

Секция «Искусственный интеллект и анализ данных в космических исследованиях»

Распознавание голосовых команд на Raspberry Pi: производительность моделей и интеграция с механическим манипулятором

Научный руководитель – Шишкин Алексей Геннадиевич

Литвинов Иван Борисович

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова, Факультет
космических исследований, Москва, Россия

E-mail: litvan007@gmail.com

На Международной космической станции (МКС) использование голосового управления механическими манипуляторами сталкивается с проблемой сильной зашумлённости аудиосигналов [2], что существенно затрудняет точное распознавание команд. Настоящая работа представляет собой финальный этап разработки прототипа механического манипулятора, предназначенного для помощи экипажу МКС и способного распознавать 18 голосовых команд.

Предложенная архитектура распознавания голосовых команд разделена на две последовательные модели: модель определения голосовой активности (VAD), реализованную с использованием сверточной нейронной сети [1] (точность: 0.982, полнота: 0.92, f-мера: 0.948), и модель распознавания речи (ASR), сочетающую сверточные слои, LSTM и механизм внимания (точность: 0.95, полнота: 0.94, f-мера: 0.94). В отличие от end-to-end решений, данный подход обеспечивает гибкость и более точное распознавание в условиях сильного шума.

Разработанные модели были конвертированы в универсальный формат ONNX [3], что позволило эффективно использовать их на Raspberry Pi 4 [4]. Тестирование показало, что среднее время обработки одной голосовой команды составило 1.5 секунды для модели VAD (размером 37 МБ) и 2 секунды для ASR-модели (размером 95 МБ). Такое время анализа подходит лишь для демонстрации прототипа. В реальных условиях необходимо использовать GPU или специализированные устройства, например NVIDIA Jetson Nano [5], что ускорит обработку, но увеличит стоимость системы.

Итоговое тестирование подтвердило работоспособность всех компонентов системы. Механический манипулятор от производителя Waveshare [6] был успешно собран и интегрирован в единый контур с Raspberry Pi и моделями в формате ONNX, продемонстрировав стабильное и корректное выполнение всех 18 голосовых команд согласно разработанному дизайну. При этом результаты, полученные ранее при тестировании моделей на тестовом наборе данных, полностью подтвердились в ходе реальных экспериментов на готовой системе.

Источники и литература

- 1) Процоров С. Д., Шишкин А. Г. Сегментация зашумленных речевых сигналов // Искусственный интеллект и принятие решений – 2021. №1. – С. 75-85.
- 2) Limardo N., Goodman J., Grosveld F. Status Crewmember Noise Exposures on the International Space Station // 22nd International Congress on Sound and Vibration. - 2015.
- 3) ONNX. Open Neural Network Exchange. Доступно по ссылке: <https://onnx.ai/>
- 4) Raspberry Pi Foundation. Raspberry Pi 4 Model B. Доступно по ссылке: <https://www.raspberrypi.org/products/raspberry-pi-4-model-b/>

- 5) NVIDIA. NVIDIA Jetson Nano. Доступно по ссылке: <https://developer.nvidia.com/embedded/jetson-nano>
- 6) Waveshare. Механический манипулятор RoArm-M1. Доступно по ссылке: <https://www.waveshare.com/product/raspberry-pi/robots/robot-arm-control/robot-arm-for-pi.htm>