

## СРАВНИТЕЛЬНЫЙ АНАЛИЗ МОДЕЛЕЙ РАСПОЗНАВАНИЯ РУССКОЯЗЫЧНОЙ РЕЧИ НА ПРИМЕРЕ ТЕЛЕФОННЫХ ЗВОНКОВ

Газизулин Р.М., магистрант  
Хартьян Д.Ю., к.тех.н., доцент,  
ТюмГУ, г. Тюмень, Россия

**Аннотация.** Рассмотрены модели Wav2Vec2-Large-Ru-Golos, Wav2Vec2-Large-Ru-Golos-With-LM, Whisper-Turbo, GigaAM-RNN, GigaAM-CTC для распознавания русскоязычной речи. Приведены результаты по точности, скорости и устойчивости к шуму. Выявлены сильные и слабые стороны каждой модели.

**Ключевые слова:** транскрибация, WER, телефонные звонки.

**Введение.** Автоматическое распознавание речи (АРР) является ключевым инструментом в современной обработке речевых данных, с широким применением в сферах обслуживания клиентов, управления интеллектуальными устройствами и других бизнес-процессах. В статье "Исследование возможностей и оценка качества распознавания звучащей речи различными нейросетевыми моделями" был проведен анализ производительности ряда нейросетевых моделей на стандартных датасетах. Авторы сделали акцент на сравнении архитектур и подходов, определяя области их эффективного применения. Ключевыми выводами стали выявление преимуществ некоторых моделей для специфических условий распознавания и рекомендации по выбору моделей в зависимости от задач [2]. Отличительной особенностью представленной статьи является анализ моделей АРР на собственных данных, что делает результаты более релевантными для реальных бизнес-процессов, где качество распознавания речи напрямую связано с особенностями речевой коммуникации в телефонных разговорах. Таким образом, в рамках данного исследования проводится:

Автор: Газизулин Р.М.,Хартьян Д.Ю.  
07.12.2024 20:31 -

---

1. Сравнительный анализ моделей на телефонных звонках — особом формате речи, отличающемся качеством записи и спецификой диалогов.

2. Оценка устойчивости моделей к шуму, что является важным фактором для применения в реальных условиях, особенно в контактных центрах и системах голосового управления.

...

полный текст во вложении