

Error Detection: Know What you Don't Know

PROJECT PITCH CS294S/W FALL 2020

Semantic Parsing

- Is the task of converting what the user says to executable code.



What is a **Chinese** restaurant in **Palo Alto**?

Restaurant, servesCuisine =~ "**Chinese**"
&& geo =~ "**Palo Alto**"

- Depending on the test questions, commercial VAs are ~70-85% accurate.
 - (And we see lower numbers in research papers)

Semantic Parsing

- Virtual assistants are far from perfect.
- The result is user frustration
 - Users have to repeat their command several times
 - Sometimes the wrong command is executed
- But the conversation does not have to end with a mistake
- **Very Big Question:** How can we build parsers that seek user's feedback and fix their own mistakes?
- **Project-size Question:** How can we build parsers that know they made a mistake?

High-Level Project Plan

- Step 1: Choose a semantic parsing dataset (Schema2QA, MultiWOZ, etc.)
- Step 2: Ideate (we have some ideas!)
- Step 3: Implement your ideas, train models
- Step 4: Iterate
- Step 5: (Bonus) Integrate your model into Almond
- Step 6: Profit!
 - i.e. go down as one of the people who helped disrupt the emerging virtual assistant oligopoly and lower the power of a few companies over consumers!

Natural Response Generation for Virtual Assistants

PROJECT PITCH CS294S/W FALL 2020

Almond The Virtual Assistant

You can try Almond version 1.99 at almond-dev.stanford.edu

For now, you can ask about the weather or restaurants or connect it to your spotify account.

The following is a conversation I had with it, without any edits.



Hi, what can I do for you?

Can you tell me about restaurants in illinois?



Sorry, there are no restaurant matching your request.

How about Stanford?



I have Evvia Estiatorio, Ramen Nagi and Zareen's. All of them are restaurants rated 4.5 star.

Can you tell me more about Evvia Estiatorio?



I have one, it is Evvia Estiatorio. It is restaurant rated 4.5 star near 420 Emerson St, Palo Alto, CA 94301.

What is the price range?



Evvia Estiatorio is a expensive restaurants.



Hi, what can I do for you?

restaurants

Can you tell me about restaurants in illinois?



Sorry, there are no restaurant matching your request.

How about Stanford?



I have Evvia Estiatorio, Ramen Nagi and Zareen's. All of them are restaurants rated 4.5 star.

stars

Can you tell me more about Evvia Estiatorio?



I have one, it is Evvia Estiatorio. It is restaurant rated 4.5 star near 420 Emerson St, Palo Alto, CA 94301.

a

an

What is the price range?



Evvia Estiatorio is a expensive restaurants.

restaurant



Hi, what can I do for you?

Can you tell me about restaurants in illinois?



Sorry, there are no restaurant matching your request.

How about Stanford?



I have Evvia Estiatorio, Ramen Nagi and Zareen's. All of them are restaurants rated 4.5 star.

?

Can you tell me more about Evvia Estiatorio?



I have one, it is Evvia Estiatorio. It is restaurant rated 4.5 star near 420 Emerson St, Palo Alto, CA 94301.

What is the price range?



Evvia Estiatorio is a expensive restaurants.

Natural Response Generation for VAs

- We have:
 - A large set of synthetic multi-turn dialogues for several domain
 - In each turn, what VA needs to say back to the user in ThingTalk code
 - A baseline model that converts ThingTalk code to natural language
 - A baseline neural network that tries to “fix” the response
- **Question:** How do we make responses more natural?



Hi, what can I do for you?

Can you tell me about restaurants in illinois?



I'm sorry, but I don't have a restaurant that matches your request.

How about Stanford?



I found Evita Estiatorio, Ramen Nagi and Zareen's, all of which have a rating of 4.5 stars .

Can you tell me more about Evvia Estiatorio?



It's a restaurant with a 4.5-star rating, located at 420 Emerson Street, Palo Alto, CA 94301 .

What is the price range?



Evita Estiatorio is an expensive restaurant.

The Problem

- The “fixes” are not always correct.
 - Pieces of information might get dropped
 - Additional information might be hallucinated by the neural network
- There seems to be a trade-off between naturalness and correctness in the current system.
- Correctness is important for VAs, especially in sensitive domains like banking

High-Level Project Plan

- Step 1: Define/find a suitable evaluation metric for correctness
- Step 2: Ideate (we have some ideas!)
- Step 3: Implement your ideas, train models
- Step 4: Iterate
- Step 5: Conduct human evaluation
- Step 6: (Bonus) Integrate your changes with Almond
- Step 7: Profit!
 - i.e. go down as one of the people who helped disrupt the emerging virtual assistant oligopoly and lower the power of a few companies over consumers!

Tools to Find a Solution

- Natural Language Processing
- Heavy use of pretrained language models like BERT, BART and GPT-2
- Human evaluation on Amazon Mechanical Turk

