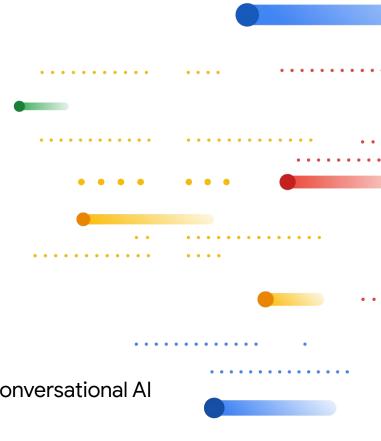# Towards scalable multi-domain conversational agents

Xiaoxue Zang, Abhinav Rastogi, Srinivas Sunkara, Raghav Gupta, Pranav Khaitan

Google Research
The 4th International Workshop on Search-Oriented Conversational AI

Overview

- Introduction
- The Schema-guided approach
- The Schema-guided dialogue dataset
- Data collection methodology
- Dialogue state tracking challenge (DSTC 8)
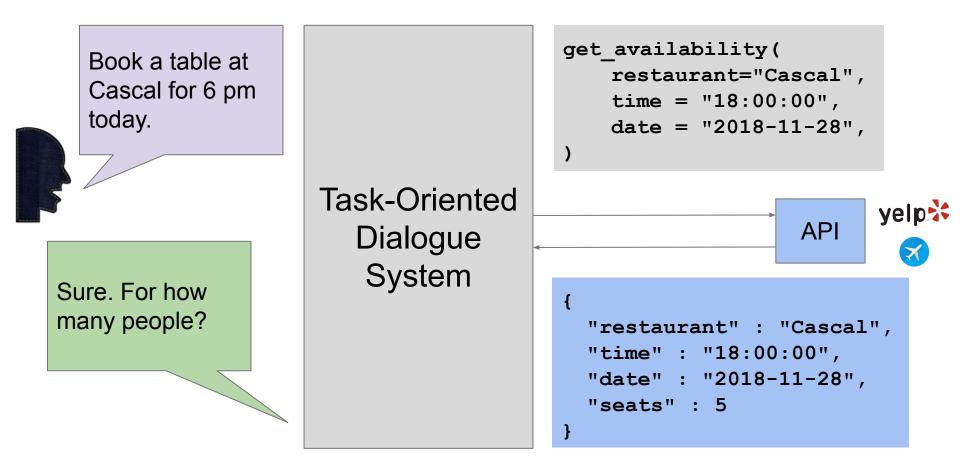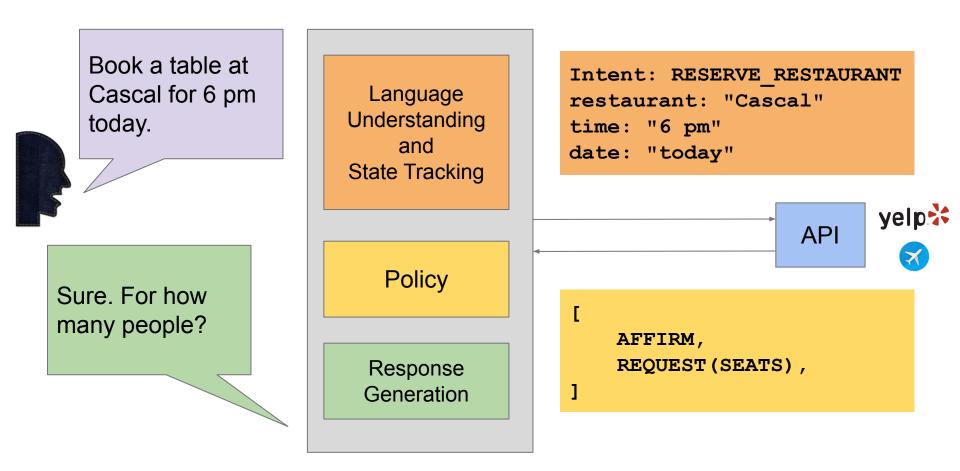
Overview

- Introduction
- The Schema-guided approach
- The Schema-guided dialogue dataset
- Data collection methodology
- Dialogue state tracking challenge (DSTC 8)

Challenges from Real World APIs

- The number of APIs to support is large

- No uniform schema or entity names
  - Slot names
    - "origin" = "from", "departing from" = "where from", ...
  - Slot values
    - "San Francisco" = "SFO" = "SF", ...

- Don't allow arbitrary calls
  - find_restaurant(*city="San Francisco"*, date="today")   ✔️
  - find_restaurant(city, *date*="today")   ❌

- The set of supported entities may be large and dynamic
  - May not expose the list of all available entities
  - Robustness to OOV problem

Requirements of Virtual Assistants

- Facilitate dialogues across multiple Services/APIs
    - Handle large universe of services spanning multiple domains
    - Data efficient: Zero-shot or few-shot learning for tail services

Domain	Hotel	Movie	Flight
Services	Hotels.com Booking.com	FANDANGO IMDb	Expedia skyscanner Google . . . . . .

Requirements of Virtual Assistants

- Can handle multi-domain conversations
  - Seamlessly switching domains
  - Carrying over relevant slots among APIs

Book a table at Cascal at 6 pm today and find a taxi to go there.

```
get_availability(
    restaurant="Cascal",
    time = "18:00:00",
    date = "2018-11-28",
)

get_cab(
    destination="Cascal",
)
```

Requirements of Virtual Assistants

- Robust to changes in schema
  - No retraining if the interface of an API changes
  - Robust to new/unseen slot values

Round trip ▾   1 passenger ▾   Economy ▾

◎ San Francisco   ⇄   📍 Try "JFK, LGA, EWR"

← Free-form input

Going to
📍 Destination, hotel name, airport, train station, landmark, or address   ✕

Existing Dialogue Datasets

Metric	DSTC2	WOZ2.0	FRAMES	M2M	MultiWOZ
#Domains	1	1	3	2	**7**
#Dialogues	1,612	600	1,369	1,500	**8,438**
#Turns	23,354	4,472	19,986	14,796	**113,556**
#Slots	8	4	61	13	**24**
#Values	212	99	3,871	138	**4,510**

● **Datasets not large enough for training generic virtual assistants**

Existing Dialogue Datasets

Metric	DSTC2	WOZ2.0	FRAMES	M2M	MultiWOZ
#Domains	1	1	3	2	**7**
#Dialogues	1,612	600	1,369	1,500	**8,438**
#Turns	23,354	4,472	19,986	14,796	**113,556**
#Slots	8	4	61	13	**24**
#Values	212	99	3,871	138	**4,510**

- Datasets not large enough for training generic virtual assistants
- **Very few slots and possible values**

Existing Dialogue Datasets

Metric	DSTC2	WOZ2.0	FRAMES	M2M	MultiWOZ
#Domains	1	1	3	2	**7**
#Dialogues	1,612	600	1,369	1,500	**8,438**
#Turns	23,354	4,472	19,986	14,796	**113,556**
#Slots	8	4	61	13	**24**
#Values	212	99	3,871	138	**4,510**

- Datasets not large enough for training generic virtual assistants
- Very few domains, slots and possible values
- **Make simplified assumptions on underlying APIs/Services**

Existing dialogue datasets do not expose all the challenges

Overview

- Introduction
- The Schema-guided approach
- The Schema-guided dialogue dataset
- Data collection methodology
- Dialogue state tracking challenge (DSTC 8)

The Principal Idea

Problems
Assuming the set of slots and intents to be handled by the dialogue system are fixed

- Slots and intents are treated as merely labels

- Cannot handle unseen/new intents or slots

Solution
Associate a semantic representation with each label using their natural language description

- Predictions are made over dynamic sets of labels using embedded representations

- Unseen/new slots and intents can be handled

The Schema-guided Approach

Service

service_name: "RideSharing"
description: "On-demand taxi calling service"

Slots

name: "destination"
description: "Destination for taxi ride"

name: "number_of_riders"
description: "Number of riders to call taxi for"

name: "shared_ride"
description: "Whether ride is shared or not"

name: "ride_fare"
description: "Total cost of the ride"

Intents

name: "GetRide"
description: "Call a taxi to head to a given destination"
required_slots: ["destination", "number_of_riders"]
optional_slots: ["shared_ride" = "False"]

Google AI

The Schema-guided Approach

Descriptions can be used to semantically identify related services and slots

Service

service_name: "RideSharing_1"
description: "On-demand taxi calling service"

service_name: "RideSharing_2"
description: "App to book a cab to any destination"

Slots

name: "number_of_riders"
description: "Number of riders to call taxi for"

name: "number_of_seats"
description: "Number of seats to reserve in the cab"

Intents

name: "GetRide"
description: "Call a taxi to head to a given destination"

name: "GetRide"
description: "Book a cab for any destination, number of seats and ride type"

The Schema-guided Approach

Annotations dependent on **the schema of API** under consideration

Flight Booking Service A

Intents	SearchFlight, ReserveFlight
Slots	origin, destination, depart, return, number_stops, trip_type, ...

Flight Booking Service B

Intents	FindFlight, ReserveFlight
Slots	depart, arrive, depart_date, return_date, direct_only, trip_type, ...

SearchFlight:
origin = *Baltimore*
destination = *Seattle*
number_stops = *0*

User:	Find direct round trip flights from Baltimore to Seattle.
System:	What dates are you looking for?

FindFlight:
depart = *Baltimore*
arrive = *Seattle*
direct_only = *True*

SearchFlight:
origin = *Baltimore*
destination = *Seattle*
number_stops = *0*
depart = *May 16*
return = *May 20*

User:	Flying out May 16 and returning May 20.
System:	I found a Delta itinerary for 302 dollars.

FindFlight:
depart = *Baltimore*
arrive = *Seattle*
direct_only = *True*
depart_date = *May 16*
return_date = *May 20*

Overview

- Introduction
- The Schema-guided approach
- The Schema-guided dialogue dataset
- Data Collection Methodology
- Dialogue State Tracking Challenge (DSTC 8)

The Schema-Guided Dialogue Dataset (SGD)

Metric	DSTC2	WOZ2.0	FRAMES	M2M	MultiWOZ	SGD
#Domains	1	1	3	2	7	**16**
#Dialogues	1,612	600	1,369	1,500	8,438	**16,142**
#Turns	23,354	4,472	19,986	14,796	113,556	**329,964**
#Slots	8	4	61	13	24	**214**
#Values	212	99	3,871	138	4,510	**14,139**

- **Largest** publicly available dataset for task-oriented dialogues
- Contains multi-domain conversations involving **34 services** over **16 domains**
- Test set contains few unseen services to test zero shot generalization

Covered Domains (Training set)

Domain	Weather	Banks	Buses	Calendar	Events	Flights	Homes	Hotels
#Dialogues	951	727	2280	1602	3509	2747	847	3353

Domain	Media	Movies	Music	Rental car	Restaurant	Ride sharing	Services	Travel
#Dialogues	1113	1617	1290	1585	2419	1703	1889	1871

For more details, please go to dataset link:
https://github.com/google-research-datasets/dstc8-schema-guided-dialogue

Example - single domain

Google AI

Okay, it's family movie night. We want to see a good movie that's family oriented.

Do you know the title of the movie you want to see?

Well, my first choice is Captain Marvel.

Okay, Captain Marvel is available. There are no subtitles with this movie. Please confirm if this is the movie you want to see.

Oh. No, I want us to see a movie with subtitles tonight. I'm thinking The Curse of La Llorona movie might be a good choice if it has subtitles.

Yes, The Curse of La Llorona film is available with subtitles. If you want to see that movie, please confirm.

Movie Domain

Example - multi domain

Please tell me the weather forecast for March 8th.

Any particular city's weather forecast?

Please check the weather in **Menlo Park**.

The expected average temperature for the day is 83 degrees Fahrenheit, with a 9 percent chance of rain.

Sounds good! Please find me a restaurant offering Latin American cuisine there.

Cafe Del Sol Restaurant, a nice restaurant **Menlo Park** is among the 3 I found.

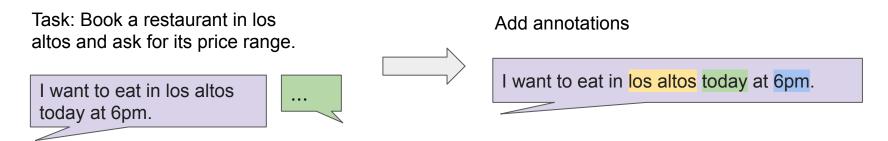**Weather Domain → Restaurant Domain**

Overview

- Introduction
- The Schema-guided approach
- The Schema-guided dialogue dataset
- Data collection methodology
- Dialogue state tracking challenge (DSTC 8)

Data Collection Methodology

- Two commonly used approaches for data collection
    - Chat based
    - Simulation based

Chat-based Data Collection

1. Crowdworkers chat with each other given the high-level intent.
2. Annotate the dialogues.

Task: Book a restaurant in los altos and ask for its price range.

I want to eat in los altos today at 6pm.

...

Add annotations

I want to eat in los altos today at 6pm.

☑ May generate more realistic conversation.

✗ Coarse or no control over dialogue flow.
→ Hard to control the dialogue variations.

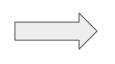✗ Annotation can be noisy and expensive [MultiWOZ 2.1 - Eric et. al.]

Simulation-based Data Collection

Approach 2:

1. Simulate dialogue outlines.

2. Paraphrase a simulated conversation

INFORM(location="los altos")
INFORM(date="today")
INFORM(time="6pm")

⟹ I want to eat in los altos today at 6pm.

☑ Annotations are cleaner as they are automatically generated

☑ Cheaper and faster to collect

☑ Easy to ensure coverage of many dialogue flows
→ Control the dialogue variation

✗ Requires a powerful simulator that simulates the dialogue outlines

Data Collection Approach

1

Generate dialogue skeletons using a multi-domain simulator

2

Use utterance templates to generate synthetic conversations

3

Get paraphrased dialogues from crowdworkers

REQUEST(location)
REQUEST(cuisine)

INFORM(location="los altos")
INFORM(date="tomorrow")
INFORM(cuisine="greek")

OFFER(restaurant="Opa!")

Where do you want to eat?
What kind of food?

I want to eat in los altos.
I want to go there tomorrow.
I want to eat greek food.

Opa! is a wonderful restaurant.

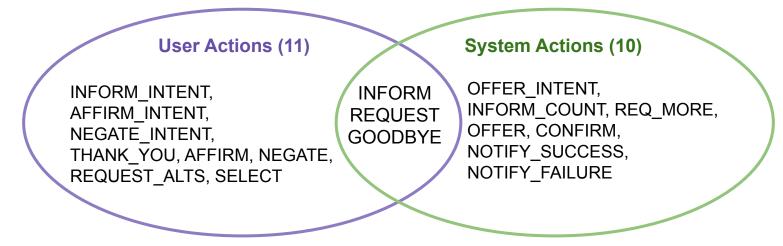Where are you located and what is your preferred food?

I am looking for a greek restaurant in los altos tomorrow.

How does Opa! sound?

The Dialogue Simulator

- Multi-domain simulator generates dialogue outlines containing **user and system actions** and **state annotations**

User Actions (11)

System Actions (10)

INFORM_INTENT,
AFFIRM_INTENT,
NEGATE_INTENT,
THANK_YOU, AFFIRM, NEGATE,
REQUEST_ALTS, SELECT

INFORM
REQUEST
GOODBYE

OFFER_INTENT,
INFORM_COUNT, REQ_MORE,
OFFER, CONFIRM,
NOTIFY_SUCCESS,
NOTIFY_FAILURE

- Implementation details to be released after conclusion of DSTC8

Utterance Templates

- User and system actions are converted into robot utterances using templates

User Actions

INFORM(where_to=**A**) INFORM(where_to=dontcare)	I want to travel to **A**. Any place works for me, I just want to travel.

System Actions

OFFER(title=**A**, **B**) INFORM(number_stops=**A**)	What about **A** or **B**? The flight has **A** stops.

- Templates make it easier to convey the meaning of actions to workers

Dialogue Paraphrasing

- Generate natural utterances from robot utterances
- Very simple, annotation preserving crowd task

SYSTEM	Please confirm the following details: The name of the hair stylist is Great Clips, the appointment is at 2 pm and the appointment is on March 2nd.	Enter the utterance text here
USER	Yes that sounds good. What is its contact number? What is its average rating?	Enter the utterance text here
SYSTEM	Your appointment is confirmed. Its average rating is 3.8. Its phone number is 408-532-1849.	Enter the utterance text here
USER	Thank you very much. I want to watch a movie around there on that day. I like Comedy movies.	Enter the utterance text here

Data Collection Challenge

- Some crowd workers try to get away with minimal work.

Short unnatural utterances	No paraphrasing - copy from the given outlines
U: Reserve hotel S: name U: one hotel andra 4th of march nine S: Ok U: Ok S: Ok	U: I am looking for something to eat. I want it in Fremont. I am looking for some intermediate priced food. S: What kind of food are you looking for? U: Some Asian Fusion kind of food will be perfect.

- Solution
  - Create a whitelist of skilled crowdworkers.
  - Automated checks to ensure alignment with annotations and variability in utterances.

Whitelist Creation

Initial paraphrasing tasks open to all workers, followed by a verification task for obtaining scores for each paraphrased dialogue.

Speaker	Outline	Dialogue	Is dialogue valid and natural?
USER	I am looking for a place to eat. I want it in Fremont. I am looking for something intermediate priced priced.	I'm looking for somewhere I can grab some food in Fremont. Something intermediate priced would be perfect.	○ True ○ False
SYSTEM	What kind of food are you looking for?	Is there any kind of food you want to eat specifically today?	○ True ○ False

Do you think the overall dialogue sounds like a natural conversation?

○ **True**

○ **False**

How to ensure data quality?

Use verification task and scripts to find skilled workers

- Give paraphrasing work only to the qualified crowd workers.
- Update the whitelist consistently
  - Run scripts to compute the similarities between outlines and paraphrased utterances.
  - Remove workers whose dialogues are of high similarity score and their dialogues.

Overview

- Introduction
- The Schema-guided approach
- The Schema-guided dialogue dataset
- Data Collection Methodology
- Dialogue State Tracking Challenge (DSTC 8)

DSTC 8 Task

- Aim
  - Dialogue state tracking in a virtual assistant

- New Challenges
  - Zero-shot learning of unseen services
  - Handling OOV slot values
  - Transfer of slot values in multi-domain dialogues

- Submission deadline: Oct 13, 2019

Thank You!

Questions?

xiaoxuez@google.com
abhirast@google.com