Project Proposal: Speech to Text Order Assist

Rutu Nanavati, Viral Patel and Rishabh Agrawal

1 Summary

In recent years, we have witnessed a revolution in the ability of computers to understand and to convert natural speech, especially with the application of deep neural networks (e.g., Google Home Mini, Alexa etc). In particular, automated speech systems are still struggling to recognize simple words and commands. They don't engage in a conversation flow and force the user to adjust to the system instead of the system adjusting to the user.

The key focus of this project is to speed up the process flow of ordering at restaurants. This is especially useful in a drive-thru and take-out service to get through the queue quickly. Currently, a lot of resources and man-power are wasted by a person taking manual orders. Eventually, this could replace the person taking orders and save the money and time for both user and business end.

We would be using the google dataset Taskmaster-1. This dataset consists of 13,215 task-based dialogues in English. We will be using the conversation from only 2 of the 6 domains (ordering pizza and ordering coffee drinks) which are relevant to Quick Service Restaurants.

The Quick Service Restaurant (QSR) Assist aims to accomplish the above. It would start out by hearing the order, transcribe it to text, map it to the items in the menu and finally create an invoice. From the converted text we would be able to classify the intent of that part of the speech such as adding, deleting, modifying and canceling an order. The extracted information is then used for keyword extraction and invoice generation.

2 Proposed plan of research

The Quick Service Restaurant (QSR) Assist will have the following key functionalities:

- Speech-to-Text : Listen for user text and transcribe for speech to text.

  Input speech: "Can I please get a Veggie Pizza and a coke bottle."

  Transcribed to: "Can I please get a Veggie Pizza and a coke bottle."

  This part of the QSR Assist will be achieved by Google Speech to text API.

- Entity-Intent Recognition: Uses the text transcribed to recognize the entity(object) and intent(action). eg: "Please add olives to the pizza".

  Entities: Pizza.

  Sub-Entities: Olives.

Intent: Customize

The object entity recognition is done to locate and classify items. This can be done by tokenizing the words and part of speech tagging to obtain noun and verb in the statements this information extraction step will help us with information extraction.

- Invoice Generation: The entity-intent is queried in the database to create order and invoice. This can be achieved by querying on our data and creating consolidated data frames. At this point, the order is mapped to the items in the menu by the unique identification code which would fetch prices and generate invoices in the end.

  Output:

  Invoice

Invoice		
Pizza	1	$10
+ olives	1	$1.5
Coke	4	$12
Sub	1	$9
Pizza [5 Cheese Pizza]	1	$12
Burger	1	$5
Fries	1	$2.5
Total	10	$53

3 Preliminary Analysis

The GOOGLE NLP API was tested to get the intent and objects from some text orders. Here are those examples:

- **"One Burger and Two Fries"**


```
{
  "entities": [
    {
      "mentions": [
        {
          "text": {
            "beginOffset": 12,
            "content": "burgers"
          },
          "type": "COMMON"
        }
      ],
      "metadata": {},
      "name": "burgers",
      "salience": 0.58150464,
```

```
      "type": "OTHER"
    },
    {
      "mentions": [
        {
          "text": {
            "beginOffset": 24,
            "content": "fries"
          },
          "type": "COMMON"
        }
      ],
      "metadata": {},
      "name": "fries",
      "salience": 0.4184954,
      "type": "CONSUMER_GOOD"
    },
    {
      "mentions": [
        {
          "text": {
            "beginOffset": 7,
            "content": "five"
          },
          "type": "TYPE_UNKNOWN"
        }
      ],
      "metadata": {
        "value": "5"
      },
      "name": "five",
      "salience": 0.0,
      "type": "NUMBER"
    }
  ],
  "language": "en"
}
```


- **I want to order a burger, with coke and fries at the side**


```
        {
  "entities": [
    {
      "mentions": [
        {
```

```
      "text": {
        "beginOffset": 18,
        "content": "burger"
      },
      "type": "COMMON"
    }
  ],
  "metadata": {},
  "name": "burger",
  "salience": 0.36076498,
  "type": "OTHER"
},
{
  "mentions": [
    {
      "text": {
        "beginOffset": 31,
        "content": "coke"
      },
      "type": "COMMON"
    }
  ],
  "metadata": {},
  "name": "coke",
  "salience": 0.33710873,
  "type": "OTHER"
},
{
  "mentions": [
    {
      "text": {
        "beginOffset": 53,
        "content": "side"
      },
      "type": "COMMON"
    }
  ],
  "metadata": {},
  "name": "side",
  "salience": 0.18460932,
  "type": "OTHER"
},
{
  "mentions": [
    {
```

```
          "text": {
            "beginOffset": 40,
            "content": "fries"
          },
          "type": "COMMON"
        }
      ],
      "metadata": {},
      "name": "fries",
      "salience": 0.11751698,
      "type": "CONSUMER_GOOD"
    }
  ],
  "language": "en"
}
```


4 References

- https://cloud.google.com/natural-language/
- https://ai.google/tools/datasets/taskmaster-1
- https://cloud.google.com/text-to-speech/
- https://www.topbots.com/ai-nlp-research-papers-acl2019/