Dialog for Language to Code

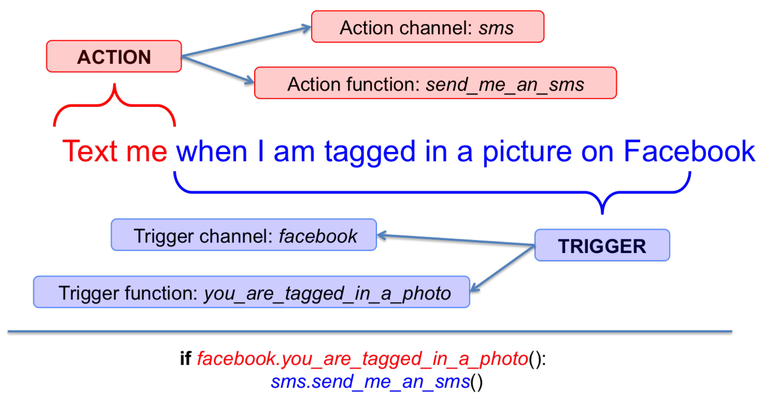
Shobhit Chaurasia and Raymond J. Mooney
University of Texas at Austin

Interactive Code Generation

We propose a system that can engage users in a dialog to clarify their intent until it has all the information to produce correct code.

IFTTT domain

IFTTT (if-this-then-that) allows users to automate simple tasks by creating short scripts, called recipes, through a GUI that enables them to connect web-services and smart devices.

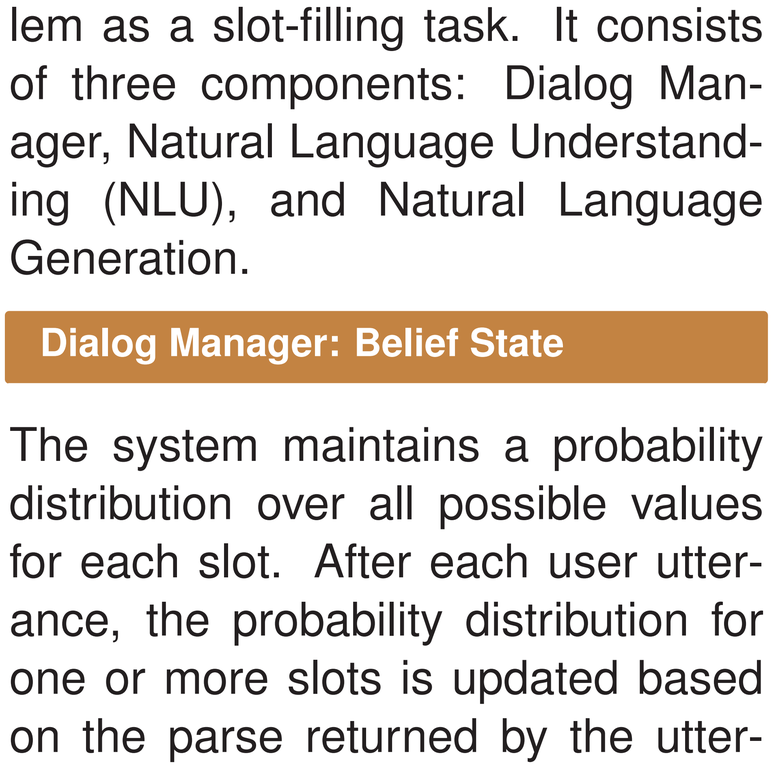


Dialog System

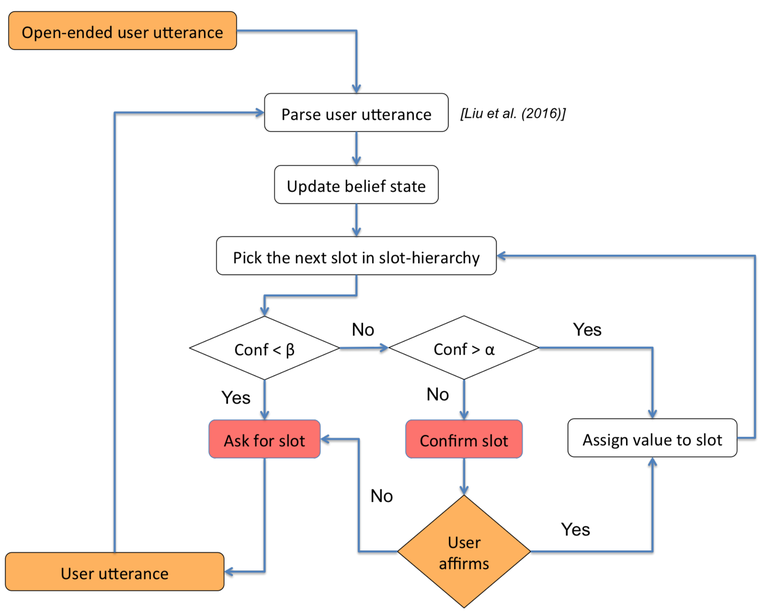
The aim of the dialog system is to determine values of channels and functions for the recipe that the user wants to create. We cast this problem as a slot-filling task. It consists of three components: Dialog Manager, Natural Language Understanding (NLU), and Natural Language Generation.

Dialog Manager: Belief State

The system maintains a probability distribution over all possible values for each slot. After each user utterance, the probability distribution for one or more slots is updated based on the parse returned by the utterance parser. The system follows a hand-coded policy over the discrete state-space obtained from the belief state by assigning the values with highest probability to each slot.



Dialog Manager: Static Dialog Policy



Natural Language Generation

The dialog system uses templates and IFTTT API documentation to translate its belief state into a comprehensible utterance. For example, the confirmation request for the `blink_lights` action function of the `hue` action channel is: 'Do you want to *briefly turn your hue lights off then back on* every time the applet is triggered?'

Retraining NLU using Dialog

A dialog approach to recipe synthesis unlocks the possibility of continuous parser improvement through conversations. Opening user utterances and user utterances for each slot after a system-initiative in successful dialogs were paired with inferred slot values to retrain the models.

Baselines

- *Liu et al. (2016)* LSTM with attention provided only with initial recipe descriptions (single-shot).
- *Concat* baseline, which uses the same model, but is provided with all the user utterances from the conversation concatenated.

Experiments

- *Constrained User-Initiative*: Users used the original recipe descriptions as their first utterance when asked to describe the recipe.
- *Free User-Initiative*: Users drove the entire conversation themselves, including the user-initiative.

Mechanical Turk Interface

CONVERSATION:

Note that the names in "Task Description" are only representative. **The bot will not directly understand words like** *"any_event_starts"*, or *"send_a_notification"*. You should describe these to the bot in your own language with proper English.

ROBOT:	Hi! Please describe the applet you want to create for automating the task you have on your mind.
YOU:	I want notification on my Google Glass about taking and sharing pictures for events
ROBOT:	Do you want an event on the Google Calendar service to trigger the applet?
YOU:	yes
ROBOT:	Which event on the Google Calendar service should cause the applet to run?
YOU:	When some event has started
ROBOT:	Do you want to trigger the applet within 15 minutes of the starting time of any event on your google calendar?
YOU:	yes
ROBOT:	The applet will trigger within 15 minutes of the starting time of any event on your google calendar. It will use the Google Calendar service to look for this event. The action taken will be to send a notification to your google glass timeline. This action will be performed using the Google Glass service. Is this what you wanted? (yes/no)
YOU:	yes
ROBOT:	Ok, bye!

Continue

TASK DESCRIPTION:

"Notify Glass to take and share pictures for Events."

Which event should cause the applet to run/trigger?
any_event_starts

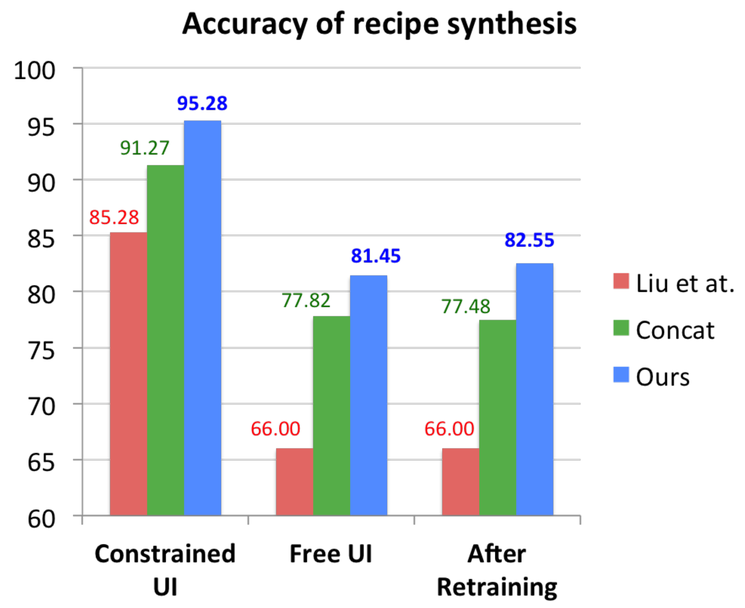
On which service should the above event happen?
google_calendar

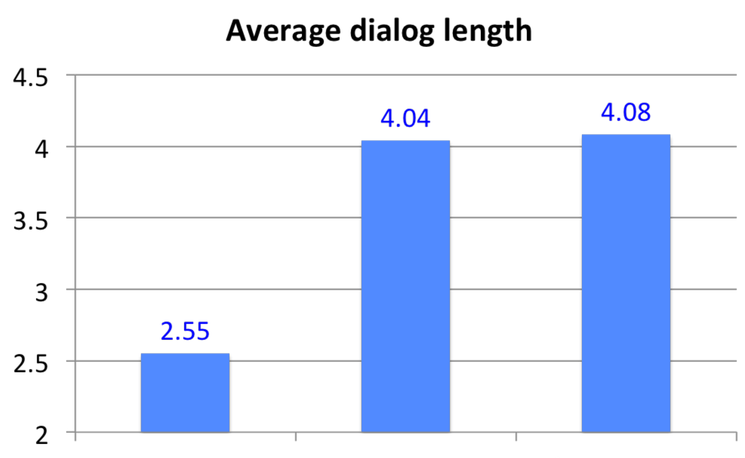
What action should be taken when the applet runs/triggers?
send_a_notification

Which service should be used to perform the action?
google_glass

Results

We trained our parsers on the IFTTT training set. Evaluation was done on the 'gold' subset of the IFTTT test set, which consists of 550 recipes on which at least three humans presented with the recipe descriptions agreed with the true labels.





Conclusion

The proposed system engaged the user in a dialog, asking questions until it was confident it its inference, thereby increasing the accuracy over the state-of-the-art models that are restricted to synthesizing recipes in one shot by 10 – 15 points.

Code available at https://github.com/shobhit6993/natural-language-to-code

mooney@cs.utexas.edu