

Рецензия на проектную работу по ASR для осетинского языка

Представленная проектная работа посвящена актуальной теме – оценке эффективности многоязычной модели автоматического распознавания речи (ASR) *facebook/mms-1b-all* при работе с малоресурсными языками, а конкретно осетинским. Автор использует интересный подход, тестируя модель на реальных полевых аудиоданных, собранных во время экспедиций во Владикавказ в 2023-2025 годах, что позволяет проверить работу модели на материале, существенно отличающемся от тренировочных данных.

Автор грамотно структурировал исследование, последовательно изложив информацию о доступных наборах данных (Mozilla Common Voice), особенностях используемой модели и характеристиках собранных полевых записей. Особенно ценным представляется привлечение разнообразных дикторов (22 человека различных возрастов, пола и диалектных групп), что позволило оценить способность модели к обобщению и провести многостороннее тестирование.

Интересен проведенный анализ распознавания отдельных фонем, где автор отмечает, что модель относительно успешно распознает сибиллянты, губные согласные и г, но испытывает трудности с гласными и многими другими согласными. Наблюдение о передаче осетинского æ как а, о или е в зависимости от контекста также представляет фонетический интерес.

В методологическом плане можно отметить, что исследование в значительной степени ограничено доступными ресурсами. Модель обучалась на небольшом объеме данных (1,2 часа) преимущественно от одного диктора, что неизбежно сказывается на качестве распознавания. Однако, в работе также недостаточно обоснован выбор архитектуры *facebook/mms-1b-all* и её преимущества перед современными моделями, такими как Whisper, которые показывают высокую эффективность для малоресурсных языков. Помимо этого, под вопрос встает предобработка данных – не совсем ясно, ограничивается ли она только ресемплингом до 16 кГц, и применялись ли стандартные техники, такие как нормализация громкости или спектральная аугментация. Также в работе не приводятся количественные метрики оценки качества ASR (например, WER или CER), что несколько затрудняет объективное сравнение результатов.

Заключительная часть работы с планами по созданию более представительного датасета для осетинского ASR демонстрирует понимание автором ключевых проблем и путей их решения. Это показывает, что проведенное исследование не только выявило ограничения существующей модели, но и заложило основу для дальнейшей работы в этом направлении.

В целом, работа представляет собой ценное пилотное исследование в области ASR для малоресурсного языка и закладывает основу для дальнейших исследований в этом направлении. Особенно ценным является внимание автора к фонетическим особенностям распознавания отдельных звуков, что может быть полезно для разработки специализированных моделей ASR для осетинского и других малоресурсных языков.