A Multi-Stage Spoken Dialogue Question-Answering System

Afeka Center for Language Processing (ACLP)

Noam Lotner, Michal Gishri, Vered Aharonson, Ami Moyal
Speech Processing Conference, June 2012
Research Description

- Ongoing research, with the goal of creating speech based human-machine dialog in a specific domain
  - Speech input and speech output
  - Textual analysis of the domain using online resources
  - Funded by MAFAT
The Challenges

- Natural Spoken Dialogue
  - Lexicon size
  - Language models
  - Acoustical models
  - Speech recognition errors

- Text Analysis
  - Parsing, disambiguation and semantic analysis
  - Inferring meaning
  - Input contains speech recognition errors
Approach

- Question and answer dialog in a specific domain.
- Integration of text-based algorithms with algorithms used for speech analysis to achieve a common decision.
- Development of a distance measure that combines approaches from text analysis and speech recognition fields.
Text Processing Stage (offline)

- Questions in database divided into several sub-topics
  - "casualties," "causes," "leaders," etc.
- Each topic undergoes analysis
  - Keywords, bigrams, collocates
- A classifier is trained for each topic, based on these features
  - SVM with linear kernel function
Text Processing Stage (runtime)

- Input text passed through topic classifiers
- Questions from the n-best matching topics are then matched to the input text
- Answers for the best-matching question(s) are then returned by the system
Human-Machine Interaction

User Voice Input

Large Vocabulary Speech Recognition

Word1 Word2 ....

Text Index

Search & Retrieval Algorithm

System Word String

Text To Speech

System Voice Output

User Voice Input
System Schematic

Speech Recognition Engine
- Acoustic Model
- Language Model
- Dictionary

Recognition Result (Text)

Text Analyzer
- Topic classifier #n
- Topic classifier #2
- Topic classifier #1

Keyword Index
- Input Index
- Response Index

Text Corpus

Human Input (Speech)

System Response (Speech)

TTS Engine

Matched Response (Text)
Infrastructure

- Textual Corpora
  - “Gulf War” database for question and answers
  - Database for LM estimation

- Speech Corpora
  - Acoustical model development
  - Testing database – recording speakers in the specific domain

- Lexicon
  - Text lexicon, for word indexing
  - Speech lexicon, for word transcription
Infrastructure

- Text analysis tool
  - Count occurrences of text features
  - Produce input for training topic classifiers

- Classification framework
  - SVMLight toolkit
  - Current features:
    - Posterior probabilities of the various features given the topic
    - $P(\text{topic} \mid \text{keywords}), P(\text{topic} \mid \text{bigrams})$ etc.
    - Summed over each type of feature
Infrastructure

- Q&A Evaluation tool
- Interactive Q&A application
- Please welcome Michal for a demonstration
Evaluation

- Text evaluation tools – developed at ACLP
- Speech evaluation – using HTK tools
- Interactive Q&A application
  - Subjective evaluation
Experimental Setup

- Language: American English
- Domain: the 1990-1991 Persian Gulf War
- Textual Index
  - 2360 user queries (questions)
  - 584 system responses (answers)
  - Lexicon 25K words
  - 26 sub-topics
- Speech Inputs
  - 5 speakers
  - 483 user queries
  - Lexicon – 537 different words
  - LM - Bigram
Results

- Effect of classifiers on system performance (text input)

<table>
<thead>
<tr>
<th># topics</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71.33%</td>
</tr>
<tr>
<td>2</td>
<td>79.33%</td>
</tr>
<tr>
<td>3</td>
<td>80.67%</td>
</tr>
<tr>
<td>4</td>
<td>82.33%</td>
</tr>
<tr>
<td>5</td>
<td>83%</td>
</tr>
<tr>
<td>10</td>
<td>84%</td>
</tr>
<tr>
<td>15</td>
<td>84.33%</td>
</tr>
</tbody>
</table>

- Previous results show over 97% match

- Conclusions
  - Reduction in performance is inhibitive
  - Need to first improve classifier performance before considering effect of speech recognition errors
Next Steps

- Improve classifier performance using better text features
- Improve distance measure between questions using additional features (e.g. bigrams, collocates..)
- Port software to Efi (our NAO robot)
Next Steps

- Integrate textual entailment tools (collaboration with Prof. Ido Dagan)
- Continue increasing number of questions in database
- Retrieve data from online sources
The End