

АВТОМАТИЧЕСКОЕ ОЦЕНИВАНИЕ СОЧИНЕНИЙ НА АНГЛИЙСКОМ ЯЗЫКЕ

Мэн Сыфэй

Студент

Факультет компьютерных наук НИУ ВШЭ, Москва, Россия

E-mail: smen@edu.hse.ru

Научный руководитель — Ильвовский Дмитрий Алексеевич

В последние годы наблюдается значительный рост популярности английского языка, что стимулирует развитие нейросетевых подходов для автоматической проверки письменных работ. Однако большинство существующих моделей предполагают, что сочинения в тестовой выборке написаны на те же темы, что и в обучающем наборе — такое допущение не соответствует реальным условиям применения.

Целью настоящего исследования является разработка нейронной сети, способной эффективно оценивать англоязычные тексты по темам, отсутствующим в обучающей выборке. Предлагаемая модель должна комплексно оценивать все критерии IELTS, включая соответствие заданной теме, структурную организацию работы, лексическое разнообразие и грамматическую корректность.

В рамках исследования был сформирован новый корпус данных, состоящий из 15,803 сочинений в формате IELTS, собранных с профильных образовательных ресурсов. В процессе анализа выборки было выявлено отсутствие примеров низкого качества, которые впоследствии были синтезированы с применением различных стратегий аугментации для имитации текстов, написанных студентами с недостаточным уровнем владения английским языком. После дополнения общий объем корпуса достиг 17,370 сочинений.

Проведенный статистический анализ распределения оценок в собранном корпусе продемонстрировал соответствие официальным данным экзаменационных комиссий. Для объективной оценки эффективности разработанных моделей были использованы квадратичный взвешенный каппа и средняя абсолютная ошибка, поскольку задача рассматривалась в регрессионной парадигме.

В качестве референсных решений были воспроизведены современные архитектуры, представленные в актуальных научных публикациях [1] [2] [3]. Кроме того, было проведено сравнительное исследование различных трансформерных архитектур: BERT, ALBERT, GPT-2, ELECTRA. Результаты экспериментов показали,

что ELECTRA демонстрирует оптимальное соотношение качества и вычислительной эффективности, что обусловило её выбор в качестве базовой модели. Внесенные архитектурные модификации позволили сохранить производительность ELECTRA на уровне state-of-the-art решений в данной предметной области.

В заключение, основные результаты исследования включают:

- Формирование нового репрезентативного корпуса IELTS-сочинений с детализированными оценками по критериям;
- Разработку и имплементацию эффективных стратегий аугментации данных для повышения робастности модели;
- Проведение комплексного анализа дискриминативных моделей и их архитектурных модификаций.

Итогом работы стала вычислительно эффективная модель, которая демонстрирует сопоставимое качество с более ресурсоемкими аналогами.

Литература

1. Cho B., Jang Y., Yoon J. Rubric-Specific Approach to Automated Essay Scoring with Augmentation Training // arXiv, Sep 2023
2. Ridley R., He L., Dai X., Huang S., Chen J. Automated Cross-prompt Scoring of Essay Traits // Proceedings of the AAAI Conference on Artificial Intelligence, 2021, Vol. 35, No. 15, P. 13745–13753.
3. Xue J., Tang X., Zheng L. A Hierarchical BERT-Based Transfer Learning Approach for Multi-Dimensional Essay Scoring // IEEE Access, 2021, Vol. 9, P. 125403–125415.