Chatbots

«Высшая школа экономики»

25.11

Outline

Чатботы

Task-oriented

Other

Chatbots

Chatbots are small computer programs used to simulate the method of human conversation and interact with real people automatically to help them with their issues and complete their tasks.

- Button/Menu-Based Chatbots
- Keyword-Based Chatbots
- Natural Language Processing Chatbots

Chatbots

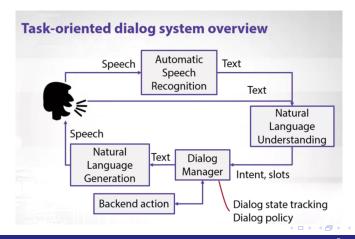
- Потребность человека в общении (болталка Алисы)
- Интерфейс приложения
- Голосовое управление (Siri, Alexa, Алиса)

- О чём люди должны спрашивать чатбота?
- Какие интенты внутри поддерживаемых доменов бот должен обрабатывать?
- Какие сущности могут быть в каждом сценарии? Какие ограничения?
- Как будут обрабатываться ошибки?
- Как будут обрабатываться неподдерживаемые домены?
- Что будет, если на бота будут материться? Оскорблять?
 Приставать?
- Проработка персонажа.



- Open domain чатбот умеет поддерживать разговор на любую тему.
- Closed domain чатбот умеет поддерживать беседу на ограниченное количество тем.

Диалог vs. беседа
Диалог — единичная реализация сценария. Беседа = диалог +
диалог + диалог
Поддержка состояний — когда бот запоминает предыдущие
реплики человека. Это можно делать как в рамках одного
диалога, так и в рамках всей беседы.



Откуда брать ответы?

- Retrieval-based
- Generative-based

Как ответить?

- Распознать интенцию
- Подготовить формы ответов со слотами для заполнения
- Заполнить форму

Пример формы:

- Я хочу взять билет на самолёт из Москвы в Лиссабон на 1 ноября.
- Билет из Москвы в Лиссабон на 1 ноября будет стоить 13 000 рублей.

Как распознать интенцию?

- Пдойдет любая модель на веторизованных текстах (bow, tf-idf, n-grams) - baseline
- RNN (LSTM, GRU)
- CNN (предпочтительно)

Как найти метки?

- regex
- CRF
- RNN (seq2seq)
- CNN (seq2seq)
- any seq2seq model..

Slot tagging

ATIS dataset

- Airline Travel Information System
- · Collected in 90s
- 4978 context independent utterances
- 17 intents, 127 slot labels
- State-of-the-art: 1.79% intent error, 95.9 slots F1

Utterance	show	flights	from	Seattle	to	San	Diego	tomorrow
Slots	0	0	0	B-fromloc	0	B-toloc	I-toloc	B-depart_date
Intent	Flight							

Поддерржка состояния:

- User: Give me directions from Los Angeles
 - Intent classifier: nav.directions
 - Slot tagger: @FROM{Los Angeles}
 - Dialog manager: required slot is missing, where to?
- Agent (assistant): Where do you want to go?
- User: San Francisco
 - Intent classifier: nav.directions
 - Slot tagger: @TO{San Francisco}
 - Dialog manager: okay, here's the route
- Agent (assistant): Here's the route





Pipelilne recap:

- domain
- intention
- slot tagging

seq2seq-нейросети

Нейросеть вида "sequence to sequence" принимает на вход некоторую последовательность чего угодно (чисел, пикселей, символов) и порождает другую последовательность чего угодно, сооветствующую первой. Нас скорее интересуют последовательности букв или слов, хотя в отдельных случаях в качестве токенов используются целиком вопросы и ответы.

seq2seq-нейросети

Как реализовать?

- Берём обучающий корпус запросов и ответов.
- Тренируем рекуррентную (LSTM или GRU) нейросеть, которая принимает на вход последовательность токенов запроса и выдаёт в качестве ответа некую другую последовательность токенов.

seq2seq-нейросети

Когда мы можем такое делать?

- Есть много данных.
- Мы уверены в качестве используемых данных.
- Нет ресурсов на разработку правил.
- Чатбот скорее развлекательный.
- Нам не очень жалко свою репутацию.