Robust Dialog Management with N-best Hypotheses Using Dialog Examples and Agenda

Cheongjae Lee, Sangkeun Jung, Gary Geunbae Lee
POSTECH
Outline

- Introduction
- Related Work
- Example-based Dialog Modeling
- Agenda Graph
- Greedy Selection with N-best Hypotheses
- Error Recovery Strategy
- Experiment & Result
- Conclusion & Discussion
Outline

- Introduction
- Related Work
  - Example-based Dialog Modeling
  - Agenda Graph
  - Greedy Selection with N-best Hypotheses
  - Error Recovery Strategy
- Experiment & Result
- Conclusion & Discussion
Introduction

- Real World of Spoken Dialog System

Spoken Dialog System

Speech-In

Noisy Environment

Speech-Out

Error Propagation
- ASR error
- SLU error
- DM error

User
Introduction

Problem

- Robustness issue of Spoken Dialog System
  - Very Important!
  - Making worse by Noisy Environments or Unexpected Inputs

![Diagram showing ASR, SLU, and DM with error handling steps]

- Noise reduction
- Adaptation
- N-best & lattice & CN

- Robust parsing
- Data-driven app.

- Error handling
- N-best support
Robust Dialog Management

- There are some approaches to develop the robust dialog manager.
  - 1) Error Handling
    - Ex) Error recovery strategies used to handle ASR/SLU errors [McTear et al., 2005, Torres et al., 2005, Lee et al., 2007].
  - 2) N-best Support
    - Ex) POMDP is superior to MDP by integrating N-best hypotheses into belief estimation [Williams and Young, 2007].

Our Goal

- To increase the robustness of Example-based Dialog Modeling [Lee et al., 2006]
  - by supporting N-best hypotheses with Agenda Graph
Related Work

- **Agenda & Task Model**
  - **Domain-specific Hierarchical Structure**
    - Manually designed in the target domain
      - Plan Tree (Collagen) \[Rich and Sinder, 1998\]
      - Dialog Task Specification (RavenClaw) \[Bohus and Rudnicky, 2003\]
      - User Goal Tree (HIS) \[Young et al., 2007\]
  - **Description**
    - Domain-specific rules in each node of the tree
      - Plan Recipe (Collagen)
      - Dialog Agent & Agency (RavenClaw)
      - Ontology Rules (HIS)

Example

- Bohus and Rudnicky, (2003), RavenClaw: Dialog Management Using Hierarchical Task Decomposition and an Expectation Agenda, EUROspeech.
- Young et al., (2007), The Hidden Information State Approach to Dialog Management, ICASSP.
Example-based Dialog Modeling

- Example-based Dialog Modeling (EBDM) Framework
  - Our previous work for generic dialog modeling.

*Lee et al., (2006), A Situation-based Dialogue Management using Dialog Examples, ICASSP.*

Architecture

**User Input**

**Dialog Corpus**

**Dialog Example Database**

**Dialog State Space**

*Lee et al., (2006), A Situation-based Dialogue Management using Dialog Examples, ICASSP.*
Problem

1) Only 1-best processing
   - Weak to ASR/SLU errors
   - Supporting N-best hypotheses is required.

2) Too flexible to track dialog state
   - It needs some heuristics to keep track and complete the task-oriented dialogs.

Solution

We can solve these problems with:

- N-best ASR hypotheses
- Agenda Graph as Heuristics
Agenda Graph

- Agenda graph $G=\{E, V\}$
  - It is represented by Directed Acyclic Graph (DAG)
    - Simply a way of encoding the domain-specific dialog control to complete the task.
  - Node = Possible intermediate subtasks
    - User goal state to achieve domain-specific subtask
    - Each node includes three components
      - State property
      - Precondition
      - Links to nodes at the subsequent turn.

- Edge = Possible dialog flows
  - For every edge $e_{ij}=(v_i, v_j)$
    - We defined a transition probability based on prior knowledge of dialog flows
Example Filtering

- Each node \( v \) should hold relevant dialog examples
  - Corresponding to user goal states by matching precondition rules
  - These are relevant and possible examples to appear in each node.
Discourse Interpretation Algorithm

- To determine the discourse state to keep track of dialog state
  - To consider how the current utterance can contribute to the current discourse purpose.

- This breaks down into five main cases.
  - **NEW_TASK**
    - Starting a new task to complete a new agenda
  - **NEW_SUB_TASK**
    - Starting a new subtask to partially shift focus
  - **NEXT_TASK**
    - Working on the next subtask contributing to the current agenda
  - **CURRENT_TASK**
    - Repeating or modifying the observed goal on the current subtask
  - **PARENT_TASK**
    - Modifying the observation on the previous subtask
Greedy Selection with N-best Hypotheses

- **Greedy Selection**
  - We should select the best node and the best example using N-best hypotheses by greedy selection.
    - Node Selection
    - Example Selection

- **Multi-Level Score Function**
  - We should propose a new score function that combines the parts of ASR, SLU, and DM.

*Overall Strategy*
Node Selection

\[ c^* = \arg \max_{h_i \in H, c_i \in C} \omega S_H (h_i) + (1 - \omega) S_D (c_i \mid S) \]

- where \( H \) is a list of n-best hypotheses and \( C \) is a set of candidate nodes to be generated by the discourse interpretation.

Hypothesis Score (Utterance-level score)

\[ S_H (h_i) = \alpha S_{rec} (h_i) + \beta S_{cont} (h_i) \]

- Recognition Score: a generalized n-gram LM score

\[ S_{rec} (h_i) = \frac{\text{lm}(h_i) - \text{lm}(h_n)}{\text{lm}(h_0) - \text{lm}(h_n)} \]

- Content Score: the degree of content coherence

\[ S_{cont} (h_i) = \begin{cases} 
N(C_{h_i}) / N(C_{prev}) & \text{if } C_{h_i} \subseteq C_{prev} \\
N(C_{h_i}) / N(C_{total}) & \text{if } C_{h_i} \not\subseteq C_{prev}
\end{cases} \]
**Greedy Selection with N-best Hypotheses**

- **Discourse Score (Discourse-level score)**
  - The degree to which candidate node $c_i$ is in focus with respect to the previous user goal state.
  - We can use each transition probability as discourse score.
    - **NEXT_TASK** case
      \[
      S_D(c_i \mid S) = P(c_i \mid c = top(S))
      \]
    - Other cases
      - Problem: No transition probability
      - Assumption: the transition probability may be lower than the **NEXT_TASK** case
        - Because an user utterance is highly correlated to the previous user goal state in task-oriented dialogs
        \[
        S_D(c_i \mid S) = \max\{P_{\min}(S) - \lambda \text{Dist}(c_i, c), 0\}
        \]
        where \[
        P_{\min}(S) = \min_{c_j \in \text{CHILD}(c=top(S))} \{P(c_j \mid c)\}
        \]
Greedy Selection with N-best Hypotheses

- **Example Selection**
  
  \[ e^* = \arg \max_{e_j \in E(c^*)} \omega S_{utter} (h, e_j) + (1 - \omega) S_{sem} (h, e_j) \]

- **Example Score**
  - *Utterance score*
    - Utterance similarity between hypothesis and example [Lee et al., 2006]
      
      \[ S_{utter} (h, e_j) = \delta S_{LSS} (w_h, w_{e_j}) + (1 - \delta) S_{DHS} (v_h, v_{e_j}) \]

  - *Semantic Score*
    - A dialog example is semantically closer to the current utterance if the example is selected with more query keys
      
      \[ S_{sem} (h, e_j) = \frac{\text{# of matching query keys}}{\text{# of total query keys}} \]
Error Recovery Strategy

- **Error Detection**
  - Out-of-coverage cases by noise or unexpected input
    - *No Agenda*
      - No possible nodes are selected by the discourse interpretation.
    - *No Example*
      - No dialog examples are retrieved by the EBDM.

- **Error Recovery**
  - Help message generation to easily rephrase [Lee et al., 2007]
    - *AgendaHelp* strategy  *Example*
    - *UtterHelp* strategy

*Lee et al., (2007), Example-based Error Recovery Strategy For Spoken Dialog System, ASRU.*
Outline

- Introduction
- Related Work
- Example-based Dialog Modeling
- Agenda Graph
- Greedy Selection with N-best Hypotheses
- Error Recovery Strategy
- Experiment & Result
- Conclusion & Discussion
Real User Evaluation

- Target Domain: Building Guidance for Intelligent Robots
  - Providing information about room and person
  - Navigating to the desired room

<table>
<thead>
<tr>
<th>Dialog Act</th>
<th>Main Goal</th>
<th>Slot</th>
<th>System Act</th>
</tr>
</thead>
<tbody>
<tr>
<td># Classes</td>
<td>9</td>
<td>13</td>
<td>9</td>
</tr>
</tbody>
</table>

Examples
- WH_QUESTION
- YN_QUESTION
- REQUEST
- STATEMENT
- SEARCH_LOC
- SEARCH_LOC_PHONE
- SEARCH_PER_MAIL
- GUIDE_LOC
- LOC_ROOM_NAME
- LOC_ROOM_TYPE
- LOC_ROOM_NUMBER
- PERSON_NAME
- INFORM (ROOM_NUMBER)
- INFORM (PERSON_NAME)
- SAY(YES_NO)
- GUIDE(ROOM)

AgendaGraph
ScenarioExample
Real User Evaluation

- Task Completion Rate
- Total Elapsed Time
- User Questionnaire

1. Reflecting Real World (+)
2. Human Factor - It looses objectivity (-)

Spoken Dialog System
Real User Evaluation

- **Experiment Environment**
  - Evaluator: 5 students
  - Size: 150 user utterances in 50 dialogs
  - HTK speech recognizer (WER = 21.03%)

### System Results

<table>
<thead>
<tr>
<th>System</th>
<th>#AvgUserTurn (turns per a dialog)</th>
<th>TCR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-best (-AG)</td>
<td>4.65</td>
<td>84.0</td>
</tr>
<tr>
<td>10-best (+AG,+ER)</td>
<td>4.35</td>
<td>90.0</td>
</tr>
</tbody>
</table>

**TCR**: Task Completion Rate  
**#AvgUserTurn**: Average number of user turns per a dialog  
**AG**: Using Agenda Graph  
**ER**: Using Error Recovery Strategy

Not enough!!!
Simulated User Evaluation

- **Simulated Interaction**
- **Noisy Environment** (+ASR error)

1. **Low Cost (+)**

Spoken Dialog System

- Task Completion Rate
- Total Elapsed Time
- Various Trial

2. **Consistent Evaluation**
- It guarantees objectivity (+)

1. **Not Real World (-)**

**Jung et al., (2008), An Integrated Dialog Simulation Technique for Evaluating Spoken Dialog Systems, COLING2008 Workshop (To be appeared).**
## Experiment & Result

### Simulated User Evaluation

#### 1000 Simulated Dialogs

<table>
<thead>
<tr>
<th>WER=10%</th>
<th>#AvgUserTurn (turns per a dialog)</th>
<th>TCR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-best (-AG,-ER)</td>
<td>5.85</td>
<td>81.01</td>
</tr>
<tr>
<td>1-best (-AG,+ER)</td>
<td>5.90</td>
<td>83.50</td>
</tr>
<tr>
<td>10-best (+AG,-ER)</td>
<td>6.43</td>
<td>92.66</td>
</tr>
<tr>
<td>10-best (+AG,+ER)</td>
<td>6.17</td>
<td><strong>93.87</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WER=20%</th>
<th>#AvgUserTurn (turns per a dialog)</th>
<th>TCR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-best (-AG,-ER)</td>
<td>5.85</td>
<td>67.94</td>
</tr>
<tr>
<td>1-best (-AG,+ER)</td>
<td>5.90</td>
<td>70.75</td>
</tr>
<tr>
<td>10-best (+AG,-ER)</td>
<td>6.55</td>
<td>90.85</td>
</tr>
<tr>
<td>10-best (+AG,+ER)</td>
<td>6.39</td>
<td><strong>91.75</strong></td>
</tr>
</tbody>
</table>

(a) Task Completion Rate  
(b) Best Rank Distribution
Agenda graph

- Cost of applying our methodology
- However, we can improve the previous EBDM framework by
  - 1) Rescoring n-best hypotheses
     - Multi-level score function - ASR, SLU, DM score
  - 2) Integrating heuristics into dialog model
     - Agenda graph - Heuristics to control dialog flow
  - 3) Generating *UtterHelp & AgendaHelp* strategy
     - Help messages about next subtask to complete the task.
1) How to reduce the cost of agenda?
   - Solution: Semi-automatic graph generation
     - Similar to learn the discourse structure of dialog corpus.
     - Clustering method
       - Dialog state clustering, dialog example clustering

2) How to estimate the optimal weight?
   - Solution: Dialog Simulation
     - Maximizing the task completion and minimizing the dialog length to complete the task
Thank you!
More detail idea in the paper!
References

- Agenda & Task Model
  - Collagen
  - RavenClaw
  - HIS
References

- **Example-based Dialog Modeling**

- **Error Handling**
Appendix List

I. RavenClaw Dialog Management
II. Example-based Dialog Modeling
III. Dialog Example Database
IV. Agenda Graph Example for Building Guidance Domain
V. AgendaXML
VI. Discourse Interpretation Algorithm
VII. Example of NEXT_TASK case
VIII. Greedy Selection Strategy
IX. AgendaHelp Strategy
X. UtterHelp Strategy
XI. Dialog Simulator Architecture
XII. Scenario Example
RavenClaw Dialog Management
- Dialog Task Specification

They have been used for various purposes.
- Ex) Dialog modeling, complexity reduction, domain knowledge, user simulation, ...

Bohus and Rudnicky, (2003), RavenClaw: Dialog Management Using Hierarchical Task Decomposition and an Expectation Agenda, EUROSPEECH
Appendix (II)

EBDM ARCHITECTURE

- Noisy Input (from ASR/SLU)
- Query Generation
  - Making SQL statement using Discourse History and SLU results.
- Example Search
  - Trying to search semantically close dialog example in DEDB given the current dialog state.
- Example Selection
  - Selecting the best example to maximize the utterance similarity measure based on lexical and discourse information.

EBDM PROCESSES

- Query Generation
  - Making SQL statement using Discourse History and SLU results.
- Example Search
  - Trying to search semantically close dialog example in DEDB given the current dialog state.
- Example Selection
  - Selecting the best example to maximize the utterance similarity measure based on lexical and discourse information.
Dialog Example Database (DEDB)

- Indexed by semantic and discourse features

<table>
<thead>
<tr>
<th>Turn #1 (Domain=Building_Guidance)</th>
<th>Dialog Corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>USER: 회의실이 어디지? (Where is the meeting room?)</td>
<td>USER: 회의실이 어디지? (Where is the meeting room?)</td>
</tr>
<tr>
<td>[Dialog Act = WH-QUESTION]</td>
<td>[System Action = inform(Floor)]</td>
</tr>
<tr>
<td>[Main Goal = SEARCH-LOC]</td>
<td></td>
</tr>
<tr>
<td>[ROOM-TYPE = 회의실 (conference room)]</td>
<td></td>
</tr>
<tr>
<td>SYSTEM: 3층에 교수회의실, 2층에 대회의실, 소회의실이 있습니다. (There are Professor Meeting Room on 3rd floor, and Conference Room, Meeting Room on 2nd floor.)</td>
<td></td>
</tr>
</tbody>
</table>

Semantically Indexing

Domain = Building_Guidance
Dialog Act = WH-QUESTION
Main Goal = SEARCH-LOC
ROOM-TYPE = 1 (filled), ROOM-NAME = 0 (unfilled), LOC-FLOOR = 0, PER-NAME = 0, PER-TITLE = 0
Previous Dialog Act = <s>, Previous Main Goal = <s>, Discourse History Vector* = [1,0,0,0,0]
User Utterance = ROOM_TYPE 이 어디지? (Where is ROOM-TYPE?)
System Action = inform(Floor)

* Discourse History Vector = [ROOM-TYPE, ROOM-NAME, LOC-FLOOR, PER-NAME, PER-TITLE]
Appendix (IV)

- Agenda Graph Example for Building Guidance Domain
AgendaXML

- This is described by XML scheme.

```xml
<agenda_graph domain="building_guidance">
  <state state_id="2">
    <property>
      <label>방 위치 검색*</label>
      <pos>DIALOG_ON</pos>
    </property>
    <precondition>
      <main_goal>SEARCH</main_goal>
      <target>LOC</target>
      <slot_status name="LOC_ROOM_NAME">1</slot_status>
    </precondition>
    <next>
      <state_id prob="0.15">4</state_id>
      <state_id prob="0.25">5</state_id>
      <state_id prob="0.60">6</state_id>
    </next>
  </state>
</agenda_graph>
```

*Search Location of Targeted Room (in English)
## Discourse Interpretation Algorithm

### Pseudo Codes

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
</table>
| **INTERPRET**(<S,H>,G) ≡ | \[
C \leftarrow \text{GENERATE}(<S,H>,G)
\]
| if \(|C| = 1\) then | return discourse state in C |
| else | \[c^* \leftarrow \text{SELECT}(<S,H>,C)\] |
| return | selected discourse state in \(c^*\) | |
| **GENERATE**(<S,H>,G) ≡ | return a union set of:
| i) **NEW_TASK**(<S,H>,G) |
| ii) **NEW_SUB_TASK**(<S,H>,G) |
| iii) **NEXT_TASK**(<S,H>,G) |
| iv) **CURRENT_TASK**(<S,H>,G) |
| v) **PARENT_TASK**(<S,H>,G) |
| **NEXT_TASK**(<S,H>,G) ≡ | \[
C \leftarrow \emptyset
\]
| E ← retrieved examples of H |
| foreach \(e \in E\) | foreach \(c_s \in \{\text{children of } \text{top}(S)|G\}\) |
| if \(e\) is an example of \(c_s\) | if \(e\) is an example of \(c_s\) |
| then | then |
| \(C \leftarrow C \cup \{c_s\}\) | \(C \leftarrow C \cup \{c_s\}\) |
| return \(C\) | return \(C\) |
| **SELECT**(<S,H>,C) ≡ | \[
c^* = \arg\max_{c \in C|S} \omega_S H(H)^+ + (1-\omega) S_D (c \in C|S)
\]
| return \(c^*\) | return \(c^*\) |
Discourse Interpretation Algorithm

- Ex) NEXT_TASK case
  - NEXT_TASK covers totally focused user behavior.
  - $v_4$, $v_5$, or $v_6$ are next subtasks to continue the current agenda.

When $v_4$ is selected as the best node.

Focus Stack ($t=i$)  
Focus Stack ($t=i+1$)
Overall Strategy

From User

<table>
<thead>
<tr>
<th>ASR</th>
<th>SLU</th>
<th>EBDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$</td>
<td>$u_1$</td>
<td>$s_1$</td>
</tr>
<tr>
<td>$w_2$</td>
<td>$u_2$</td>
<td>$s_2$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$w_n$</td>
<td>$u_n$</td>
<td>$s_n$</td>
</tr>
</tbody>
</table>

Focus Stack

Discourse Interpretation

<table>
<thead>
<tr>
<th>$v_1$</th>
<th>$v_2$</th>
<th>$v_3$</th>
<th>$v_4$</th>
<th>$v_5$</th>
<th>$v_6$</th>
<th>$v_7$</th>
<th>$v_8$</th>
<th>$v_9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>$e_2$</td>
<td>$e_k$</td>
<td>$v_6$</td>
<td>$v_3$</td>
<td>$v_4$</td>
<td>$v_6$</td>
<td>$v_8$</td>
<td>$v_9$</td>
</tr>
</tbody>
</table>
AgendaHelp strategy

The system could inform what the user achieves next by tracking the agenda.

- ex) SYSTEM: Next, you can do the subtask 1) Asking the room's role, or 2) Asking the office phone number, or 3) Selecting the desired room for navigation.
**UtterHelp strategy**

- If none of the dialog example were retrieved
  - The system gives in-coverage example of what the user could say at the current dialog state.

1) Find the possible examples based on the previous dialog example
2) Select the closest example
3) Generate help message with example's utterance

ex) SYSTEM: I am sorry that I could not understand. **You can say "Where is ROOM-NAME".**

1. Previous User Input
2. Previous Example
3. Current User Input
Appendix (XI)

- Dialog Simulator Architecture

![Diagram of Dialog Simulator Architecture]

- Argmax Selector
- Discourse Interpreter
- EBDM
- Error Handler
- SLU
- Intention Simulator
- Utterance Simulator
- ASR Simulator

Symbols:
- $a_m^*$, $c^*$, $a_e^*$
- $a_u$
- $w$
- $\tilde{a}_{u,1}$, $\tilde{a}_{u,2}$, ..., $\tilde{a}_{u,n}$
- $\tilde{w}_1$, $\tilde{w}_2$, ..., $\tilde{w}_n$
## Scenario Example

<table>
<thead>
<tr>
<th>Turn</th>
<th>Utterance (in Korean)</th>
<th>Utterance (in English)</th>
</tr>
</thead>
<tbody>
<tr>
<td>USER₁</td>
<td>회의실은 어디죠? (Hoe-uisil-eun eodijyo?)</td>
<td>Where is the meeting room?</td>
</tr>
<tr>
<td>SYS₁</td>
<td>교수 회의실은 3층에, 대회의실, 중회의실은 2층에 있습니다. (Gyosu hoe-uisil-eun 3-cheung-e, dae-hoe-uisil, jung-hoe-uisil-eun 2-cheung-e iss-seubnida.)</td>
<td>There are Professor’s meeting room on 3rd floor, conference room and meeting room on 2nd floor.</td>
</tr>
<tr>
<td>USER₂</td>
<td>대회의실은 몇 명이 들어가죠? (Daehoe-uisil-eun myeoch myeong-i deul-eogaj yo?)</td>
<td>How many people can the conference room accomodate?</td>
</tr>
<tr>
<td>SYS₂</td>
<td>대회의실은 20명 내외가 회의를 할 수 있는 곳 입니다. (Daehoe-uisil-eun 20myeong nae-oega hoe-uleul hal su isseun gos-ibnida.)</td>
<td>The conference room is the meeting room for about 20 people.</td>
</tr>
<tr>
<td>USER₃</td>
<td>그럼 대회의실로 안내해주세요. (Geuleom daehoe-uisil-lo annae-hae-juse-yo.)</td>
<td>Let me bring to the conference room.</td>
</tr>
<tr>
<td>SYS₃</td>
<td>대회의실로 안내해드리겠습니다. (Daehoe-uisil-lo annae-hae-deuligess-seubnida.)</td>
<td>I will navigate to the conference room.</td>
</tr>
</tbody>
</table>