Dialogue systems I

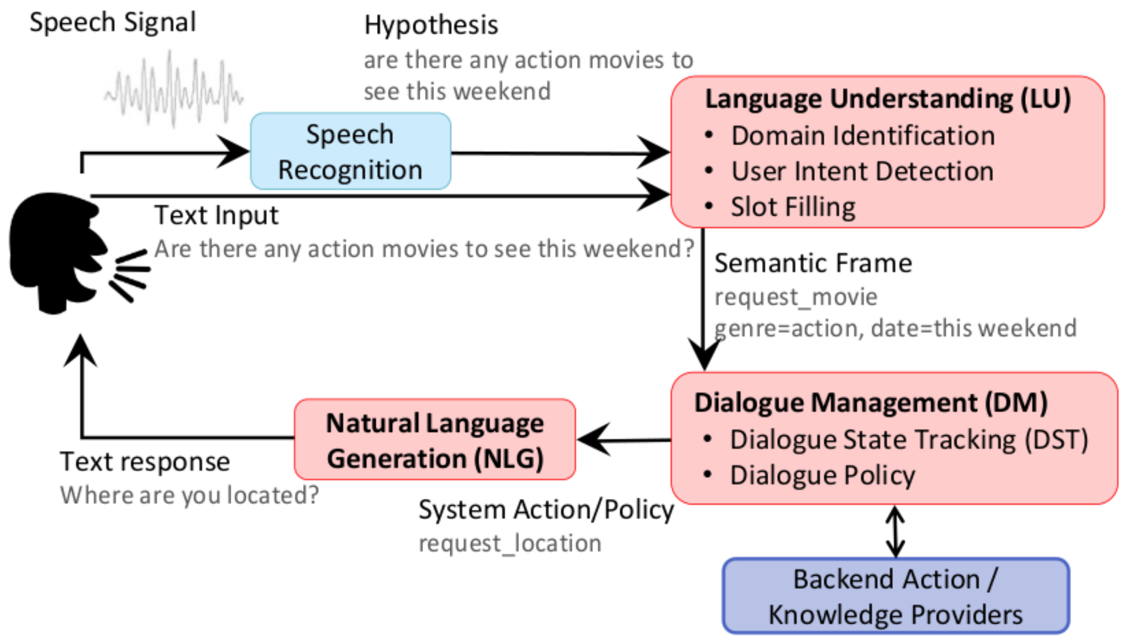
Why dialogue systems?

- Natural language interface to computers
- Real-life applications and business demands
- Many open research questions

Types of dialogue systems

- Basic categories:
  1. Task-oriented dialogue systems
  2. Chit-chat
- Communication domains
- Application areas
- Mode of communication
- Dialogue initiative

Architecture of a Typical Dialogue System



Source: medium

Practice: walk through one of the example domains in PyOpenDial framework

Dialogue datasets

▼ ATIS

[url][paper]


```
{
    "rasa_nlu_data": {
        "common_examples": [
            {
                "text": "i would like to find a flight from charlotte to las vegas that makes a stop in st. louis",
                "intent": "flight",
                "entities": [
                    {
                        "start": 35,
                        "end": 44,
                        "value": "charlotte",
                        "entity": "fromloc.city_name"
                    },
                    {
                        "start": 48,
                        "end": 57,
                        "value": "las vegas",
                        "entity": "toloc.city_name"
                    },
                    {
                        "start": 79,
                        "end": 88,
                        "value": "st. louis",
                        "entity": "stoploc.city_name"
                    }
                ]
            },
            ...
        ]
```

```
    }
}
```

▼ Dialogue state tracking challenge (DSTC)

[url][paper]

User: I'm looking for a cheaper restaurant
`inform(price=cheap)`
System: Sure. What kind - and where?
User: Thai food, somewhere downtown
`inform(price=cheap, food=Thai, area=centre)`
System: The House serves cheap Thai food
User: Where is it?
`inform(price=cheap, food=Thai, area=centre); request(address)`
System: The House is at 106 Regent Street

Annotated dialogue states in a sample dialogue. Underlined words show rephrasings which are typically handled using semantic dictionaries.

▼ CLINC

[url][paper]

Domain	**Intent**	**Query**
BANKING	TRANSFER	*move 100 dollars from my savings to my checking*
WORK	PTO REQUEST	*let me know how to make a vacation request*
META	CHANGE LANGUAGE	*switch the language setting over to german*
AUTO & COMMUTE	DISTANCE	*tell the miles it will take to get to las vegas from san diego*
TRAVEL	TRAVEL SUGGESTION	*what sites are there to see when in evans*
HOME	TODO LIST UPDATE	*nuke all items on my todo list*
UTILITY	TEXT	*send a text to mom saying i'm on my way*
KITCHEN & DINING	FOOD EXPIRATION	*is rice ok after 3 days in the refrigerator*
SMALL TALK	TELL JOKE	*can you tell me a joke about politicians*
CREDIT CARDS	REWARDS BALANCE	*how high are the rewards on my discover card*
OUT-OF-SCOPE	OUT-OF-SCOPE	*how are my sports teams doing*
OUT-OF-SCOPE	OUT-OF-SCOPE	*create a contact labeled mom*
OUT-OF-SCOPE	OUT-OF-SCOPE	*what's the extended zipcode for my address*

Sample queriesfrom CLINC 150 dataset. The out-of-scope queries are similar in style to the in-scope queries.

▼ SNIPS

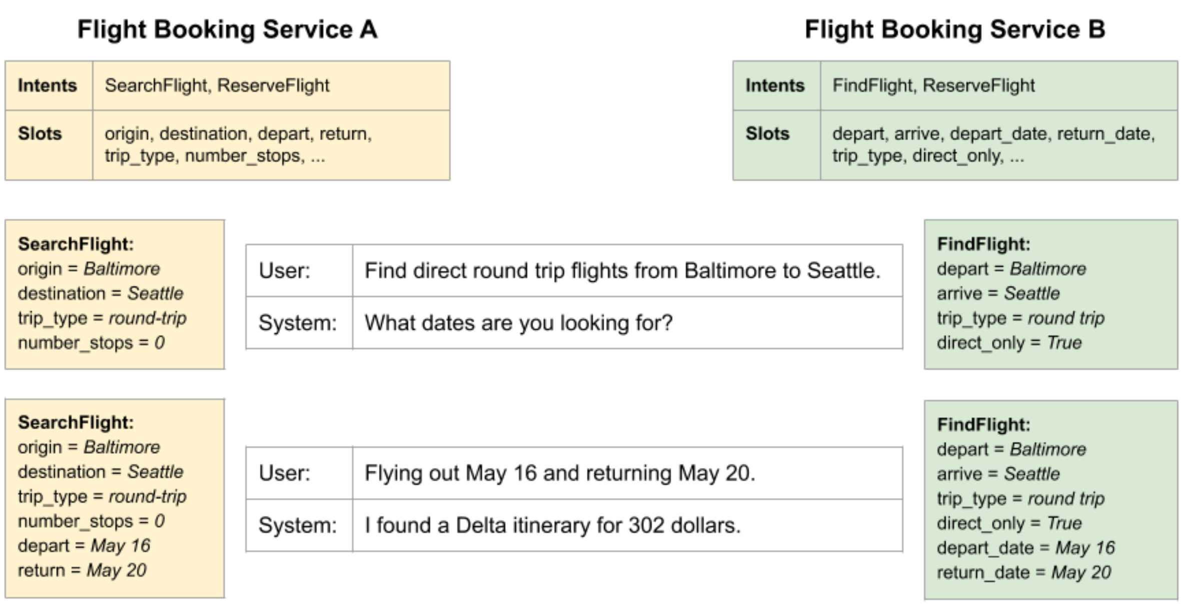
[url][paper]

```
{
     "data": [
       {
```

```
          "text": "add "
        },
        {
          "text": "clem burke",
          "entity": "artist"
        },
        {
          "text": " in "
        },
        {
          "text": "my",
          "entity": "playlist_owner"
        },
        {
          "text": " playlist "
        },
        {
          "text": "Pre-Party R&B Jams",
          "entity": "playlist"
        }
      ]
    },
```


▼ SGD

[url][paper]



The predicted dialogue state for the first two user turns for an example dialogue, showing the active intent and the slot assignments. Note that the representation is conditioned on the schema under consideration.

▼ OpenSubtitles

[paper]

Practice: watch one Poly AI demos, discuss annotation schema

Dataset survey

Multilingual dialogue datasets

The magic triangle of dialogue data collection

Challenges:

- Source of unlabelled data → we need to create the datasets from scratch / re-use logs
- Scripting whole dialogue → controlled environment
- Crowdsourcing → quality control, inter annotation agreement
- Multiple steps: (i) utterances, (ii) labels

Main tasks

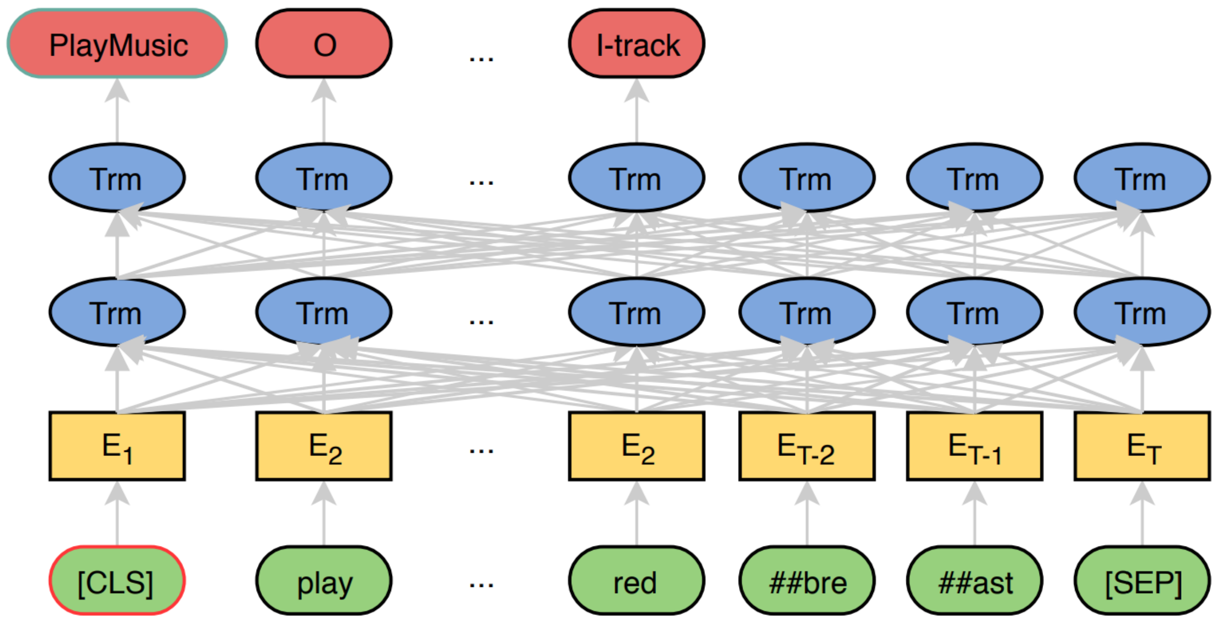
- Intent recognition / detection: at the utterance level classify the utterance into one of intents
- Slot filling: at the token level detect all tokens to fill in the slot

> 📖 Weld, H., Huang, X., Long, S., Poon, J., & Han, S. C. (2021). A survey of joint intent detection and slot filling models in natural language understanding. *ACM Computing Surveys (CSUR).*

query	find	recent	comedies	by	james	cameron
slots	O	B-date	B-genre	O	B-dir	I-dir
intent	find_movie					
domain	movies					

An example of an utterance as semantic frame with domain, intent and IOB slot annotation (from [Hakkani-Tür et al. 2016])

Baseline approach to IR + SF



`JointBERT`: BERT for Joint Intent Classification and Slot Filling

Practice: JointBERT

Practice: Godel demo

Issues in IR

- Ambiguity in interpretation

  `buy bus tickets` VS `buy movie tickets`

  Solution: contrastive learning / triplet loss: maximise the margin between positive and negative pair

$$d(\texttt{anchor}, \texttt{positive}) - d(\texttt{anchor}, \texttt{negative}) \to \max$$

- Emerging intents detection

  Solutions: zero-shot methods, including MRC and capsule networkds

- Unseen intents

  Solution: out of domain techniques

  Пусть $\mathcal{D}_{ID} = \{(x_1, y_1), \dots, (x_n, y_n)\}$ – размеченные пары (высказывание, метка класса, $y_i \in \Upsilon$). Υ , $|\Upsilon| = N$, – множество ID классов, т.е., таких классов, которые диалоговый помощник поддерживает. Допустим, что ID высказывания примерно похожи друг на друга и приходят из какого-то одного распределения. Пусть существует другое, OOD распределение, из которого приходят другие высказывания. Модель детектирования OOD высказываний должна решать следующую задачу:

$$R(x) = \begin{cases} \textbf{reject}, & \text{если } d(x) \geq \theta \\ \textbf{accept}, & \text{иначе} \end{cases}$$

где θ – порог, d – некоторая функция, которая моделирует совместное $d(x, y)$ или условное $d(x|y)$ распределение. В идеале, мы хотим, чтобы $d(x) < d(\hat{x})$ для всех $x \sim P_{IN}$ и $\hat{x} \sim P_{OOD}$.

Maximum Softmax Probability (MSP) [4] использует предобученный классификатор f с *softmax* на выходном слое. Обозначим через $p_y(x)$ вероятность, предсказанную классификатором f для высказывания x и класса y. Чем ниже confidence классификатора, тем вероятность того, что высказывание пренадлежит к OOD распределению.

$$d(x) = 1 - \max_{y \in \Upsilon} p_y(x).$$

Оценки качества OOD детектора: AUROC, AUPR-OOD, FPR@X.

- Multiple intents in an utterance

  Exampe: *find Beyonce's movie and music*

- Imbalanced data

- Lack of labelled training data, small training sets, costly annotations

  Solution: few-shot and zero-shot techniques

- Multi-domain/multi-lingual generalisability

Evaluation metrics

- Intent recognition
  - Intent accuracy
  - Intent precision, recall and F1
  - Recall@k for multi-label intent recognition
  - Student t-test for significance of difference between two models
- Slot filing
  - Span slot precision, recall and F1
  - Micro-averaged F1
  - Slot accuracy.
- Semantic accuracy
  - The number of correctly analysed sentences (IR + SF) divided by the number of sentences

What we covered today

- Types of dialogue systems
- Intro to ToD models
  - two core tasks, intent recognition and slot filling
  - basic approaches to these tasks and issues

Open questions

- Generalisation to new domains and languages
- Efficient architectures for on-device deploy
- User temporal and local context