Data Driven Approaches for Spoken Dialog Processing

Gary Geunbae Lee, Ph.D., Professor
Dept. CSE, POSTECH
Europe-Korea SDS workshop
Contents

• **SLU**
• DM
• On-going researches
Ubiquitous spoken dialog interface?

Telematics Dialog Interface (POSTECH, LG, DiQuest)
What’s hard – ambiguities, ambiguities, all different levels of ambiguities

John stopped at the **donut store** on his way home from work. He **thought** a coffee was good **every few hours**. But it turned out to be **too expensive** there. [from J. Eisner lecture note]

- donut: To get a donut (doughnut; spare tire) for his car?
- Donut store: store where donuts shop? or is run by donuts? or looks like a big donut? or made of donut?
- From work: Well, actually, he stopped there from hunger and exhaustion, not just from work.
- Every few hours: That’s how often he thought it? Or that’s for coffee?
- it: the particular coffee that was good every few hours? the donut store? the situation
- Too expensive: too expensive for what? what are we supposed to conclude about what John did?
“I need a flight from Washington DC to Denver roundtrip”

ORIGIN_CITY: WASHINGTON
DESTINATION_CITY: DENVER
FLIGHT_TYPE: ROUNDTrip
Spoken Language Understanding
Spoken language understanding (SLU) is to map natural language speech to frame structure encoding of its meanings.

- <frame domain='ATIS'>
  - <utt>Show me flights from Denver to New York on Nov. 18th</utt>
  - <slot type='DA' name='Show_Flight'/>
  - <slot type='NE' name='FROM.CITY'>Denver</slot>
  - <slot type='NE' name='TO.CITY'>New York</slot>
  - <slot type='NE' name='MONTH'>Nov.</slot>
  - <slot type='NE' name='DAY_NUMBER'>18th</slot>
</frame>

- <frame domain='EPG'>
  - <utt>I want to watch LOST</utt>
  - <slot type='DA' name='Search_Program'/>
  - <slot type='NE' name='PROGRAM'>LOST</slot>
</frame>
Non-local Features

Algorithm 1 Trigger Selection

1: Initialize training data $\mathcal{D}$ with local features and a trigger set $t = \emptyset$
2: while $t$ is increased do
3: Learn a ME classifier on $\mathcal{D}$: $\Lambda \leftarrow \text{TrainME}(\mathcal{D})$
4: Make candidates: $g \leftarrow \text{GenerateTriggers}(\mathcal{D}, \Lambda)$
5: Optimize $\mu$: $\hat{\mu} \leftarrow \text{OptimizeGain}(\mathcal{D}, g, \Lambda)$
6: Select triggers: $g^* \leftarrow \text{SelectTrigger}(g, \hat{\mu})$
7: Update training data: $\mathcal{D} \leftarrow \text{UpdateData}(\mathcal{D}, g^*)$
8: Update a trigger set: $t \leftarrow t \cup g^*$
9: end while
10: return $\mathcal{D}, t$

• Outline of Trigger Selection Algorithm
Non-local Features

<table>
<thead>
<tr>
<th>Inducer type</th>
<th>Time</th>
<th># features</th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Induction</td>
<td>Training</td>
<td>Trigger</td>
</tr>
<tr>
<td>None</td>
<td>3:47:43</td>
<td>5:12:11</td>
<td>1,320</td>
</tr>
<tr>
<td>Approx. 1</td>
<td>0:47:30</td>
<td>2:14:02</td>
<td>800</td>
</tr>
<tr>
<td>Approx. 2</td>
<td>0:11:10</td>
<td>1:34:37</td>
<td><strong>467</strong></td>
</tr>
<tr>
<td>Approx. 1+2</td>
<td><strong>0:05:30</strong></td>
<td><strong>1:29:26</strong></td>
<td>608</td>
</tr>
</tbody>
</table>

Approx. 1: Only Non-local
Approx. 2: ME-based

- Learning & Precision-Recall Curves
  on Communicator Data
Joint SLU

\[ p(y, z|x; \Theta) = \frac{1}{Z(x)} \exp \left( \phi(x, y, z) \right) \]

\[ \phi(x, y, z) = \sum_{t} \sum_{k} \left\{ \lambda_k f_k(y_{t-1}, y_t, z) + \mu_k g_k(y_t, x) \right\} + \sum_{k} \nu_k h_k(z, x) \]

\[ f_k(y_{t-1}, y_t, z) = f_k^1(y_{t-1}, y_t) \cdot f_k^2(y_t, z) \]

• Search Space
• Factor Graph of Triangular-chain CRF
### Joint SLU

<table>
<thead>
<tr>
<th>Method</th>
<th>Air-Travel</th>
<th>Robot-Café</th>
<th>Telebank</th>
<th>TV-EPG</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DA NE</td>
<td>DA NE</td>
<td>DA NE</td>
<td>DA NE</td>
<td>DA NE</td>
</tr>
<tr>
<td>Indep</td>
<td>91.29 86.54</td>
<td>88.59 96.11</td>
<td>94.53 88.55</td>
<td>96.84 96.29</td>
<td>92.81 91.87</td>
</tr>
<tr>
<td>Cascade1</td>
<td>92.76 -</td>
<td>89.40 -</td>
<td>96.58 -</td>
<td>96.91 -</td>
<td>93.91 -</td>
</tr>
<tr>
<td>Cascade2</td>
<td>- 88.96</td>
<td>- 96.13</td>
<td>- 88.85</td>
<td>- 96.76</td>
<td>- 92.68</td>
</tr>
<tr>
<td>Rerank1</td>
<td>92.85 86.36</td>
<td>87.74 96.00</td>
<td>96.57 88.29</td>
<td>96.86 96.31</td>
<td>93.51 91.74</td>
</tr>
<tr>
<td>Rerank2</td>
<td>92.18 88.94</td>
<td>88.64 96.13</td>
<td>94.65 88.88</td>
<td>96.60 96.51</td>
<td>93.02 92.62</td>
</tr>
<tr>
<td>Joint</td>
<td>93.77 88.24</td>
<td>89.47 96.12</td>
<td>96.92 88.91</td>
<td>96.99 96.94</td>
<td>94.29 92.55</td>
</tr>
</tbody>
</table>

- **Log-likelihood for DA, NE and Joint Optimization**

- **A description of four dialog data sets**

<table>
<thead>
<tr>
<th>Data set</th>
<th># of utt</th>
<th>DA</th>
<th>NE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air-Travel</td>
<td>1,178</td>
<td>12</td>
<td>54</td>
</tr>
<tr>
<td>Robot-Café</td>
<td>626</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>Telebank</td>
<td>2,239</td>
<td>25</td>
<td>17</td>
</tr>
<tr>
<td>TV-EPG</td>
<td>1,917</td>
<td>16</td>
<td>7</td>
</tr>
</tbody>
</table>
Contents

• SLU
• DM
• On-going researches
Example-based Dialog Management
Introduction

• Pipeline Architecture for SDS

• Dialog Management
  – A central component to select correct system actions based on observed evidences and inferred beliefs.

• Goal
  – Dialog modeling for practical deployment of multi-domain dialog system
Introduction

• Related Works
  – **Knowledge-based approach**
    • Regularized Language & Grammar
    • Hand-crafted Rules & Finite State Automaton
    • Pros & Cons
      – This approach have been deployed in many practical application. (+)
      – Can be controlled by developers (+) (-)
      – Not good for domain portability and flexibility. (-)
  – **Data-driven approach**
    • A stochastic modeling using reinforcement learning [Levin et al., 2000]
      – Formalization : fully or partially observable Markov Decision Processes.
    • Pros & Cons [Paek, 2006]
      – The training is done automatically (+)
      – Theoretically formalized (+)
      – Time-consuming corpus collection & annotation (-)
      – High Complexity & Hardly practical deployment (-)
Example-Based Dialog Modeling

• Concept
  – From Example-Based Machine Translation (EBMT)
    • Nagao (1984) introduced the EBMT methodology.
      – Source Sentence $\rightarrow$ Processing $\rightarrow$ Finding Similar Sentence $\rightarrow$ Retrieval Target Sentence
  – The idea of EBMT can be extended to determine the next system actions by finding the similar dialog example in the corpus.
  – *Example-Based Dialog Modeling (EBDM) [Lee et al., 2006; Lee et al., 2007]*
    • A dialog model in which the system action can be selected by the similar user utterance within dialog corpus.
      – User Utterance $\rightarrow$ Processing $\rightarrow$ Finding Similar Utterance $\rightarrow$ Retrieval System Action
  • Dialog Example Database (DEDB)
    – Indexed by state variables chosen by system designer.
Example-Based Dialog Modeling

- Indexing and Querying
  - DEEB is semantically indexed and queried to generalize the data.

<table>
<thead>
<tr>
<th>Goal-Oriented Dialog Corpus (Domain = Navigation)</th>
<th>DEEB (Dialog Example Database)</th>
</tr>
</thead>
</table>
| #1  
  **User:** Where is the Korean restaurants?  
  [Dialog Act = wh-question]  
  [Main Goal = search-location]  
  [LOC_TYPE = Korean restaurant]  
  **System:** There are A, B, and C in D and E and F in G.  
  [System Action = inform(name,address)] |  
  **User Utterance** = Where is the LOC_TYPE?  
  Domain = navigation  
  Dialog Act = wh-question  
  Main Goal = search-location  
  LOC_TYPE = 1 (filled)  
  LOC_ADDRESS = 0 (unfilled)  
  LOC_NAME = 0  
  ROUTE_TYPE = 0  
  Previous Dialog Act = <s>  
  Previous Main Goal = <s>  
  Discourse History Vector = [1,0,0,0]  
  System Action = inform(name,address) |
| #2  
  **User:** Let me go the A in D.  
  [Dialog Act = request]  
  [Main Goal = guide]  
  [LOC_NAME = A]  
  [LOC_ADDRESS = D]  
  **System:** Ok. You selected the A in D.  
  [System Action = select(name,address)]  
  **System:** Choose the route type of the fastest or the easiest path.  
  [System Action = specify(rotue_type)] |  
  **User Utterance** = Let me go the LOC_NAME in LOC_ADDRESS.  
  Domain = navigation  
  Dialog Act = request  
  Main Goal = guide  
  LOC>Type = 0  
  LOC_NAME = 1  
  LOC_ADDRESS = 1  
  ROUTE_TYPE = 0  
  Previous Dialog Act = wh-question  
  Previous Main Goal = search-location  
  Discourse History Vector = [1,1,1,0]  
  System Action = select(name,address); specify(route_type) |

* Discourse History Vector = [LOC_TYPE, LOC_ADDRESS, LOC_NAME, ROUTE_TYPE]
Example-Based Dialog Modeling

• Relaxation
  – Once there is no example, dialog experts have some relaxation strategies according to the genre and domain of dialog.
  • State can be approximated by relaxing particular state variables for avoiding data sparseness.

• Utterance Similarity
  – Select the best one among the retrieved dialog examples
  • Considering the lexical and discourse history information

Current User Utterance
Where is the LOC_TYPE?
Discourse History Vector : [1,0,0,0]

Retrieved Examples
Where is the LOC_TYPE in LOC_ADDRESS?
Discourse History Vector : [1,1,0,0]

Let me know where the LOC_TYPE is?
Discourse History Vector : [1,0,0,0]

Lexico-Semantic Similarity (by edit distance)
Discourse History Similarity (by cosine measure)
Example-Based Dialog Modeling

- User’s Utterance
  - Domain Expert
  - User Intention
  - Semantic Frame
  - Discourse History

- Dialogue Corpus
  - Automatic Indexing

- Dialogue Example DB

- System Responses
  - Best Dialogue Example
  - Utterance Similarity
    - Lexico-semantic Similarity
    - Discourse history Similarity

- Query Generation
  - Retrieval
  - Dialogue Examples

- Tie-breaking
User: When do the KBS dramas start?

Linguistic Analysis

When/WRB do/VBP the/DT KBS/NNP drama/NN start/VB
Last Word = start | Last Marker = ? | Last Tag = VB
Last Verb = start | Last Noun = drama | First Noun = KBS

Semantic Analysis

Dialog Act = WH-QUESTION
Main Goal = SEARCH

Keyword Analysis

Best Keyword = drama | Second Keyword = start
Best Class = EPG | Second Class = Navigation

Agent Spotter

Task Agent

Domain Spotter

EPG domain
Experimental Results

• Spotter Evaluation
  – Domain Spotter

<table>
<thead>
<tr>
<th>Feature Set</th>
<th>Accuracy (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (only Linguistic Feature)</td>
<td>96.65</td>
<td>-</td>
</tr>
<tr>
<td>+ Pragmatic Feature</td>
<td>97.57</td>
<td>0.004</td>
</tr>
<tr>
<td>+ Keyword Spotting Feature</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TFxIDF</td>
<td>97.47</td>
<td>0.533</td>
</tr>
<tr>
<td>Salience</td>
<td>97.75</td>
<td>0.449</td>
</tr>
</tbody>
</table>

– Agent Spotter

<table>
<thead>
<tr>
<th>Feature Set</th>
<th>Accuracy (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (only Linguistic Feature)</td>
<td>84.74</td>
<td>-</td>
</tr>
<tr>
<td>+ Pragmatic Feature</td>
<td>85.23</td>
<td>0.016</td>
</tr>
<tr>
<td>+ Keyword Spotting Feature</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TFxIDF</td>
<td>89.86</td>
<td>0.004</td>
</tr>
<tr>
<td>Salience</td>
<td>87.11</td>
<td>0.006</td>
</tr>
</tbody>
</table>
Experimental Results

• Dialog Modeling Evaluation
  – Success Turn Rate

  ![Graph showing success turn rate for different domains.]

  - ChatBot
  - Car Navigation
  - Weather Information
  - EPG

  Success Turn Rate (%)

  - Task Completion Rate

<table>
<thead>
<tr>
<th>Application Domain</th>
<th>TCR(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPG</td>
<td>80</td>
</tr>
<tr>
<td>Weather Information</td>
<td>88</td>
</tr>
<tr>
<td>Car Navigation</td>
<td>90</td>
</tr>
</tbody>
</table>
Error Recovery System

Noisy Input (from ASR/SLU) → Query Generation → Example Search → Example Selection

Relaxation Strategy

S = (# of Examples, # of Contents, # of Slots)

Error Detected = YES

A = {HelpType, Content, Template}

Error Detected = NO

Containing some errors.
Error Detection of EBDM

• **No Example**
  – No dialog example is retrieved by both exact and partial matches are used.
    • OOV, OOU, and unexpected input may be out-of-example.
    • This is a potential error because out-of-example is hard to understand and manage.

• **No Content**
  – No information is retrieved using the slot values of the current dialog frame.
    • In this case, the user does not know the slot values of interest or unexpected errors in SLU module occurs.

• **No Slot**
  – The understanding module cannot extract any slot value from the user utterance
    • The utterance without slot information may be erroneous when *No Example* is detected.
    • Because, in goal-oriented dialogs, most utterances contain slot information to search contents.
Error Recovery Strategy

• Our Goal
  – The system should give help messages to learn what and how to say.

• Error Recovery Strategy
  – *UtterHelp*
    • A help message for possible utterance template
      – Triggered by No Example detection
    • The system gives an example template of what the user could say at this situation.
  – *InfoHelp*
    • A help message for domain content database
      – Triggered by No Content detection
    • The system recommends relevant contents which can be retrieved with the current dialog frame
  – *UsageHelp*
    • A longer help message for a system usage
      – Triggered by all (No Example/No Content/No Slot) detection
    • This system provides guidelines of using the system
Recovery Example

**UtterHelp**

**User:** Please inform me a category of the restaurants that serves Korean food.  
**ASR output:** Please me a car of a restaurant that Korean food  
**SLU output:** [REQUEST, GUIDE_LOC, FOOD-TYPE=Korean food]  
**[Error Detection: No Example]**  
**System:** You can say "Please give me a category of the restaurants that serve [FOOD-TYPE]" to search restaurants of [FOOD-TYPE].

**InfoHelp**

**User:** Please inform me a category of the restaurants that serves Korean food.  
**ASR output:** Please inform me a category of restaurant that Korean fish  
**SLU output:** [REQUEST, SEARCH_LOC, FOOD-TYPE=Korean fish]  
**[Error Detection: No Content]**  
**System:** There is no content on "Korean fish". You can say "Korean food", "Korean beef", and "Japanese fish" as the slot of [FOOD-TYPE].
Experimental Result

• **User Evaluation** *[Lee et al., 2007]*
  
  - Car navigation domain for 4 slot-filling dialog
    - Provides support for the information and selection of the desired destination in Pohang city.
    - About 160 places names related to the restaurants, hospital, etc.
  
  - 10 novice users with prior knowledge about the system
    - 50 dialogs of evaluation set (5 dialogs for each user)
  
  - Measure
    - TCR = Task Completion Rate
    - $\#AvgTurn = \# \text{of user utters} / \# \text{of dialogs}$
    - $\#DetErr = \# \text{of detected errors} / \# \text{of dialogs}$

<table>
<thead>
<tr>
<th>System</th>
<th>Input</th>
<th>TCR (%)</th>
<th>$#AvgTurn$</th>
<th>$#DetErr$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recovery (-)</td>
<td>Text</td>
<td>84</td>
<td>5.16</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td>Speech</td>
<td>76</td>
<td>5.68</td>
<td>1.56</td>
</tr>
<tr>
<td>Recovery (+)</td>
<td>Text</td>
<td>92</td>
<td>5.88</td>
<td>2.12</td>
</tr>
<tr>
<td></td>
<td>Speech</td>
<td>84</td>
<td>6.28</td>
<td>2.12</td>
</tr>
</tbody>
</table>
Dialog for EPG (POSTECH)

Unified Chatting and Goal-oriented Dialog (POSTECH)
Contents

• SLU
• DM
• On-going researches
Context-sensitive Chatbot
Advanced ChatBot

• Discourse-Sensitive ChatBot
  – Most of chatbot cannot reflect the discourse context of previous user turns. [ALICE, 2002; B. Shawar, 2002]

Context-Free ChatBot

<table>
<thead>
<tr>
<th>USER</th>
<th>Did you watch a Korean team’s game yesterday?</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYSTEM</td>
<td>What game?</td>
</tr>
<tr>
<td>USER</td>
<td>How come you didn’t know that?</td>
</tr>
<tr>
<td>SYSTEM</td>
<td>Well, I don’t know.</td>
</tr>
</tbody>
</table>

Obscure answering

Context-Sensitive ChatBot

<table>
<thead>
<tr>
<th>USER</th>
<th>Did you watch a Korean team’s game yesterday?</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYSTEM</td>
<td>What game?</td>
</tr>
<tr>
<td>USER</td>
<td>How come you didn’t know that?</td>
</tr>
<tr>
<td>SYSTEM</td>
<td>I am not interested in sports.</td>
</tr>
</tbody>
</table>

What’s better?

More intelligent answering
Discourse History in ChatBot

 USER : I had a date last week
 Dialog Act = statement-non-opinion
 Main Goal = perform
 Agent = chat
 Domain = friend
 date = last week
 event = date
 SYSTEM : Did you like her?

 USER : Well, I think so, but.
 Dialog Act = statement-opinion
 Main Goal = judge
 Agent = chat
 Domain = friend
 SYSTEM : but, what?

 USER : I don’t know her real feeling.
 Dialog Act = statement-opinion
 Main Goal = know
 Agent = chat
 Domain = friend
 theme = real feeling
 SYSTEM : Why? What happen?
POSTECH Chatbot Demo
Multimodal Dialog Management
Multi-Modal Dialog

• A system which supports human-computer interaction over *multiple different input and/or output modes*.
  – Input: voice, pen, gesture, face expression, etc.
  – Output: voice, graphical output, etc.

• Applications
  – GPS, Information guide system, Smart home control, etc.

• Task performance and user preference for multi-modal over speech interfaces [Oviatt et al., 1997]
  – 10% faster task completion,
  – 23% fewer words,
  – 35% fewer task errors,
  – 35% fewer spoken disfluencies

Hard to represent using only uni-modal !!
Multi-Modal Dialog System Architecture

• Components of multi-modal dialog system [Chai et al., 2002]

![Diagram showing components of multi-modal dialog system](image)
N-best re-ranking for improving speech recognition performance

• Using multi-modal understanding feature [Kim et al., 2007]

Speech: bring this to here

Pen

ASR: bring his to here

SLU: Speech Act: request
Main Goal: move
Component Slots:
Target.Loc: here

Missing the slot!!!

→ Source.item: this
Experimental Result

- Experimental Result [Kim et al., 2007]
  - Word error rate
    - Relative error reduction rate: 7.95 (%)
    - Re-ranking model has significantly smaller word error rates than that of baseline system. \( p < 0.001 \)
  - Concept error rate
    - Relative error reduction rate: 10.13 (%)
    - Re-ranking model has significantly smaller concept error rates than that of baseline system. \( p < 0.01 \)

<table>
<thead>
<tr>
<th></th>
<th>WER (%)</th>
<th>CER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>17.74</td>
<td>14.28</td>
</tr>
<tr>
<td>+ Speech recognizer features</td>
<td>17.38</td>
<td>13.81</td>
</tr>
<tr>
<td>+ SLU features</td>
<td>16.43</td>
<td>13.11</td>
</tr>
<tr>
<td>+ Multi-Modal reference resolution features</td>
<td>16.33</td>
<td>12.83</td>
</tr>
</tbody>
</table>
Multi-Modal Dialog Management Using Hidden Information State Manager

- Hidden Information State Dialog Manager for Multi-Modal Dialog System [Kim et al., 2007]
  - POMDP based dialog manager
    - Uncertainty inherent framework
    - Maintains probability distribution of dialog states

- Scaling POMDPs for dialog system
  - The state space of a practical dialog system is very large.
    - E.g.) 3-city tourist domain
      - \( n( b(su, au, sd) ) = 6 \) user goals * 18 user acts * 18 dialog states = 1944 states
  - Hidden information state dialog manager...
    - Groups the equivalent states to a partition
    - Generates system action hypotheses according to each partition
POSTECH multimodal Dialog System Demo
Dialog Studio
Introduction to Dialog Studio

• Motivation
  • Dialog system development and maintenance involves
    • System Tutoring
    • Model adaptation
    • Model Synchronization
      • ASR ↔ SLU ↔ DM
    • Human Effort & Time reduction
    • Dialog Simulation
      • For auto dialog evaluation
      • For automatic massive corpus building
      • For finding flaws of the dialog system
Experiments – EPG Domain

![Bar chart showing comparison between Not Using DialogStudio and DialogStudio in various stages of experiment.

- SLU corpus Annotation
- Dialog Example Annotation
- Preparing New Models
- SLU Tuning
- DM Tuning
- Knowledge Management Tuning

Ex.1
Ex.2
Ex.3]
Dialog Simulation for SDS

Simulated User
(Speech/Text/Action)

Spoken Dialog Management System
Where can we use dialog simulation?

- **Strategy Learning**
  - State Exploration + Reinforcement Learning
    - POMDP

- **Evaluation**
  - Strategy Evaluation
  - DM Performance Evaluation

- **Corpus Expansion**
  - User Side corpus expansion
    - SLU corpus
  - User + System Side corpus expansion
    - Dialog corpus
    - Small corpus → Large corpus
Intention/Utterance Simulation

Discourse Info.

User Intention

Semantic Frame

User Utterance Simulation

User Semantic Frame

<table>
<thead>
<tr>
<th>Dialog_Act</th>
<th>Request</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main_Goal</td>
<td>Move_Channel</td>
</tr>
<tr>
<td>Component.[Genre]</td>
<td>--</td>
</tr>
</tbody>
</table>

Generated User Utterance

<table>
<thead>
<tr>
<th>Generated User Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td>그냥 [genre] 프로 좀 보자</td>
</tr>
<tr>
<td>야 [genre] 좀 들어</td>
</tr>
<tr>
<td>[genre] 좀 들어</td>
</tr>
<tr>
<td>그냥 [genre] 프로그램 보자</td>
</tr>
<tr>
<td>그럼 그냥 [genre] 나 보자</td>
</tr>
<tr>
<td>에이 [genre] 프로 좀 보자</td>
</tr>
</tbody>
</table>

User Utterances
Linear CRF Model for Intention Modeling

- Assumption
  - An user utterance has only one intention
- $U_I$: User Intention
- $DI$: Discourse Information
  - Previous System Response + Discourse History
Structure Tags: Component Slot Names + Part of Speech Tags
S: member of Structure Tags given space
W: member of vocabulary given space

Utterance generation

Dialog_Act _X_ Main_Goal

Structure Tag Transition

Emission Prob.

Sampling

Sampling

Sampling

Sampling

Sampling
Naturalness

- **BLEU (N-gram Precision)**

\[
p_n = \frac{\sum_{ngram \in G} Count_{clip}(ngram)}{\sum_{ngram \in G} Count(ngram)},
\]

\[Count_{clip}(ngram) = \min(\text{count, maximum reference count}).\]

\[BP = \begin{cases} 
1 & \text{if } g > r \\
e^{(1-r/g)} & \text{if } g \leq r 
\end{cases}, \quad \text{BLEU} = BP \cdot \exp\left(\sum_{n=1}^{N} \frac{1}{N} \log p_n\right).
\]
USER : What’s the phone number of ‘Gajokjung’?
SYS : The phone number of ‘Gajokjung’ placed in ‘Daeidong’ is 275-7543
USER : OK. Let’s go to ‘Daeidong’
SYS : You chose ‘Gajokjung’ in ‘Daeidong’. Please select the way of path finding. You can select ‘Shortest path’ and ‘Optimal path’.
USER : Shortest path.
SYS : You selected ‘Shortest path’. May I guide you to ‘Gajokjung’ in ‘Daeidong’ in shortest path way?
USER : Ok. Thank you.
POSTECH Dialog Studio Demo
English Tutoring Dialog Management
Dialogue System for English Conversation Tutoring

• Motivation
  – Huge demand for education of English conversation, but “One-to-one human tutoring is very expensive”
  – Spoken dialogue system can be used instead at low cost

• Research Issues
  – Speech recognition for the non-native
  – Language understanding for the non-native
  – Flexible dialogue modeling for diverse conversation themes
  – Feedback generation about performance of students
Research Issues

• Language understanding for the non-native
  – Improve robustness by adopting statistical methods
  – Use features designed for common errors of non-native speakers

• Flexible dialogue modeling for diverse conversation themes
  – Improve scalability by adopting data-driven methods
  – Example-based dialogue management can support various domains easily

• Feedback generation about performance of students
  – Implicit tutoring helps students learn appropriate expressions without stress
  – Explicit tutoring helps students continue the conversation when they do not know what to say
  – Final feedback indicate week points of an individual student
System architecture

User Utterance

ASR

SLU

DM

NLG

TTS

Semantic frame

System action

System Utterance

Domain experts

SupervisorExpert

ImmigrantExpert

TransportExpert

RestaurantExpert

Assessment data
Implicit Tutoring

• Even though the expressions of students are not perfect, the system can provide sound examples on screen.

Diagram:
- Dialog situation
- Dialogue Example DB
- Dialog Example Match
- Tutoring Example DB
- Ask user to repeat
- Examples
- Best Example

Lexical similarity
Explicit Tutoring

- When students do not know what to say, the system provides explicit tutoring so that students can continue the conversation.
Assessment

- Assessment
  - Collect assessment data for learning session and calculate final score to give feedback
  - Feedback message reflects individual perspective to improve
- Criteria
  - Mission Completion
  - Elapsed Time
  - Utterance Suitability
  - Help Frequency
  - Filler Word Frequency
• English Tutoring Dialog System Demo (POSTECH)
Automatic Content Feeding For Dialog Management
Information Sources

- Information Access Agent
- KB Access Module
- Question Answering Module
- Knowledge-bases (RDB or Ontology)
- Inferencing
- Ontology building
- WEB
Schema Design for Ontology Building

- Example ontology schema
Automatic Instance Population

Lost
- hasSurface: ABC
- hasName: “Lost”
- hasChannel: ABC
- hasGenre: Drama
- hasPerson: Kate, Kim, Fox
- hasWeekday: Sat, Sun
- hasStartTime: 20071215 13:00:00, 20071215 14:00:00
- hasEndTime: 20071216 13:00:00, 20071216 14:00:00

ABC
- hasName: “American Broadcasting Company”
- hasSurface: “ABC”
- hasName: “ABC”

dur1
- hasStart: 20071215 13:00:00
- hasDuration: dur1

dur2
- hasStart: 20071216 13:00:00

Kate
- hasFirstName: Yunjin
- actsAs: Kim

Fox
- hasFirstName: Tim

Sun
- hasFirstName: Sun

"Lost"
- hasName: “The Lost”
- hasName: "The Lost"
- hasName: “lost”

"The Lost"
- hasName: “The Lost”
- hasName: “The Lost"
- hasName: “lost”
Inferencing for Dialog system

• We cannot show the adequate answers for some utterances with RDB knowledge bases
  – Let’s watch Wayne Rooney’s game (In English Premier League, football)
  – Who is Sun’s husband? (In “Lost”, drama)
  – Let’s watch the game of Yankees (In Major League, baseball)

• RDB do not have the detail knowledge of the specific domain

• We need to inference to get the right result for the queries
Querying with Inference Engine

Let’s watch Wayne Rooney’s game

SLU

Wayne Rooney: Person name

Query Generation

SELECT ?match ?entry ?channel
FROM <sportsOntology.owl#>
WHERE { ?match owl:hasMonth owl:Dec .
  ?match owl:hasDay owl:d_12 .
  owl:Rooney owl:isMemberOf ?t .
  ?match owl:hasTeam ?t .
  ?match owl:hasEntry ?entry
  ?match owl:hasChannel ?channel
}

Result

<table>
<thead>
<tr>
<th>Match</th>
<th>entry</th>
<th>Channel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feb 5 ManU vs Chelsea</td>
<td>football</td>
<td>KBS</td>
</tr>
</tbody>
</table>
Knowledge Acquisition & Expansion

• Previous Approaches
  – Manually managed by human experts
  – Reliable (+)
  – High-cost (-)

• Automatic Knowledge Management
  – Based on the information extraction form natural language documents
Relation Extraction

- A task that detects and classification relations between named-entities

Hillary Clinton moved to New York last year.

AT.Residence

Person

Geo-Political Entity
Wayne Rooney currently plays for the English Premier League club Manchester United.
Semi-supervised Information Extraction

• The overall procedure
  a. Prepare seed data set S and unlabeled document D.
  b. Find the sentences containing lexical data in S in the document D.
  c. Extract and generalize context pattern P from the selected sentences in step b.
  d. Measure the score of each pattern P, and rank it.
  e. Extract more data S’ using high-ranked patterns, and add S’ to S
  f. Repeat from step b to step e.
the character `<ROLE>` portrayed by `<ACTOR>` in the television series `<PROGRAM>` is

character Michael Scofield portrayed by Wentworth Miller in the TV series Prison Break is
Information Extraction for Multiple Arguments

Seed Data → n arguments

Seed Data → Extracting Context Patterns → Relation Extraction

Seed Data → Extracting Context Patterns → Relation Extraction

Seed Data → Extracting Context Patterns → Relation Extraction

Seed Data → Extracting Context Patterns → Relation Extraction

Seed Data → Extracting Context Patterns → Relation Extraction

Validation & Integration

Results → n arguments
Experimental Results

• Named-Entity Recognition [Kim et al. 2007]
  – Data
    • CU-Communicator corpus
    • 13,983 utterances
  – Target
    • The three most frequently occurring semantic categories in the corpus
    • CITY_NAME, MONTH, DAY_NUMBER
  – Seed
    • One-tenth of the entities in the corpus
  – Result

<table>
<thead>
<tr>
<th>Category</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>CITY_NAME</td>
<td>91.30</td>
<td>86.83</td>
<td>89.01</td>
</tr>
<tr>
<td>MONTH</td>
<td>98.98</td>
<td>87.24</td>
<td>92.74</td>
</tr>
<tr>
<td>DAY_NUMBER</td>
<td>92.00</td>
<td>82.03</td>
<td>86.73</td>
</tr>
<tr>
<td>Overall</td>
<td>93.24</td>
<td>85.53</td>
<td>89.22</td>
</tr>
</tbody>
</table>
Experimental Results

• Relation Extraction for Multiple Arguments [Kim et al. 2008]
  – Data
    • Korean news documents about TV series
    • 930 documents which consist of 13,175 sentences
  – Seed
    • A tuple of the 4 arguments

<table>
<thead>
<tr>
<th>types of relations</th>
<th>with only binary relations</th>
<th>with all intermediates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># of tuples</td>
<td>precision</td>
</tr>
<tr>
<td>(PROGRAM,ACTOR,ROLE)</td>
<td>9</td>
<td>77.78</td>
</tr>
<tr>
<td>(CHANNEL,PROGRAM,ROLE)</td>
<td>11</td>
<td>81.82</td>
</tr>
<tr>
<td>(CHANNEL,PROGRAM,ACTOR)</td>
<td>12</td>
<td>58.33</td>
</tr>
<tr>
<td>(CHANNEL,PROGRAM,ACTOR,ROLE)</td>
<td>8</td>
<td>87.5</td>
</tr>
</tbody>
</table>
• Automatic Knowledge Acquisition and Inference for Dialog System Demo (POSTECH)
References

• M. Jeong, and G. G. Lee. 2006. Exploiting non-local features for spoken language understanding. COLING/ACL
References

References

• Minwoo Jeong, Gary Geunbae Lee. Tri-angular chain conditional random fields. To appear in IEEE Transactions on Audio, Speech and Language Processing (TASLP)
Thanks To

- Minwoo Jeong
- Cheongjae Lee
- Sangkeun Jung
- Kyungduk Kim
- Seokhwan Kim
- Hyungjong Noh
- Sungjin Lee
- And others…