Multi-participant Dialog Systems

Satanjeev Banerjee
Dialog Seminar: 12/04/2003
Papers we shall see today

• Embodied Agents for Multi-party Dialogue in Immersive Virtual Worlds
  – Traum and Rickel, AAMAS 2002

  – Matsusaka, Fujie and Kobayashi, Eurospeech 2001
Embodied Agents in Multi-party Dialog in Immersive Virtual Worlds

• Aim: To create agents that can participate in virtual 3D conversation with humans

• Challenges in “face-to-face” conversation:
  – How to handle non-verbal cues?
  – How to handle focus of attention? Or, Who is talking to whom? Who is listening to whom?
  – How to handle multi-party conversations? Or, Who heard what? Who is aware of what?
Scenario: Mission Rehearsal

Large screens, immersive audio equipment

Scene: Village in Bosnia

Hollywood-style story of what has just happened (pg 2, col 1)

Sample dialog: Figure 2 (page 3) of paper.
Agents based on “Steve”

• Steve = Earlier work by authors on animated virtual tutoring agent

• Uses path planning algorithms to move around virtual worlds

• Uses human’s gaze to
  – Reference objects in 3D world
  – Regulate turn-taking in multi-party dialog
Multi-party Dialog Layers

• Contact Layer: Concerned with whether others are accessible to conversation thru:
  – Visual
  – Audio (shout, normal, whisper)
  – Radio

• Attention Layer: Concerned with the agents’ object of attention
Multi-party Dialog Layers (2)

- **Conversation**: Concerned with all the aspects that make up a specific conversation
  - Participants: Speakers, addressees, overhearers
  - Turn: Person who has the right to speak now
  - Initiative: Person who is controlling the dialog
  - Grounding: Tracks how information is added
  - Topic: Tracks relevance of speech to topic
  - Rhetorical: The structure of the dialog (?)

- Social commitments and negotiation: **Obligations** to act, **restrictions** on actions, and how that is decided by the agents through **negotiation**.
Contact Layer Implementation

• Vector for all participants that specify if the participant is “in ear-shot”

• *Make-contact*: agent walks into ear-shot, turns on radio, turns to look at speaker

• *Break-contact*: agent walks out of ear-shot, turns off radio, turns around to look away

• Contact is essential for conversation to start
Attention Layer Implementation

- Vector similar to contact layer
- Specifies if agent is paying attention or not!
- *Give-attention:* Agent does back-channeling or gazes at object of attention
- *Withdraw-attention:* Agent looks away
- *Request-attention:* Agent asks for attention – by calling another agent by name, raising hand, etc
- *Release attention:* Agent signals attention no longer needed – like at end of conversation
- *Direct attention:* Agent can command another agent to pay attention
Conversation Layer Implementation

• Conversation opening is intricate
  – First need “contact” – signal intent to converse with an agent by gazing longer than normal
  – Next, physically approach agent, avoiding gaze
  – On close proximity, resume eye contact

• Conversation closing can be
  – Abrupt – just move on to other jobs
  – Highly formal in military setting – saying “out”
Turn-taking Actions

• **Take-turn**: Take turn by starting to speak.
• **Request-turn**: Ask for turn by opening mouth, raising hand, avoiding speaker’s gaze at phrase boundaries(!), etc.
• **Release-turn**: Indicate end of turn by falling intonation, removing hands from gesture space, gazing at listener at end of utterance.
• **Hold-turn**: Continue to hold turn using filled pauses, gaze avoidance at phrasal boundaries, etc.
• **Assign-turn**: Give turn to another agent by verbal signal, or gaze.
Other Layers

- **Initiative**: Usually only one agent controls conversation, but others can take-initiative, hold-initiative or release-initiative.
  - Not currently implemented

- **Grounding acts**: repair, request repair, display (repeat word, perform action etc to show understanding), acknowledge (backchannel), request-acknowledge (gaze at addressee at phrasal boundaries)

- **Topic**: Change topic with words like “now”, “anyway”, and with head movements, body shifts

- **Negotiation**: Involves figuring out what needs to be done, who will do it, and whether everyone agrees to plan.
Implementation? Evaluation?

• Implementation in process.
• Evaluation may come later on…
• What can we learn from this paper?
  – Sounds intuitive
  – Sounds like a lot of hand tuning
    • But if the acts are domain independent, that may not be so bad
  – Need to see if this got implemented, and evaluated.
A Robot Participating in a Group Conversation

• Aim: To build a robot that can take part in a multi-party conversation

• Issues to solve:
  – “Understanding” the conversation
  – Distinguish which questions are directed at him and which to the other participants
  – Decide what role the robot should play
    • Keep quiet and observe?
    • Jump in with comments?
Some Terminology

Parties concerned

Primary receiver

What can you ...

Focusee

Observers
Two Levels of Modeling Turn Taking

• Micro-level: The “local dynamics” of the conversation
  – Who’s turn to speak is it now?
  – Who’s turn is it next?

• Macro-level: Higher “social dynamics”
  – Attitude, motivation, contribution (more later)
Rules for Micro Level Turn Taking

• From Sacks et al.

• Rule 1: If current speaker (C) selects next speaker (N), then N gets the turn
  – Bano: So what do you think about this Alex?

• Rule 2: If current speaker doesn’t select anyone, and ends his turn, anyone can start
  – Bano: And that’s all I had to say. *pause*
  – June: I have something to add…

• Rule 3: If current speaker (C) stops, and no one picks the turn, C can restart
  – Bano: And that’s all I had to say. *pause*
  – Bano: Okay, what is the next thing on the agenda?
How Implemented

- Who is C giving his turn to?
  - Use C’s gaze direction at end of speech.
- How to detect rule 2 is being used?
  - Detect pause in speech (if C wants to keep turn, he’ll use filled pauses)
Macro Level Strategy

• Rule of thumb: Take turn only when sure that you can contribute to the dialog

• Ways to contribute to conversation:
  – Provide topics to talk about
    • Robot: Did you guys see Conan last night?
  – Arouse interest by intellectual support
  – Support the decision
    • Robot: Sure, let’s talk about Jay Leno instead.
  – Point out mistakes
    • Robot: No Conan didn’t get the Emmy this year.
  – Take conversation minutes, collect information, create user model, learn human manner
    • Robot (to himself): Hmm, so Bano likes Conan.
Their Scenario: Talk about Baseball

B: What team played with Giants?
A
[m0 <= play(Giants, x)]
R: (understand) > B
[buffer <= query(m0)]

R: (interest) > A
A: The Giants played the Hiroshima. > B
[m1 <= play(Giants, Hiroshima)]
R: (agree) > A
[compatible(butter, m1)]
Scenario (2)

R: (interest) -> B
B: Did Giants win? -> A
   [m2 <= win(Giants, Hiroshima)?]
R: (understand) -> B
   [buffer <= query(m2)]

R: (interest) -> A
A: Yes, of course. -> B
   [m3 <= yes]
Scenario (3)

A: (interest) > R
B: (interest) > R
A: Is that really? > R
[m4 <= right?]
R: Yes, Giants lose by 2 to 4. > A
[utter(query(m4))]

R: (disagree)
[!compatible(buffer,m3)]
No, Giants lose. > A
[utter(buffer)]
Hardware used

• Two cameras on head
• Used to capture images, and also as “eyes”
  – The robot can look at you squarely in the eye
  – Or give you a sideways glance to mean “temporary attention”
• Image processing software and audio processing software run in parallel on distributed processors
  – Communicate through black board.
“Dialoging”

- Robots knows a lot about baseball
- Parses input sentences to pick up phrases
- If a question has been asked, it finds the answer from its database
  - If question was directed at someone, and he gives the wrong answer, the robot corrects him.
  - If question was thrown out in the open, or directed at the robot, the robot answers
  - If the robot can’t understand the question, it utters some random sound like “Hmmm”
Limitations of the System

• Currently only handles 2 participants besides the system
  – Authors expect more participants = problems figuring out who is paying attention to who

• Doesn’t handle interruptions well
  – Probably just keeps speaking even if interrupted

• Speech recognition is a problem
  – Not very big, since the user can just ignore what the robot says, but still something that needs fixed
What do we learn?

• Does much of the things we would like to do in multi-party conversations
  – But restricted-domain system → how will this scale up?

• Nice combination of gaze and speech!

• How did users like the system? Evaluation needed!