SPOKEN HUMAN ROBOT INTERACTION PROJECT

Artificial Intelligence (2A)

MARCH 5, 2019

AHMED EL SHEIKH - 1873337
Aly Amr - 1848399

Problem Definition

Human-robot interaction is the field of study dedicated to understanding, designing, and evaluating robotic systems for use by or with humans.

Spoken Language Understanding (SLU)

I. Interpreting commands

Basic Pipeline

1. Automatic Speech Recognition
   - Recognition and transcription of user utterances.
2. Morpho-Syntactic Analysis
   - Morphological information and syntactic structures.
3. Semantic Analysis
   - Extraction of meaning from sentences for grounding.

How to

1. Grammar Based
2. Data Driven

But, let us take about the differences

Grammar-based	Data-driven
Hand-crafted grammar development	Learning from data
Parsing	Statistical models rather than grammars
Limited coverage (structure+lexicon)	Generalization (structure+lexicon)
High performance on specific domains	Over/Under-fitting risk

Understanding tools

- Speaky 4 Robots (S4U)
- A tool for generating spoken command interpretation
- Grammar-based

- Language is domain-specific
- Interpretation is application-dependent
- Adaptive spoken Language Understanding 4 Robots (LU4R2)
- A Spoken Language understanding tool for robotic commands
- Data-driven
- Language is domain-driven
- Interpretation is context-sensitive

Automatic Speech Recognition [ASR]

It is simply the process of converting audio signals into text, also known as transcribing.

Burdens

There is never a free lunch, so what are the issues that we might encounter

1. Segmentation (Missing spaces).
2. Coarticulation (Merging sounds).
3. Homophones (different words that sound the same).
4. Not to mention that, input voice is highly noisy and that the natural language is naturally inherently ambiguous (highly dependent on the context).

Approaches

1. Classical Hidden Markov Models (HMM)
2. Deep Learning Connectionist Temporal Classification (CTC)
    - Evaluation Metrics [Word Error Rate (WER)]

II. Dialogic Interaction

- It improves by interacting with the user - by instructing the robot to fill in the gaps -
- Interactions require Spoken Dialog System (SDS)

Command interpretation	Dialogic interactions
Communication is an atomic processing of sentences	Communication is a sequence of turns (sentences)
Each sentence is independent	Each sentence depends (at least) on the previous one
Linguistic/Physical context irrelevant	The dialogue state influences the flow

The SLU process is enough to carry out the task	Implies: dialogue manager, SLU, dialogue state tracking, natural language generation

Spoken Dialog Systems (SDS)

- SDS falls under the umbrella of Dialogue management
- SDSs are intelligent agents able to help users to finish tasks more efficiently through spoken interactions, added to, they are being integrated into various devices.
- Based on Dialogic interactions
- From question-answering to dialog
- Communication takes place in a sequence of turns
- Each sentence depends (at least) on the previous one, in general the dialogue state influences the flow
- Language in dialog is more “implicit” and unstructured
- Disfluencies can be relevant (including other sound features)
- Text-To-Speech is required

Types of SDS

1. Conversational Agents
    - Personal Assistants
    - Help user to achieve manifold tasks
2. (Chit-chat agents) chatbots
    - User entertainers
    - No specific goal
    - Focus is on producing coherent responses
3. Task oriented SDSs
    - Personal Assistants
    - Help user to achieve a specific tasks

Task Oriented SDSs

Classified based on:

- Domain ontology: a characterization of the user intentions that can be dealt with
- Frames: structured representations of the semantics associated to each intension
- Slots (frame elements): containers for the values that need to be filled in to achieve the intention

Dependencies and Tools

Dialogflow

Dialogflow is a conservation agent development platform grounded on natural language conversations. It lets app developers create task-oriented dialog systems that can either be rule-based, machine learning based or a hybrid approach combining both.

With Dialogflow, developers have access to inbuilt machine learning and natural language processing technologies. This lets them focus on other integral parts of app creation rather than on delineating in-depth grammar rules.

How it works

1. Agent
Agents are best described as NLU (Natural Language Understanding) modules. These can be included in your app, product, or service and transform natural user requests into actionable data. Agent is the name of your app you are creating.

2. Intent
Whenever the user ask a question, it will try to match in corresponding Intent. Intent plays vital role in the assistant app. In Dialogflow, an intent houses elements and logic to parse information from the user and answer their requests.

To understand the question better by intent we need to feed as much as data we can. The more variations added to the intent, the better the agent will comprehend the user. Developer need to think of different variations of same question.

3. Entity
The Dialogflow agent needs to know what information is useful for answering the user's request. These pieces of data are called entities. Assumed to be dynamic variables.

4. Context
Context plays vital role in the success of assistant. Context helps the assistant to talk more like human by maintaining the context and replying in the context to end users.

Google APIs

Speech-to-text

Google Cloud Speech-to-Text enables developers to convert audio to text by applying powerful neural network models in an easy-to-use API. Cloud Speech-to-Text is tailored to work well with real-life speech and can accurately transcribe proper nouns (such as, names) and appropriately format language (such as, dates, phones numbers).

Text-to-speech

Google Cloud Text-to-Speech converts text into human-like speech by applying groundbreaking research in speech synthesis (WaveNet).

Spacy & NLTK

Spacy was used a dependency parser and NLTK was used to build the dependency tree.

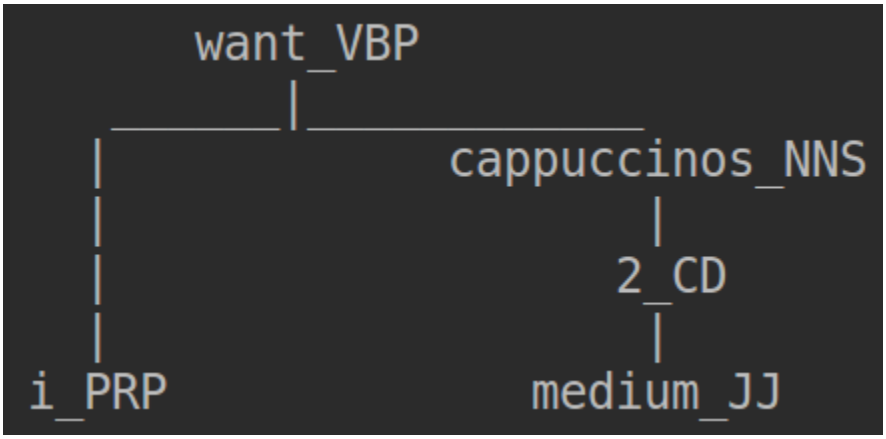

Figure: Dependency Tree for “I want 2 medium cappuccinos”

Scenario

You are at a coffee shop where you can order drinks or snacks. Your last order is also saved for future requests. You can go pick-up your order or have it delivered, while to choosing to pay with a credit card or in cash.

Dialog Sample

User: Good Morning
Bot: Good day! I have a lot coffee and snacks. What can I get you to drink?
User: I want 2 medium cappuccinos
Bot: Would you like a delivery or pick-up?

User: Delivery
Bot: You want 2 medium Cappuccino and you want a delivery. Is that right?
User: Yes
Bot: Will you use the same card?
User: No
Bot: What is the new card number?
User: 12345
Bot: All set, enjoy!