and determined by numerous cues
  - Barging in
  - Verbal signals of continuation or turn termination
    - Prosody
    - Sentence structure
    - Filled pauses (discontinuities)
  - Non-verbal signals
    - Gaze
    - Gesture

Turn Management

- In MPD, turn-taking complexity increases due to increased number of possible turn takers
  - New types of actions possible in addition to **take-turn** and **release-turn**
  - e.g., **assign-turn <participant>** in case of arbitration role
  - e.g., **request-turn** in case of newly entering participant

Channel Management

- Single channel is common, easy to manage
- Multiple channels can be used
  - Uni-modal
    - All channels use same modality (e.g., speech)
  - Multi-modal
    - Different channels use different modalities
    - e.g., video feed for gesture recognition, audio feed for speech information
  - Must decide which channels used for which content
    - Use video channel for sign language speech, audio channels as backchannels?
    - Use audio channel for English speech, video channels as backchannels for attention detection?

Thread/Conversation Management

- What is being communicated when?
- How do topics progress from one to another?
- MPD issues
  - Possible to have multiple topics open in parallel
  - Concurrent topics can involve different subsets of participant set
  - Multiple successive conversations can be related to each other but interrupted by activity/silence
    - How to maintain important pieces of context?
  - Conversations can dynamically split and merge between/from different participant groups
  - How to match utterances to conversations?
    - Speaker involvement in many concurrent convos

Initiative Management

- Who determines dialogue flow?
- Two-party: user-, system-, mixed-initiative
- MPD systems are less symmetric
  - The more participants, the less likely each participant will have same amount of initiative
  - Tendency to develop leaders
  - Cross-initiative possible: participant assigns initiative to someone else, or third party interjects
  - Cross-conversation initiative: conversation dependencies (C1 depends on C2) cause initiative-holder in C1 to take direction from initiative-holder in C2

Attention Management

- Two-party
  - Attention assumed to be present
  - Attention assumed to be on agent or conversation
- MPD
  - More complex model of attention
  - Not much discussion in Traum 2004
  - Promising ideas
    - Gesture, gaze, orientation features to detect attention to conversation
    - Engagement in other, parallel conversations to signal inattention to yet other conversations

Grounding and Obligations

- Grounding: Adding to common ground of assumed knowledge between participants
- Obligation: Determining obligations of participants to act or respond depending on utterances and other cues

Grounding

- Common Ground Units (CGU) model used in MRE
  - CGU consists of initiator, responder, contents, state
  - State calculated using FSA
    - Updated by *grounding acts* performed on CGU
    - States include those in which CGU contents are *grounded* and *ungroundable*, as well as intermediate states which require further action before reaching *grounded/ungroundable* state
  - Should this model be used in multiparty dialogue?
    - Must be careful to not overestimate common ground -- carefully decide who gets to ground
    - Maybe don't need all agents to ground either!

Obligations

- How do you determine obligations in response to utterances by speech or action?
  - Does the utterance require verbal response? Physical action?
  - Should the request be rejected/clarified/negotiated?
  - Does the utterance require response from a specific individual? All participants individually?
  - Does the utterance assign an indefinite obligation to the group?
    - What motivates a specific individual to fulfill the obligation?

Obligations

- How do you handle transfer of obligation to other participants?
  - If participant A is asked a question by B, and C answers in B's presence, is A's obligation to answer fulfilled?
  - If A is asked a question by B, and A delegates the question to C, who answers in B's presence, is A's obligation fulfilled then?
  - How do you determine that C can barge in?
  - Answers to this can vary by setting, e.g., classroom, court, office meeting

References

- Traum, D. (2004). Issues in multiparty dialogues. In *Advances in agent communication* (pp. 201-211). Springer Berlin Heidelberg.
- Bohus, D., & Horvitz, E. (2011, November). Decisions about turns in multiparty conversation: from perception to action. In *Proceedings of the 13th international conference on multimodal interfaces* (pp. 153-160). ACM.
- Bohus, D., & Horvitz, E. (2009, September). Models for multiparty engagement in open-world dialog. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*(pp. 225-234). Association for Computational Linguistics.