Order Buddy

Aanchal Anil Samdariya: samdariya.a@husky.neu.edu
Apoorva Kasoju: kasoju.a@husky.neu.edu
Karan Gulur Muralidhar: gulurmuralidhar.k@husky.neu.edu
Krishna Sai Rohith Kanuparti: kanuparti.k@husky.neu.edu

September 30, 2019

I. Summary

Traditional way of placing an order involves repetitive manual labour, be it giving/taking orders face to face at an eatery or using mobile/web ordering.Leveraging increase in popularity of hands free assistants such as Google Home and Alexa, we enable conversational agent to work as a Speech-to-text Order Assistant.

The user will be given the ability to navigate through all phases of order management(i.e., placing an order, modifying the existing order and cancel an item/order), in effect, have a conversation and be able to place an order hands free.

The following could be an example of the kind of conversation possible with Order Buddy:

OB: Hey there! What do you want to order?
Cust: I want a Mocha Latte.
OB: Got it, one Mocha Latte, Do you want to customize this item?
Cust: Can you add extra sugar please
OB: Got it, one Mocha Latte with extra sugar. Do you want to order anything else?
Cust: No.
OB: Order Confirmed, have a good day!

A subset of the menu from Starbucks will be used for proof of concept and later aim to expand to any menu.

II. Proposed Plan of Research

Google Cloud Speech-to-Text[1] cloud service enables to convert prerecorded or real-time streaming audio to text by applying powerful neural network models in an easy-to-use API. The API recognizes 120 languages and variants to support global user base. This service will be incorporated in order to convert orders placed by user in voice to text.

With text in hand, our next objectives include

- **Item Extraction**: Each restaurant comes with their own naming standards for an item. The ability to identify an item from the user text and match with an existing menu item is not a straightforward task. Current NLP services like **Google Cloud Natural Language**[2] have state of the art solutions to extract people, location and organization names but to recognize domain-specific entities like item names, a custom entity extraction model may be needed.

- **Intent Identification**[3]: A user can have one of the following intents while placing an order:

  - **Order Intent**:The users wishes to place an order
  - **Customize Intent**:The user wishes to customize add-ons for previously placed order
  - **Modify Intent**:The user wishes to modify previously placed order
  - **Cancel Intent**:The user wishes to cancel a previously placed order
  - **Exit Intent**:The user wishes to complete order/end conversation
  - **Conversation Intent**: The user requests information/expects answer from a casual communication(outside
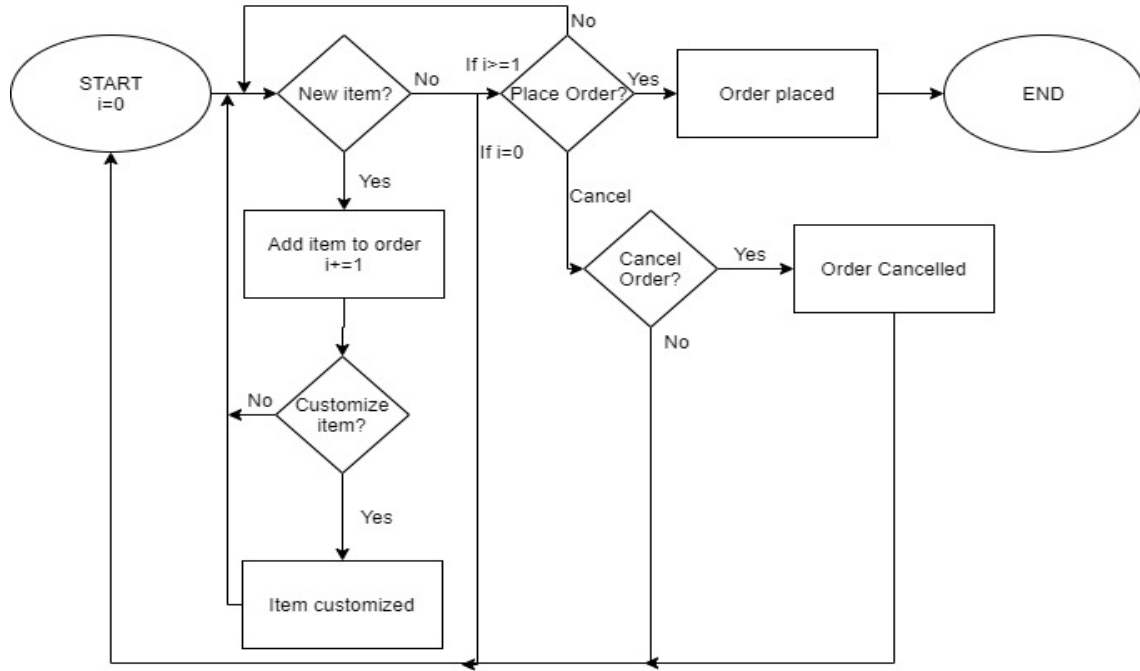
Figure 1: *Flow Chart representing state transitions for various Intents*

the order scope).

To achieve a conversation flow, like a chatbot, we would like to implement a state transition in the back end to navigate between states and respond to each utterance of the user.

Figure 1 explains the flow of conversation Order buddy uses to process to place an order. At the start, number of items *(i)* as part of the order is *0*. The customer is asked if a new item is to be added to the cart along with customisation. If he/she does not wish to add additional items, an order confirmation is requested, after which the order is placed. The customer is also given the flexibility to cancel the order or part of it before confirming the order.

We also plan to explore Google's DialogFlow service[4] to build voice powered conversational interface in the later part of the project.

To evaluate[5] the system, we plan to create a list of examples which will be labelled manually to the intent they refer to. Each example would be common phrases used by people to express their intent. This dataset will be used to calculate F1-score for each intent.

III.   Preliminary results

Using the Google Cloud Speech to Text API, *Table 1* displays the text transcription along with the confidence level of example dialogue provided as an input to the API. As seen in the results, the Order and Customize intents have a higher confidence level compared to other intents.

References

[1] Google Cloud Speech to Text: *https://cloud.google.com/speech-to-text/*.

[2] Google Cloud Natural Language: *https://cloud.google.com/natural-language/*.

[3] Homa B. Hashemi, and Amir Asiaee, Reiner Kraft
*Query Intent Detection using Convolutional Neural Networks*Intelligent Systems Program,University of Pittsburg and Yahoo Inc.,Sunnyvale,CA

Intent	Input	Transcription	Confidence Score
Order	can I have one tall white chocolate mocha	"can I have one tall white chocolate mocha"	0.9806791
Customize	can you please add whipped cream to my coffee	"can you please add whipped cream to my coffee"	0.9849021
Modify	I want a cafe latte instead	"I want a cafe latte instead"	0.935874
Cancel	I don't want the coffee	"I don't want the coffee"	0.83714724
Exit	I'm all set	"I'm all set"	0.971368
Conversation	I'm waiting for the fall to taste seasonal flavors	"I'm waiting for the fall today seasonal flavors"	0.95654744

Table 1: *Examples of transcription using Google's Speech-to-Text API*

[4] Google Cloud DialogFlow: *https://cloud.google.com/dialogflow/docs/reference/rest/v2-overview*.

[5] Brenes, David Gayo-Avello, Daniel and Pérez-González, Kilian *Survey and evaluation of query intent detection methods* Proceedings of Workshop on Web Search Click Data, WSCD'09. 10.1145/1507509.1507510