

Сравнительный анализ языковых моделей

Афанасьев Алексей Игоревич 217

3 апреля 2025 г.

1 Введение

В данном отчете проводится сравнительный анализ языковых моделей Qwen-2.5-1.5B и RuadaptQwen2.5-1.5B по показателю точности ответов на двух различных датасетах. Результаты представлены в виде сравнительных таблиц.

2 Методология

- Оценка проводилась по метрике точности (Accuracy)
- Использованы два независимых датасета:
 - **RuCola**: Датасет, который характеризует задачу понимания естественного языка и ответов на вопросы на русском языке. Разделен на части Validation, Train, Test
 - **glue/Cola**: То же самое, но для Английского. Является оригинальным датасетом. Разделен на части Validation, Train, Test
- Для каждой модели вычислен процент правильных ответов

3 Результаты

3.1 Датасет RuCola

Таблица 1: Точность моделей на RuCola

Метрика	Qwen	RuadaptQwen
Точность (%)	58.5	71

3.2 Датасет Cola

Таблица 2: Точность моделей на glue/cola

Метрика	Qwen	RuadaptQwen
Точность (%)	74	77.5

4 Сравнительный анализ

Таблица 3: Сравнение моделей по обоим датасетам

Характеристика	Qwen	RuadaptQwen
Средняя точность (%)	66.25	74.25
Разница в точности (%)	-8	+8

Комментарий: Ruadapt версия модели значительно лучше справляется с задачей понимания естественного языка как на русском так и на английском.

5 Заключение

На основании представленных данных можно сделать следующие выводы:

- Модель RuadaptQwen показала более высокую точность на обоих датасетах
- Разница в точности составила в среднем 8%
- Наибольшее расхождение наблюдалось на датасете Rucola (12.5%)