Building a Robust Dialogue System with Limited Data

Sharon Goldwater
Elizabeth Owen Bratt
Jean Mark Gawron
John Dowding

SRI International

other collaborators:
Harry Bratt, SRI International
Beth Ann Hockey, RIACS
Manny Rayner, RIACS
Frankie James, RIACS

Robustness and Limited Data

- Robustness: the ability to handle a wide range of input, including unexpected input
  - recognition level
  - parsing level
  - dialogue level
- Talk focuses on methods for building recognition and understanding language models when statistical models are impossible

CommandTalk Overview

- Spoken-language interface to ModSAF (Modular Semi-Automated Forces) battlefield simulator

- Interactions similar to commanding live forces

- English commands and mouse gestures can
  - create forces, points, lines
  - assign missions to forces
  - modify missions during execution
  - control system functions, e.g. the map display
  - inquire about the state of the simulation

General Approach

- Use one handwritten grammar for understanding, generation, and recognition language models
  - interviewed subject matter experts
  - used Gemini, a unification-based grammar formalism
  - single Gemini grammar used for both understanding and generation
  - developed Gemini-to-recognizer compiler to derive equivalent context-free grammar for Nuance speech recognizer

Advantages of the One-Grammar Approach

- Changes to grammar are automatically reflected in understanding, generation, and recognition models

- No discrepancy between recognizable and parseable strings

- Symmetry of input and output
  - system can echo user language
  - system utterances may guide user's choice of language

Disadvantages of the One-Grammar Approach

- Handwritten grammar lacks robustness of statistical models

How to Increase Robustness?

- Various approaches: fragment combining, constraint relaxation, etc.

- Our approach: string editing

  – try to process strings which are grammatical except for a few added or missing words

Insertions and Deletions

- User can insert extra words anywhere in string

- User can leave out words with no semantic contribution

- Example:
  - "Establish a base of fire at 963 472"
  - "Establish your base of fire position at 963 472"
  - "Establish... [a] base of fire... at 963 472"

- Insertions also useful for disfluencies
  - "Proceed to Obje–, uh, Checkpoint 1"

Word Insertions: Implementation

- Use "reject" model in recognizer
  - "reject phone" model trained on set of all English phones
  - "reject region" will therefore match any segment of speech
  - reject regions with longer duration accrue higher penalties
- Modified Gemini-to-recognizer compiler so recognition grammar contains optional reject region between every other word

Word Deletions: Implementation

- Based on Aho & Peterson's O(n^3) minimal-edit-distance parsing method (1972)

- Compiled new Gemini grammar
  - consulted semantics to determine omissible words
  - compiled new rules with words deleted
  - each rule has associated score equal to number of deleted words

- Modified natural language parser to prune hypotheses with score $>$ n

- Modified Gemini-to-recognizer compiler to add penalties in recognition grammar for rules with deleted words

Initial Experiments

- First, ensure that in-grammar performance doesn't decrease significantly
  - test set: 800 utterances read by SRI employees
  - ran on original and robust versions
  - collected error rates for recognition and logical forms

Results

- Recognition results:

	Control	Robust
Time, xRT	0.621	1.068
Sentence Rejects	2.56%	1.71%
Adjusted Word Errors	1.69%	2.94%
Sentence Errors	10.00%	12.07%

- Parameters affecting recognition results:
  - recognizer pruning threshold
  - penalties assigned for inserting or deleting words

- Logical form results:

	Control	Robust
Logical Form Errors	7.68%	8.17%

What about Out-of-Grammar Utterances?

- Problem: no user data, no access to subjects

- Possible solution: obsolete in-grammar data
  - 120 utterances collected at same time as in-grammar utterances above
  - now out-of-grammar due to changes in grammar

Results

- Recognition results:

	Control	Robust
Time, xRT	0.603	0.918
Sentence Rejects	70.2%	13.2%
Adjusted Word Errors	34.8%	34.6%
Sentence Errors	100%	89.3%

- Logical form results:
  - see examples

Example Utterance: Insertions

- "Red victor company slow down by three."
- Recognized as "Red victor company slow down."

Example Utterance: Deletions

- "B028 halt."
- Recognized perfectly
- Interpreted as "B02 at 8 hundred hours halt."
- Reasons:
  - valid call sign is letter-digit-digit
  - semantics for "at 8 hundred hours" depends only on "8"

Lessons

- Out-of-grammar data isn't realistic
- Semantically "meaningless" words may still constrain which rules apply

Hot off the Press!

- Preliminary results on data gathered from test subjects for RIACS PSA system

- In-grammar recognition:

	Control	Robust
Time, xRT	0.35	0.35
Sentence Rejects	4.79%	4.00%
Adjusted Word Errors	2.99%	3.48%

- Out-of-grammar recognition:

	Control	Robust
Time, xRT	0.35	0.37
Sentence Rejects	56.0%	44.2%
Adjusted Word Errors	30.1%	33.6%

Conclusions

- Allowing word insertions and deletions does not significantly affect in-grammar performance.

- Insertions: show promise, but need further testing on realistic out-of-grammar user data and disfluent utterances.

- Deletions: current implementation seems problematic in domains with high sortal ambiguity. May need to rethink definition of "meaningless."

- Bottom line: These techniques have good potential for adding robustness to grammar-based speech recognition and dialogue systems. Experimentation in other domains is clearly necessary.