How to improve clarification strategies in spoken dialogue systems? - Two proposals.

Verena Rieser

vrieser@coli.uni-sb.de

`http://homepage.mac.com/verenarieser/papers/igk_colloquium`

My work up-to-day & Outline of the talk

Hermine, the talking washing machine (MA):

- An example for sate-of-the art error handling strategies.

Confidence-Based Fragmentary Clarifications on Several Levels for Robust Dialogue Systems (MSc):

- What kind of Clarification Requests (CRs) do occur in human-human dialogue?

Error handling in task-oriented, robust (and multi-modal) dialogue systems (PhD):

- When is clarification relevant for the overall task-success?
- What are natural error handling strategies in a multi-modal setting?

To establish the common ground...

- Common ground: "[T]he sum of [two people's] mutual, common, or joint knowledege, beliefs, and suppositions", [Clark, 1996, page 93]
- Grounding: Process of adding to the common ground
- Error handling: Use of dialogue strategies to handle the situation in which the system fails to recognise the user's utterance.
- Clarification: Ensures and maintains mutual understanding given an error, i.e. helps to keep the common ground.

Current error detection strategies:

- Compare the ASR confidence score against a (manually set) threshold.
- Assign: accept, clarify, reject

Current error handling strategies:

- Complete non-understanding (reject): Incremental prompting:
  - ▷ Wie bitte?
  - ▷ Ich habe diese Formulierung nicht verstanden.
  - ▷ Sie sind im Hauptmenu. Hier haben Sie folgende Optionen...
- Uncertain understanding (clarify): Explicit and implicit confirmations:
  - ▷ Wollwaschgang. Ist das korrekt?
  - ▷ Wollwaschgang - bei wieviel Grad?

Critic

Problems with current error detection strategies:

- The (pragmatic) plausibility is not taken into account.
  - ▷ see [Gabsdil, 2004]

Problems with current error handling strategies:

- A clarification sub-dialogue is always initiated once an error was detected.
  - ▷ The decision process needs to be refined in order to keep task-related dialogues robust and efficient.
- The same surface form is used for every kind of problem.
  - ▷ What are naturally occurring forms of CRs?

Clarification Strategies in Human-Human Dialogues

- People address several levels of grounding.
  - ▷ How to model those levels within ISU-based systems? (The FRAGLE system)
- Most CRs are fragmentary.
  - ▷ How to do generation? (The FRAGLE system)
- People tend to point out the most problematic part of an utterance.
  - ▷ How to evaluate “most problematic"? (PhD)
- People tend to present their hypothesis rather than signaling non-understanding.
  - ▷ How do people behave in a multi-modal setting? (PhD)

Example

I would like to book a flight to Japan on the third.

- To Japan?!?
- The third of May?

PROPOSAL I: When to engage in Clarification Dialogue?

Error detection returns value “clarify", or “ reject"

“...when faced with ambiguity it is better to choose one specific interpretation and run the risk of making a mistake as opposed to generating a clarification subdialogue" [Allen, 1995]

→ Should we skip (some) clarification dialogues?

"Task-oriented dialogues should be efficient, robust and should lead to task-success."

- Efficient & robust: Robust methods lose fidelity and we risk misunderstanding.
- Task-success: Cautious clarification strategies ensure fidelity (mutual understanding), but dialogues get less efficient.

→ We should only engage in clarification if the fidelity of the informational content is relevant for the task-success.

→ We only have to clarify the relevant parts!

Overanswering:

- O: Where do you want to depart from?
  (most active node: **`depature`**)
- U: clarify(LF_1(From Frankfurt)) clarify(LF_2(on the 6th)).

→ **`depature`** is most relevant to clarify.

Deep semantic analysis:

- U: clarify(LF_1(I)) clarify(LF_2(to book)) accept(LF_3(a flight)) .

→ **`agent`** is not under discussion.

Extensions of Contextual RELEVANCE

- User model: Novice, expert
- Costs of performing CR: Tasks in parallel, further repairs required, number of clarifications asked already...
- Costs of misunderstanding: Probability and consequences of a late repair.
- Costs of missed opportunities*: Not performing some other actions at this point.
- Utility of other actions*: Can other actions help to solve the problem (e.g. asking task related questions)?
- User satisfaction: PARADISE evaluation & reinforcement learning (learning from experience)

* taken from [Traum, 1999]

PROPOSAL II: Error Handling in Multi-Modal Settings

- Grounding behavior varies widely on the media used, [Clark and Brenna

- For persistent mediums like visual feedback on a screen much less explicit grounding behaviour is observed, [Traum, 1999].

→ How to find the <u>optimal</u> clarification strategy in mutli-modal settings?

Method

- Setting up a Wizard-of-Oz experiment with a simulated ASR channel. $\longrightarrow$ How do humans react to different WERs?

- Using the Wizard-of-Oz simulation to bootstrap a reinforcement-learning-based dialogue system.

- Learning optimal dialogue strategies with a combination of reinforcement learning and empirical evaluation techniques (PARADISE).

- Use the context in the information state to determine the *type* of clarification.

How to introduce WERs?

- [Fraser and Gilbert, 1991] simulate errors by randomly substituting words in the input.
  - ▷ The type of errors that really do occur cannot be simulated like that.
- [Skantze, 2003] introduces a speech recognizer in the WOZ setting.
  - ▷ Need for a trained speech recognizer for this domain.
  - ▷ The performance of a speech recognizer cannot be controlled.
- [Stuttle et al., 2004] simulate an ASR channel by controlling the word error rate based on phonetic confusion matrix.
  - ▷ Confusion matrix is trained on a phone-labeled test output data from a speech recognizer.

Discussion

- Proposal II: How feasible is project II? How difficult is it to simulate an ASR? Are simpler methods sufficient for our purposes?

- Proposal I: The concept of "relevance" is defined with respect to the current domain. Is there a chance to generalize it?

- In general:
  - ▷ Both projects suppose an extended model of grounding.
  - ▷ Can we learn grounding behaviour from the representation in the Information State?

References

[Allen, 1995] Allen, J. (1995). the trains project: A case study in building conversational planning agents. Journal of Experimental and Theoretical AI (JETAI).

[Clark, 1996] Clark, H. (1996). Using Language. Cambridge University Press.

[Clark and Brennan, 1991] Clark, H. and Brennan, S. E. (1991). Grounding in communication. Perspectives on Socially Shared Cognition.

[Fraser and Gilbert, 1991] Fraser, N. M. and Gilbert, G. N. (1991). Simulating speech systems. Coputer Speech and Language, (5):81–99.

[Gabsdil, 2004] Gabsdil, M. (2004). Automatic Classification of Speech Recognition Hypotheses Using Acoustic and Pragmatic Features. PhD thesis, Computational Linguistics, Saarland University.

[Skantze, 2003] Skantze, G. (2003). Exploring human error handling strategies: Implications for spoken dialogue systems. In ISCA Tutorial and Research Workshop on Error Handling in Spoken Dialogue Systems.

[Stuttle et al., 2004] Stuttle, M. N., Williams, J. D., and Young, S. (2004). A framework for dialogue data collection with a simulated asr channel. In ICSLP.

[Traum, 1999] Traum, D. R. (1999). Computational models of grounding in collaborative systems. In Proceedings of the AAAI Fall Symposium on Psychological Models of Communication.