

УДК 004.8

РАСПОЗНАВАНИЕ ГОЛОСОВЫХ КОМАНД ДЛЯ УПРАВЛЕНИЯ УСТРОЙСТВАМИ УМНОГО ДОМА

Пантюхов А.Ю. (Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Научный руководитель – к.т.н., доцент Муромцев Д.И.

(Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

В работе рассматривается реализация сервиса распознавания голосовых команд для управления устройствами умного дома. Описывается подготовка датасета и обучение моделей.

Введение.

Системы голосового управления стали одним из важных и активно развивающихся направлений индустрии, эти системы предназначены для обеспечения более естественного взаимодействия человека и компьютерной системы. Однако большинство готовых решений является проприетарными.

Для интерпретации команд система должна быть в состоянии извлечь информацию из команды на естественном языке.

Подготовка датасета.

Для обучения моделей необходимы данные, поэтому был собран и размечен собственный датасет. При составлении датасета часть команд была написана вручную, а затем при помощи инструмента Chatito DSL датасет был автоматически расширен, путём вариации формулировок команд и использования синонимов. Датасет включает команды на английском и русском языках, для разметки использовалась схема BIO. Полученный датасет содержит более 11000 фраз.

Методы понимания естественного языка.

Для анализа команды пользователя необходимо сначала выполнить перевод голосового сигнала в текст, для этой цели был использован Google Speech Recognition. Вторым этапом является распознавание текстового представления команды (NLU).

В основе модуля NLU лежит языковая модель BERT (мультиязычная), также есть возможность использовать более лёгкую и быструю модель ALBERT (только английский язык). Модель одновременно выполняет две задачи: определяет намерение (интент) пользователя и выделяет из текста команды значения конкретных параметров (слоты).

При обучении моделей применялась кросс-валидация, а качество обученных моделей оценивалось по метрикам точности (ассигасу) – для классификации интенгов и F-меры (f1-score) – для заполнения слотов.

Экспериментальная проверка.

Средние значения метрик, полученных на тестовом наборе: точность классификации интенгов – 99.6%, F-мера определения слотов – 99.8%. Однако, несмотря на высокое среднее значение F-меры, для наиболее сложных фраз (неявные интенты) наблюдалось снижение значения метрики вплоть до 78%.

Полученные модели способны распознавать не только прямые команды вроде «Включи свет», но и команды, заданные в неявной форме, например, «Как-то здесь холодно». Распознанные моделью интент и слоты преобразуются в команду стандартизированного вида (device, action, parameter, value).

На основе обученных моделей реализован REST сервис, анализирующий входящий звуковой сигнал и возвращающий команду в виде стандартизированной структуры. Также подготовлен веб интерфейс для демонстрации возможностей системы.

Выводы.

Таким образом, был разработан сервис распознавания команд для управления устройствами умного дома, всего сервис способен распознать 24 типа команд. Для обучения моделей, лежащих в основе сервиса, был подготовлен датасет из команд на английском и русском языках.

Пантюхов А.Ю. (автор)

Подпись

Муромцев Д.И. (научный руководитель)

Подпись