Natural Language and Dialog

Artificial Intelligence for Interactive Media and Games

Professor Charles Rich
Computer Science Department
rich@wpi.edu

Outline

- Computational theory of human language and dialog
  - background
  - terminology

- Language and dialog in games
  - current common industry practice
  - emerging trends

- Speech
What is Dialog?

- a conversation between two participants
  - verbal communication
    - spoken or written
    - what about non-verbal components?
  - at least two turns
    - each turn consists of one or more utterances
    - not necessarily complete sentences
    - backchannels (uh-huh), overlapping, interruptions
  - what about more than two participants?
    - more complex turn-taking rules
    - dialog is a two-person discourse
- in a shared context
  - not just any random utterances
  - e.g., a story or collaboration

What is the Purpose of Dialog?

- contributes to participants’ goals in the context
  - example
    - my goal (desired world state) is for the window to be open
    - I say “Please open the window” to person standing next to the window
    - if person is cooperative, she says “Ok”
    - she opens the window
  - is it still a dialog if she skips saying “Ok”?  
    - yes, but a bit degenerate
  - notice interleaving/coordination of communication (utterances) and action (world state changes)
What is the *Immediate* Purpose of Dialog?

- the *speaker* is trying to achieve a change in the *mental state* of the *hearer*
  - emotional state
    - “your mother wears army boots”
    - “I love you”
  - beliefs
    - “roses are red”
    - “I’m scared”
  - goals (intentions)
    - “please open the window”
    - “don’t look in there”
What about Questions?

- “What time is it?”
  - speaker’s goal is to change his own mental state
    - to one in which he knows the time
  - speaker *could* achieve goal by looking at clock
  - but if no clock, can achieve goal indirectly
    - by changing mental state of hearer (with wrist watch)
    - to include goal of telling speaker the time
  - e.g., “Please tell me the time.”

---

Levels of Language Representation

1. Surface form (realization)
2. Syntax
3. Semantics
4. Pragmatics

*many-to-one mappings from each level to next*

- *multiple surface forms with same syntax*
- *multiple syntactic forms with same semantics*
- *etc.*
1. Surface Form (Realization)

- The sequence of words that are actually written, read, spoken or heard
- Two utterances, e.g., in two different languages, may differ only in their surface forms
  - English: “the roses are red”
  - French: “les roses sont rouges”

2. Syntax

- Parsing (diagramming) a sentence
  - part of speech tags: adjective, preposition, noun, etc.
  - syntactic roles: subject, verb, (direct/indirect) object, etc.

```
[S
  [NP The quick brown fox]
  [VP jumped
    [PP over
      [NP the lazy dog]]]]
```
3. Semantics

- The meaning of a sentence in isolation
- Much less standardized than syntax
  - frame-based semantics
  - logical (axiomatic) semantics
  - probabilistic semantics
  - etc., etc.
- Two sentences with different surface form and syntax may have same semantics
  - "John kissed Mary."
    [S [NP John] [VP kissed [NP Mary]]]
  - "Mary was kissed by John."
    [S [NP Mary] [VP was kissed [PP by [NP John]]]]
  - frame semantics:
    {action: kiss, agent: John, theme: Mary, time: past}

4. Pragmatics

- Everything else about how the utterance functions in its context
- Even less standardized than semantics
  - E.g., goal/belief modification semantics
Language Understanding

- Start with surface form
- Compute pragmatic function

```
“Please open the window”
```

```
Goal(WindowOpen)
```

- Perhaps mapping through syntactic and semantic forms along the way...

Language Generation

- Start with pragmatic (deep) representation
- Output surface form

```
Goal(WindowOpen)
```

```
Goal(WindowOpen)
```

```
“Please open the window”
```

- Perhaps mapping through semantic and syntactic form along the way...
State of the Art in Academic Research

- unrestricted language (speech) understanding input
- constrained domain
- full syntactic parsing, semantic interpretation
- pragmatics, planning
- general-purpose language generation

http://
www.cs.rochester.edu/
research/cisd/projects/trips

Language Understanding Challenges

- **Coverage**
  - you can make almost anything work if you restrict the domain enough
    - know all the words that will be used
    - know all the purposes (pragmatics)
  - e.g., airline reservation system
  - but not the Turing Test

- **Semantics**
  - lack of agreement inhibits generalization and sharing of results
Language Generation Challenges

- **Expressiveness**
  - how to say the same thing with different styles, emotional content, etc.
  - e.g., “Hello” vs. “Yo, dude”
  - need computational theory which separates style and content

- **Coherence**
  - generation needs to have wider window than single utterance
  - planning a sequence of utterances (anaphora, etc.)

Dialog in Games

- In what genres is dialog most important?
  - role playing games (RPG)
  - text adventure (interactive fiction - IF)
  - first person shooters (FPS)
  - real-time strategy (RTS)
  - sports? casual? serious? ...
Dialog between Whom?

- **player ⇔ NPC**
  - main challenge and focus
  - “dialog trees” commonly used
- **NPC ⇔ NPC**
  - player is bystander
- **player ⇔ player**
  - e.g., in MMO’s
  - no problem for humans on both ends
  - system/NPC as bystander?

Player-NPC Dialog

- Two computational problems to solve
  - generating NPC utterances
  - understanding player utterances
- Dialog trees
  - common solution to *both at the same time*
  - all possible player and NPC utterances authored in advance
  - decision tree based on user choices
Dialog Trees

Speak("Welcome stranger. What brings thee among us gentle folk?")

reply = player.SpeakOption(
    1, "Yo dude, wazzup?",
    2, "I want your money, your woman and that chicken")

if reply == 1 then
    Speak("Wazzuuuuup!")
else if reply == 2 then
    Speak("Well, well. A fight ye wants, is it? Ye can't just go around these parts demandin' chickens from folk. Yer likely to get that ugly face smashed in. Be off with thee!")
end

[From Buckland, Chapter 6]

Dialog Trees

- **Advantages**
  - *fast & flexible* – code can do anything
  - *reliable* – no misunderstandings
  - *expressive* – author has complete control to create desired style, character, atmosphere, etc.

- **Disadvantages**
  - restricts player
  - very labor intensive
  - doesn't scale well to complex interactions
    - must keep variability down to keep labor down
    - leads to lack of replayability
    - can help somewhat by designing special editors and engines for executing dialog trees
Morrowind (TES) Dialog Editor

- database of utterances
  - organized around extensible set of “topics”
  - each utterance has conditions and results
  - including menu choices
  - simple templating via variables (name, etc.)
  - sort of like programming a rule-based system

http://tommyshideout.net/files/srikandi/DialogueTut
Text Adventure Games

- also called “interactive fiction” (IF)
  - confusing with “interactive drama” and “interactive storytelling”
- started with Adventure in 1975
  - Infocom’s Zork series in 70’s and 80’s
- Interactive Fiction Competition (2008)
- restricted syntax and semantics
  - few hundred standard verbs
  - open-ended nouns
  - simple (learned) command syntax
  - highly evolved set of ad hoc techniques

>WAIT
Time passes...

The scooter glides into the station's docking port. The retro-thrusters bring the scooter to a halt. As the docking port fills with air the scooter's hatch opens.

>LOOK
Scooter, in the pilot's couch
You are in the cramped one-man space scooter. Through the viewport, you see the docking port of Space Station BG-12. The scooter's hatch is open. You can see a toolbox here.

>OPEN THE TOOLBOX
You can't reach it from the pilot's couch.
Text Adventures Games

>STAND UP
You are standing again.

>OPEN THE TOOLBOX
It's locked.

>LEAVE THE SCOOTER
Docking Port
This is the huge docking port of the space station. Only one ship is here at the moment, a one-man scooter, so the docking port seems unusually empty. A tube leads down toward the heart of the station,

>DOWM
Main Hallway, Sector M
This is the station's main corridor, which continues to port and starboard. A tube leads up toward the docking port.

Emerging Trends in Game Dialog

- Natural language understanding
  - replacing fixed menu choices
  - give player more flexibility to express herself

- Natural language generation
  - generating NPC utterances procedurally
  - reduces authoring labor

- Speech
Façade

State of the art in experimental game NLU
- unrestricted text input
- micro-domain (very constrained)
- go directly from surface form to pragmatic effect
- broad, shallow, author-intensive techniques
- cheating strategies when doesn’t understand

Façade – Surface Text Rules
- word spotting and pattern matching rules
  \(\Rightarrow\) **dialog acts** (pragmatic)
  
  \(\text{"hello" | "hi" | "there"} \Rightarrow \text{Hello}\)
  
  \(\text{"grace"} \Rightarrow \text{Character(Grace)}\)
  
  \(\text{Hello} \&\& \text{Character(?char)} \Rightarrow \text{Greet(?char)}\)

- example dialog acts:
  - Agree(?char), Disagree(?char)
  - Express(?char, ?emotion)
  - ReferTo(?char, ?object)

http://www.interactivestory.net
(2005)
ANDI-Land

http://www.andi-land.com

"Logical Agents for Language and Action",
M. Magnusson & P. Doherty,
Linkoping U., Sweden, AIIDE’08

- restricted natural language text input
  - using context-free grammar
  - shows user possible syntactic completions as player types
  - underlying logical theorem-prover
  - all output generated procedurally

Magni: “Who owns the axe?”

↓ parsing
[S Who [VP owns [NP the axe]]]

↓ semantic interpretation
informRef(magni, value(12:15, owner(axe)))

↓ theorem proving
inform(magni, Id(value(12:15, owner(axe)), smith))

↓ reversible grammar

Smith: “I own the axe.”
ANDI-Land

Magni: “Sell the axe to me.”

- **parsing**
  
  \[ S \ [VP \ \text{sell} \ [NP \ \text{the axe}] \ [PP \ \text{to me}]] \]

- **semantic interpretation**
  
  \[ \exists t_1, t_2 \ [Oc\text{cur}(\text{smith}, (t_1, t_2), \text{sell(axe, magni)}))] \]

- **theorem proving**
  
  \[ \text{Committed}(\text{smith, } t_1, \text{Oc\text{curs}...}) \land \]
  
  \[ \text{Executable}(\text{smith, } (t_1, t_2), \text{sell(axe, magni)})) \land \]
  
  \[ \text{Believes}(\text{smith, } t_1, \text{ActionId}(\text{sell(axe, magni)}, \text{sell(axe, magni)})) \Rightarrow \]
  
  \[ \text{Oc\text{curs}(smith, } (t_1, t_2), \text{sell(axe, magni)})) \]

---

**Natural Language Generation**

- **Generating NPC to NPC dialog for Interactive Storytelling**
  
  - no pre-authored dialog
  
  - situations generated by autonomous planning agents
  
  - using logic and templates to generate surface forms

[http://www-scm.tees.ac.uk/f.charles](http://www-scm.tees.ac.uk/f.charles)
Speech

- Speech recognition
- Speech generation
- Speech in games
  - experiments with player speech input
  - NPC speech output almost always recorded

Speech Recognition

- widely available commercial systems
  - all based on HMM (Hidden Markov Models) trained on large corpora
  - built into Mac Leopard and Windows Vista [demo]
- easier vs. harder versions
  - isolated word vs. continuous
  - speaker trained vs. speaker independent
  - small vs. large vocabulary
  - grammar-based vs. dictation
  - push-to-talk vs. open-microphone
Speech Generation

- text to speech
- widely available commercial systems
  - many different “voices”
  - never sounds as good as recorded voices, especially if mixed
  - built into Mac Leopard and Windows Vista [demo]
- two approaches
  - concatenative
    - chops up and stitches back together recorded voices
    - usually sounds pretty good
    - a lot of labor to produce each voice
  - model-based
    - uses mathematical model of vocal tract
    - easy to adjust parameters to get different voices
    - less natural sounding

Emotional Speech Generation

- research of Catherine Pelachaud
- same words but different affect and gestures for different emotional states
Lifeline

- Sony 2003
- single word commands
- not too successful
- speech on Wii recently announced
  - waiting for first games

Star Trek Bridge Commander

- Totally Games 2002
- more natural use of voice commands
- “shields up!”
Brothers in Arms – Hell’s Highway

- Gearbox 2008
- more natural use of voice commands
- “follow me!”

Tactical Language

- 2008 spinoff of USC research
- very successful serious game

http://tacticalanguage.com
Summary

- Natural language and dialog in games
  - academic research techniques mature
  - a lot of interest at points of overlap between academia and industry (e.g., AIIDE)
  - initial experimentation in games mixed
  - potential for breakthrough application in games in next few years