

Question-Answering Systems for Chat Bots

Антон Киселев, EORA Data Lab



Категории вопрос-ответных систем (1/2)

По форме вопроса:

- Выбор из вариантов
- Структурированный вопрос
- Произвольный вопрос

По форме ответа:

- Да / Нет
- Готовые ответы: FAQ
- Извлечение факта из текста
- Генерация ответа: машинный перевод

Категории вопрос-ответных систем (2/2)

По типу ответа:

- Factoid / Non-Factoid QA: ответить конкретным фактом или нет
- Open Domain / Closed Domain: ответы на произвольную тему или на заданную

По представлению фактов:

- Текст
- Сущности с полями
- Сущности со связями между собой

Вопросы и ответы в архитектуре чат-бота

Общий подход:

- Пользователь говорит реплику
- Реплика классифицируется по интену (намерению)
- Если интен - задать вопрос, то он передается вопрос-ответной системе

Модификации:

- Несколько отвечающих систем под каждую тематику вопроса
- Ветвление при ответе: уточняющие вопросы

Ответ на вопрос с точки зрения ML

Задача поиска ответа сводится к следующему:

- Классификация
 - Какая категория у вопроса?
 - Насколько данный ответ релевантен вопросу?
- Обучение представлений
 - Какие похожие вопросы уже задавались?
 - Какие ответы были даны на похожие вопросы?
- Ранжирование
 - В каком порядке предложить готовые ответы пользователю?

Классификация для вопросов

Задача: классификация вопроса либо пары вопрос-ответ

Модели:

- Логистическая регрессия
- Метод ближайших соседей
- Градиентный бустинг

Признаки:

- Мешок слов
- Мешок представлений слов
- Представление предложения
- Внешние признаки: поведение пользователя

Представления предложений (1/3)

Задача: нахождение векторных представлений, в которых похожие предложения близки, а разные далеки.

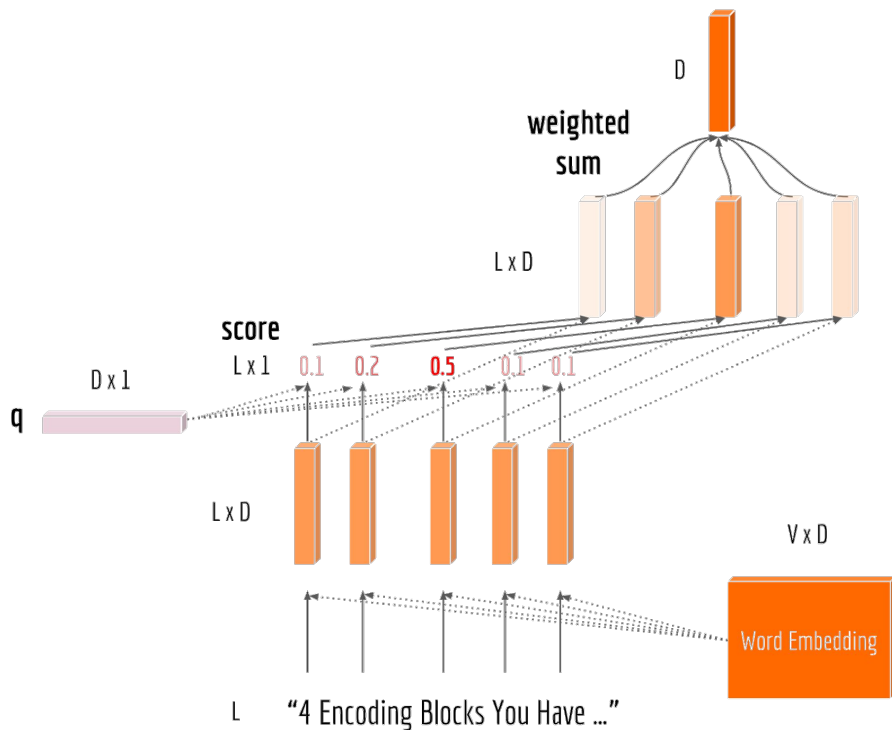
Архитектура модели:

- представления слов: FastText, ELMo, ULMFiT
- преобразование представлений: InferSent, Shallow-and-Wide CNN
- агрегация представлений: concat pooling, attentive pooling [1], sum + projection [2]

Представления предложений (2/3)

Attentive Pooling [1]

Агрегирует
представления
отдельных слов в вектор
всего предложения



Представления предложений (3/3)

Преимущества:

- Ранжирование зашито в модель
- Можно использовать индексы для поиска ближайших соседей
- Представления можно использовать для других задач

Недостатки:

- Не отражают всю суть предложения
- С трудом работают с редкими словами

Сопоставление представлений

Идея: комбинировать представления отдельных слов вопроса и ответа

Варианты реализации:

- Парное комбинирование представлений слов вопроса и ответа
- Механизм внимания между представлением вопроса и представлениями слов ответа

Реализации моделей

- AllenNLP [3]
 - Библиотека со множеством моделей глубокого обучения для решения NLP задач
 - Архитектура с агентами, отвечающими на реплики
- DeepPavlov [4]
 - Библиотека для создания чат-ботов
 - Много различных моделей для различные задач, в том числе для ответа на вопросы
 - На момент доклада трудно комбинировать модели друг с другом
- MatchZoo [5]
 - Модели для сопоставления предложений
 - Нет примеров для встраивания в свои решения

Спасибо за внимание!



Антон Киселев
Telegram: @a_kiselev
strawberrypie@eora.ru

ИСТОЧНИКИ

[1] [Sequence Encoding Blocks You Must Know Besides RNN/LSTM in Tensorflow](#)

[2] HyperQA: <https://arxiv.org/abs/1707.07847>

[3] AllenNLP: <https://github.com/allenai/allennlp>

[4] DeepPavlov: <https://github.com/deepmipt/DeepPavlov>

[5] MatchZoo: <https://github.com/faneshion/MatchZoo>