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
fine the blue mug in a crisper for Susan
finer blue mugging a kitchen 4 season

Wave ↓ **Text** ↓ Parse Tree ↓ Concept Graph

7

Stage 2: Syntactic Parsing

find the blue mug in the kitchen for Susan

Parser: Charniak's Statistical Parser

(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*))))))

Wave ↓ Text ↓ **Parse Tree** ↓ Concept Graph

8

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)

Wave ↓ Text ↓ Parse Tree ↓ **Concept Graph**

9

Uninstantiated Concept Graph (UCG)

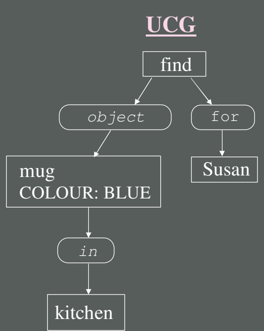
- A UCG links *lexical concepts* to each other using the *relationships* in the parse tree
- A UCG is deterministically produced from a parse tree
  - one parse tree yields one UCG, but
  - one UCG can have multiple parents
- A UCG is domain independent
  - It is not associated with any concepts in DORIS's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

Wave ↓ Text ↓ Parse Tree ↓ **UCG** ↓ ICG

10

UCG Example

find the blue mug in the kitchen for Susan



Wave ↓ Text ↓ Parse Tree ↓ **UCG** ↓ ICG

11

Instantiated Concept Graph (ICG)

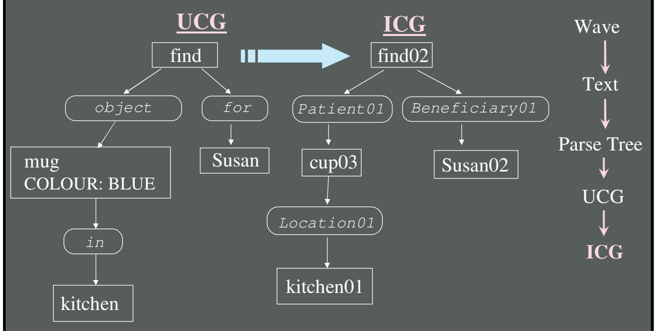
- Every concept and relationship in an ICG corresponds to an *instance* in the system's knowledge base

Wave ↓ Text ↓ Parse Tree ↓ UCG ↓ **ICG**

12

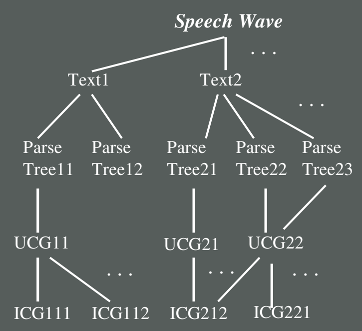
ICG Example

find the blue mug in the kitchen for Susan



13

Interpretation Search Graph



14

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

15

Interpreting a Sequence of Sentences

1. For each sentence
   a. Generate UCGs
   b. Determine mode (declarative, imperative)
   c. Determine coreferents
2. Generate UCG sequences *US*
3. Generate mode sequences *MS*
4. Generate coreference sequences *CS*
5. While there is time
   a. Select a promising tuple $\{US_i, MS_j, CS_k\}$
      - update US_i by merging its UCGs as specified by its mode and coreference sequence
   b. (Generate ICG sequences from the most promising UCG sequences)

16

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

17

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
     a. Identify pronouns and one-anaphora
        - Pronouns: heuristics from [Lapin and Leass 1994]
        - One-anaphora: heuristics based on [Ng *et al.* 2005]
     b. Construct a list of potential referents from the head nouns in the target sentence

18

Interpreting a Sequence of Sentences – Example

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

U00 go-(to-desk-(near-computer)) 0.6 → US0
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)) 0.45 → US1

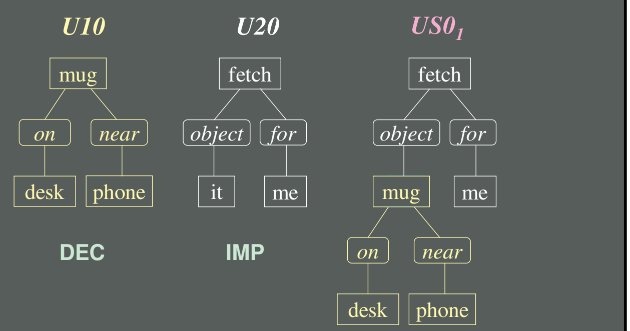
IMP S2: *Fetch it for me.*"

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

19

Merging Two UCGs – Example

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



20

Estimating the Probability of a Merged UCG Sequence

$$\Pr(\mathbf{U_1,\ldots,U_m}) = \Pr(\underbrace{U_1,\ldots,U_n}_{\text{UCGs}}, \underbrace{M_1,\ldots,M_n}_{\text{modes}}, \underbrace{C_1,\ldots,C_n}_{\text{coreferents}} \mid T_1,\ldots,T_n)$$

- After some conditionalization and incorporating Texts and Parse Trees

$$\Pr(\mathbf{U_1},\ldots,\mathbf{U_m}) = \prod_{i=1}^{n} \underbrace{\Pr(U_i \mid T_i)}_{\text{single UCG}} \underbrace{\Pr(M_i \mid P_i, T_i)}_{\text{mode}} \underbrace{\Pr(C_i \mid P_1,\ldots,P_i)}_{\text{coreferents}}$$

21

Estimating the Individual Probabilities (I)

- Single UCG [Zukerman et al., 2008]

$$\Pr(U \mid T) \propto \sum_P \underbrace{\Pr(P \mid T)}_{\text{Charniak parser}} \times \underbrace{\Pr(U \mid P)}_{1}$$

- Mode of a sentence
  - Maximum Entropy classifier

22

Estimating the Individual Probabilities (II)

- Coreferents for pronouns, one-anaphora and NPs

$$\Pr(C_i \mid P_1,\ldots,P_i) = \prod_{j=1}^{k_i} \Pr(C_{ij} \mid P_1,\ldots,P_i)$$

probability of coref for A_{ij} (anaphor j in sentence i)

- After some conditionalization

$$\Pr(C_{ij} \mid P_1,\ldots,P_i) = \Pr(\text{Type}(A_{ij}) \mid P_i) \times \Pr(A_{ij} \text{ refto sent}_b \mid \text{Type}(A_{ij}), i) \times \Pr(A_{ij} \text{ refto noun}_a \mid A_{ij} \text{ refto sent}_b, \text{Type}(A_{ij}), P_i, P_b)$$

heuristics
corpus statistics

23

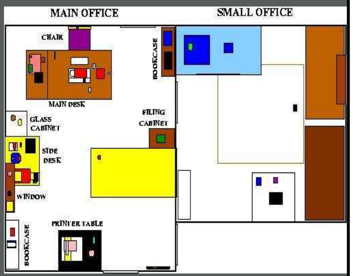
Evaluation – Experimental Set Up

- 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.*

24

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

	# Gold ICGs in top 1	# Gold ICGs in top 3	Avg rank	Med rank	75%-ile rank	Not found
Text	80 (75%)	91 (86%)	2.17	0	0	1 (1%)
Speech	45 (42%)	53 (50%)	1.75	0	1	42 (40%)

Compound ASR error

PP-attachment

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

Experiment 2 – Results

	# Gold in top 1	Avg rank	Med rank	75% rank	Not found	Total #
Requests	59 (51%)	3.14	0	1	36 (31%)	115
UCGs	146 (62%)	NA	NA	NA	55 (23%)	234

- 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?