Towards the Interpretation of Utterance Sequences in a Dialogue System

Ingrid Zukerman, Patrick Ye and Enes Makalic
Faculty of Information Technology
MONASH UNIVERSITY
Australia

DORIS (Dialogue Oriented Roaming Interactive System)
- A dialogue module for a robotic agent in a home environment
- *Scusi?* is DORIS’s language interpretation module
  - It will eventually combine spoken and visual information

Outline of this Talk
- Motivation for our main design decisions
- Interpreting a single utterance
- Interpreting a sequence of utterances
  - Estimating the probability of an interpretation
- Evaluation
- Conclusion
- Future work

Motivation
- DORIS will eventually
  - make decisions on the basis of the results of the interpretation process
    - dialogue actions and physical actions
  - modify decisions on the fly, given new information
  - recover from flawed or partial interpretations
- To support these activities, a speech interpretation module should
  - maintain multiple interpretations
  - apply a ranking process to assess the relative merit of each interpretation

*Scusi?* (DORIS’s Speech Interpretation Module)
- Maintains multiple interpretations
  - a multi-stage interpretation mechanism
  - each stage maintains multiple options
  - employs an anytime algorithm
- Applies a ranking process to assess the relative merit of each interpretation
  - a mechanism which estimates the probability that an interpretation matches the speaker’s intention

Interpreting a Single Sentence
- **Speech Recognition**
- **Syntactic Parsing**
  - **Semantic Interpretation**

Wave
↓ Text
↓ Parse Tree
↓ Concept Graph
Stage 1: Speech Recognition

ASR: Microsoft SAPI

- find the blue mug in the kitchen for Susan
- find the blue mat in the kitchen for Susan
- find the blue mug in a crisper for Susan
- finer blue mugging a kitchen 4 season

Stage 2: Syntactic Parsing

Parser: Charniak’s Statistical Parser

find the blue mug in the kitchen for Susan  
(S1 (S (VP (VB find) (NP (NP (DT the) (JJ blue) (NN mug)) (PP (IN in) (NP (DT the) (NN kitchen))))) (PP (IN for) (NP (NNP Susan)))))

Stage 3: Semantic Interpretation

- Relies on Concept Graphs
  - represent entities and relationships between them
-Performed in two stages
  - Uninstantiated Concept Graph (UCG)
  - Instantiated Concept Graph (ICG)

Uninstantiated Concept Graph (UCG)

- A UCG links lexical concepts to each other using the relationships in the parse tree
- A UCG is deterministically produced from a parse tree
  - one parse tree yields one UCG, but
  - one UCG can have multiple parents
- A UCG is domain independent
  - It is not associated with any concepts in DORIS’s KB

UCG Example

find the blue mug in the kitchen for Susan

Instantiated Concept Graph (ICG)

- Every concept and relationship in an ICG corresponds to an instance in the system’s knowledge base
### Extending Scusi? to Sentence Sequences

- People often utter several sentences to convey their wishes
  - Example: “Go to my office. Get my mug. It is on the table.”
  - Extensions to our mechanism for interpreting single utterances
    - Determine which sentences in a sequence are related, and combine them into an integrated representation
    - Provide a formulation for estimating the probability of a sentence sequence

### Determining Sentence Mode

- Employ Maximum Entropy Classifier
- Input features:
  - top parse-tree node
  - position and type of top-level phrases
  - regular expression for top-level phrases
  - top VP head
  - top NP head
  - first three tokens of the sentence
  - last token of the sentence
- Performance:
  - Accuracy of 99.2% – leave-one-out X-validation

### Determining Coreferents

- Handle pronouns, one-anaphora and NP identifiers
- Two steps:
  1. Identify a sentence being referred to
     - 4 types of referent sentences: current, previous, first, other
  2. Determine a referent within the sentence
     - Identify pronouns and one-anaphora
       - Pronouns: heuristics from [Lapin and Leass 1994]
       - One-anaphora: heuristics based on [Ng et al. 2005]
     - Construct a list of potential referents from the head nouns in the target sentence
Interpreting a Sequence of Sentences – Example

**IMP S0:** "Go to the desk near the computer."

- **U00:** go-(to-desk-(near-computer)) 0.6
- **U01:** go-(to-desk)-(near-computer) 0.4

**DEC S1:** The mug is on the desk near the phone.

- **U10:** mug-(on-desk)-(near-phone) 0.55
- **U11:** mug-(on-desk-(near-phone)) 0.46

**IMP S2:** Fetch it for me.

- **U20:** fetch-(object-it)-(for-me) 0.8
- **U21:** fetch-(object-it-(for-me)) 0.2

Merging Two UCGs – Example

**The mug is on the desk near the phone. Fetch it for me.**

- **U10:** mug on near desk phone fetch object-it for me
- **U20:** fetch object-it for me

Estimating the Probability of a Merged UCG Sequence

\[
Pr(U_1,\ldots,U_m) = Pr(U_1,\ldots,U_n, M_1,\ldots,M_n, C_1,\ldots,C_n | T_1,\ldots,T_n)
\]

After some conditionalization and incorporating Texts and Parse Trees

\[
Pr(U_1,\ldots,U_m) = \prod_{i=1}^{n} Pr(U_i | T_i) Pr(M_i | P_i, T_i) Pr(C_i | P_i, T_i)
\]

Estimating the Individual Probabilities (I)

- **Single UCG** [Zukerman et al., 2008]
  \[
  Pr(U | T) \propto \sum_P Pr(P | T) \times Pr(U | P)
  \]

- **Mode of a sentence**
  - Maximum Entropy classifier

Estimating the Individual Probabilities (II)

- **Coreferents for pronouns, one-anaphora and NPs**
  \[
  Pr(C_j | P_1,\ldots,P_k) = \prod_{j=1}^{k} Pr(C_j | P_1,\ldots,P_k)
  \]

- **After some conditionalization**

\[
Pr(C_j \mid P_1,\ldots,P_k) = Pr(\text{Type}(A_j) \mid P_1) \times \frac{Pr(A_j \text{ refn noun} \mid \text{Type}(A_j), P_1, P_k)}{Pr(A_j \text{ refn noun})}
\]

Evaluation – Experimental Set Up

- **People asked DORIS to do something in their office**
- **115 requests comprising sentence sequences**
- **Sequence length between 1-9 sentences**

- **Systematic manual changes to simplify the sentences in the requests**

**Example**

- my mug is in my office
- go to my office.
- get my mug. It is on top of the cabinet on the left of my desk.
Two Experiments

Exp 1: sentence pairs
- 106 pairs (1 dec, 1 imp)
- Text to ICG and speech to ICG
- Virtual environment
- 183 instantiated concepts

Exp 2: sentence sequences
- 115 sequences
- Text to UCG

Experiment 1 – Results

<table>
<thead>
<tr>
<th></th>
<th># Gold ICGs in top 1</th>
<th># Gold ICGs in top 3</th>
<th>Avg rank</th>
<th>Med rank</th>
<th>75%-ile rank</th>
<th>Not found</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>80 (75%)</td>
<td>91 (66%)</td>
<td>2.17</td>
<td>0</td>
<td>0</td>
<td>1 (1%)</td>
</tr>
<tr>
<td>Speech</td>
<td>45 (42%)</td>
<td>53 (50%)</td>
<td>1.75</td>
<td>0</td>
<td>1</td>
<td>42 (40%)</td>
</tr>
</tbody>
</table>

ASR top-ranked 54 correct texts
→ Scusi? overcomes some of the ASR error for utterance pairs

Experiment 2 – Results

<table>
<thead>
<tr>
<th></th>
<th># Gold in top 1</th>
<th>Avg rank</th>
<th>Med rank</th>
<th>75% rank</th>
<th>Not found</th>
<th>Total #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requests</td>
<td>59 (51%)</td>
<td>3.14</td>
<td>0</td>
<td>1</td>
<td>36 (31%)</td>
<td>115</td>
</tr>
<tr>
<td>UCGs</td>
<td>146 (62%)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>55 (23%)</td>
<td>234</td>
</tr>
</tbody>
</table>

- PP-attachment
- Anaphora resolution
- Not merging object-specs from imperative UCGs

Conclusion

- Speech interpretation module motivated by the requirements of a practical dialogue module
  - keeps track of (sub)interpretations at each stage of the process
  - provides a probabilistic formalism to handle the uncertainty inherent in the interpretation process

- Extension to utterance sequences
  - merge UCGs on the basis of sentence mode and coreference resolution
  - incorporate sentence mode and coreference resolution into our probabilistic formalism

Future Work

- Interleave UCG and ICG generation
- Deal with ASR error
- Extend Scusi?’s grammatical capabilities
- Consider additional dialogue acts
- Dialogue
- Integrate with vision

Questions?