Corpus-based Discourse Understanding in Spoken Dialogue Systems

Ryuichiro Higashinaka, Mikio Nakano, Kiyoaki Aikawa

NTT Communication Science Laboratories
Overview

• A new discourse understanding method in spoken dialogue systems
  - discourse understanding means utterance understanding taking the context into account
  - retains the ambiguity of a user utterance and resolves it by subsequent utterances
  - uses statistical information derived from dialogue corpora
Objective

• Spoken dialogue systems that can
  - accurately understand user intention using the context of a dialogue

Benefits:
  - more efficient dialogue
  - robust to misrecognitions
Discourse Understanding

User Utterance: e.g., "I’d like to go to Tokyo"

Speech Recognition

Recognition hypothesis: e.g., "I’d like to go to Tokyo"

Syntactic and Semantic Analysis

Dialogue act: e.g., [refer-destination: Tokyo]

Discourse Understanding

Update: e.g., (before) (After)

Dialogue State

|          | Origin | Destination |
|----------|--------|-------------|
| (Before) | --     | --          |
| (After)  | --     | Tokyo       |
Problem

• Ambiguities in discourse understanding
  - Speech recognizer outputs multiple recognition hypotheses (N-best)
  - Syntactic and semantic analysis produce multiple parsing results
    - multiple dialogue act candidates and thus multiple dialogue state candidates are derived from a user utterance

System has to appropriately rank the dialogue state candidates to obtain the most plausible user intention
Problem: an example

Example dialogue:
User: "To Sapporo"
System: "uh-huh"
User: "From Tokyo"

Which one is more plausible?

| Origin | Destination |
|--------|-------------|
| Tokyo  | --          |

Which one is more plausible?

| Origin | Destination |
|--------|-------------|
| Sapporo| --          |
Related Work

• ISSS Method (Nakano et al., 1999)
  - rank multiple dialogue states by hand-crafted scoring rules
    □ creating rules by hand is costly

• Estimation of dialogue act type (Nagata et al. 1994)
  - estimate the most probable dialogue act from previous dialogue act sequences
    □ mainly aims at improving recognition accuracy; not applied to dialogue systems
Approach

• Use of statistical information derived from dialogue corpora to score the dialogue states
• Keep the low-ranked dialogue states to allow possible understanding in the future
Statistical Information

- N-gram probability of a dialogue act type sequence (as Nagata et al.)
  - represents brief (superficial) flow of a dialogue
- Collocation probability of a dialogue state and the next dialogue act
  - deals with more detailed information about the dialogue
  - such as dialogue state changes including grounding information
Dialogue State Scoring

• Update the score of dialogue states by the following formula

Score of the updated dialogue state =

Score of the dialogue state before update

+ \( \theta \) \cdot Score of a dialogue act
  (from Speech Recognition and the Syntactic and Semantic Analysis)

+ \( \psi \) \cdot N\text{-}gram probability score of dialogue act type sequences

+ \( \eta \) \cdot Collocation probability score of a dialogue state and the next dialogue act

(\( \theta \), \( \psi \), and \( \eta \) are weighting factors)
### Progress of Understanding

#### Example dialogue:

**User**: “To Sapporo”  
**System**: “uh-huh”  
**User**: “From Tokyo”

#### Table: Flight Reservation Domain

| Origin  | Destination |
|---------|-------------|
| Tokyo   | Sapporo     |
| Sapporo | --          |

**Score**: $X > Y$  
**Plausible**: True

| Origin  | Destination |
|---------|-------------|
| Tokyo   | Sapporo     |
| Sapporo | --          |

**Score**: $Z < W$  
**Plausible**: True

---

**“To Sapporo”**  
[refer-origin: Sapporo]

**“From Tokyo”**  
[refer-origin: Tokyo]

---

**“To Sapporo”**  
[refer-destination: Sapporo]

**“From Tokyo”**  
[refer-origin: Tokyo]
Data Collection

• Corpus
  - 240 dialogues collected in the meeting room reservation domain
  - 26 dialogue act types
  - Vocabulary of 168 words
  - All the utterances transcribed and converted to dialogue acts

• Extraction of statistical information
  - Trigram probability of dialogue act types
  - Collocation probability
    • Classify the way of collocation into 64 classes
    • Use occurrence probability of each class
    • 17 classes found in the corpus
Implementation

• Scoring formula
  Score of the updated dialogue state =
  Score of the dialogue state before update
  + \( \beta \cdot \log \left( \frac{1}{\text{N-best-rank}} \right) \)
  + \( \gamma \cdot \log(\text{dialogue act type trigram probability}) \)
  + \( \delta \cdot \log(\text{collocation probability}) \)  \( \beta = \gamma = \delta = 1 \)

• Maximum number of dialogue states
  - Enables real-time processing by avoiding explosion of dialogue states

• Response generation
  - Rule-based response generation based on the highest-ranked dialogue state
Experiment (1)

• Verification of our approach
  - Collected 256 dialogues with the implemented system
  - 5-best recognition hypotheses as input
  - Maximum number of dialogue states: 15
  - Task completion rate: 88.3% (succeed in reservation within 5 minutes)

- Sufficiently high percentage of task completion rate suggests that system based on our approach works sufficiently
Experiment (2)

- Effectiveness of holding multiple dialogue states
  - **System1** (maximum number of dialogue states: 1)
    vs.
  - **System30** (maximum number of dialogue states: 30)
  - 224 dialogues collected with each system
  - System30 outperformed System1 both in task completion rate and task completion time
  - Average task completion time of System30 (95.86 sec.) was significantly shorter than that of System1 (107.66 sec.)

- Holding multiple dialogue states is effective
Conclusion

- A new discourse understanding method that
  - retains the ambiguity of a user utterance and resolves it by subsequent utterances
  - uses statistical information derived from dialogue corpora
- Experimental results show the validity of our approach