Natural Language Command Interpreter for Robot Control

Brandon Gottlob
The College of New Jersey
Department of Computer Science
Class of 2016

NSF REU Site Smart Spaces
Lehigh University

Advised by:
Dr. Mooi Chuah
Lehigh University
Department of Computer Science and Engineering
Problem Overview

● Limitations of LILI’s command control
  o Hard coded commands
  o Difficult to expand and maintain
  o Use of very limited vocabulary
  o Variety in sentence structure not tolerated

● Leads to less natural human-robot interaction (HRI)
Project Goals

● Generalized natural language (NL) command interpreter
  o Flexibility to add new commands
  o Utilize techniques from natural language processing (NLP) research
  o More natural HRI

● Provide video-based instructional coaching
Synonym Set Generation

- Compile set of words LILI can respond to
  - Create knowledge base
- Inputs:
  - Set of "known" words and their intended synset
  - Set of "unknown" words
- Output:
  - Final set of words grouping words with similar meanings together
Synonym Set Generation - WordNet

- Lexical database of English word senses
- Word sense - single meaning of a word
- Each sense grouped into “synset” with its synonyms
- Synsets connected based on similarities
  - Network structure
Synonym Set Generation - Implementation

\[
\text{Known Words} \ + \ \text{Unknown Words} \ = \ \text{Final Word Set}
\]

- \text{move, move.v.01}  \quad \text{twist}  \quad \text{move, go, teach}
- \text{turn, turn.v.01}   \quad \text{go}  \quad \text{turn, twist}
- \text{show, show.v.01}  \quad \text{move}  \quad \text{show}

- Use of WordNet’s path similarity calculation
  - Measure of how similar two synsets are in meaning

- For each unknown word, all its synsets are checked against each known synset
  - Known word with greatest path similarity is mapping of unknown word

- Words on same line considered synonyms
  - First word of line called \text{base word}
Synonym Set Generation - Limitations

- Assumes synset of unknown word is unknown
  - Counterproductive to determine a priori
  - Unintended synset may have max path similarity

- Path similarity depends on immutable topography of WordNet
  - Errors result from subtle distinctions
Action Identification

- Tokenize NL command - parse into list of words
- Search for each word in list of known actions
  - Start with first word in sentence
  - Stop when a known action is found
- Keep track of known action and its base action
## Action Identification

<table>
<thead>
<tr>
<th></th>
<th>Example 1</th>
<th>Example 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NL Command</strong></td>
<td>‘Please teach me how to clean my hands’</td>
<td>‘Can you move over to the left’</td>
</tr>
<tr>
<td><strong>1st Search</strong></td>
<td>‘please’</td>
<td>‘can’</td>
</tr>
<tr>
<td><strong>2nd Search</strong></td>
<td>‘teach’ -&gt; ‘show’</td>
<td>‘you’</td>
</tr>
<tr>
<td><strong>3rd Search</strong></td>
<td>-</td>
<td>‘move’ -&gt; ‘move’</td>
</tr>
</tbody>
</table>
Action Identification - Limitations

- Part-of-speech tagger not accurate for commands
  - Trained on declarative sentences
  - Thinks first verb of command is noun

<table>
<thead>
<tr>
<th>Characteristics of Declarative Sentences</th>
<th>Characteristics of Imperative Sentences (Commands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Begin with noun (subject)</td>
<td>Begin with verb phrase</td>
</tr>
<tr>
<td>Verb appears in conjugated form</td>
<td>Verb appears in base form; base form verbs can also be nouns</td>
</tr>
</tbody>
</table>
Object Extraction and Resolution

- Identify important words after action
  - Compile dictionary (key-value pairs) of these words
  - Map to semantic labels, resolve to base words

- Part-of-speech tagged command passed to object extractor function
  - One function per base action
  - Function implements specific grammar rules based on detected parts-of-speech
Object Resolution

● Has access to list of known objects
  o In same format as known actions
  o Object list can also be preprocessed

● Objects resolved to base objects
  o Root word of object searched for in known list

● Base action of command added to extracted object dictionary
## Object Extraction and Resolution

<table>
<thead>
<tr>
<th></th>
<th>Example 1</th>
<th>Example 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NL Command</strong></td>
<td>‘Please teach me how to clean my hands’</td>
<td>‘Can you move over to the left’</td>
</tr>
<tr>
<td><strong>Extractor result</strong></td>
<td>{‘person’: ’me’, ‘show_action’: ’clean’, ‘object’: ’hands’}</td>
<td>{‘direction’: ’left’}</td>
</tr>
<tr>
<td><strong>Resolution result</strong></td>
<td>{‘person’: ’me’, ‘show_action’: ’wash’, ‘object’: ’hand’, ‘action’: ’show’}</td>
<td>{‘direction’: ’left’, ‘action’: ’move’}</td>
</tr>
</tbody>
</table>
Object Extraction and Resolution - Limitations

● Rules for each action predetermined
  o Programmed into extractor module
  o Acceptable for small range of actions

● Relies on accurate part-of-speech tagging
  o Generally accurate for words after action

● Object resolution has same limitations as synonym set generation
Instructional Video Display

● Use of VLC player command line interface
  ○ Options for fine-tuning playback
    ▪ i.e. “play-and-exit”

● Videos titled ‘show_action-object’
  ○ Add new videos with name in this format

● Can show images if show_action not found
  ○ Will show image titled ‘object’
Project Outcomes

● Robust NL interpretation framework
  ○ Easy to extend
● Generation of synonym sets
  ○ Expanded vocabulary and awareness of semantic similarity
● Adaptive to various NL sentence structures
● Basis for instructional video feature

https://www.youtube.com/watch?v=KyvcuYmMMAQ&feature=youtu.be
Future Work

● Generalize interpreter using more advanced NLP concepts
  o Address discussed limitations at each step

● Improve synonym set generation
  o Utilize to create larger knowledge base
Contact

Email
gottlob1@tcnj.edu

Personal website
http://bgottlob.com

Project website
http://bgottlob.com/lili-interpreter