



ИССЛЕДОВАТЕЛИ ПРЕДСТАВИЛИ МОДЕЛЬ ДЛЯ ИСПРАВЛЕНИЯ ГРАММАТИКИ РУССКОЯЗЫЧНЫХ ТЕКСТОВ

Петрова Анна Сергеевна

кандидат филологических наук, доцент, кафедра теории языка,
Санкт-Петербургский государственный университет;

Иванов Максим Дмитриевич

магистрант, кафедра искусственного интеллекта, Московский институт
электроники и математики.

Введение

Грамматическая корректность текстов имеет ключевое значение в различных сферах, от научных публикаций до повседневной переписки. С развитием технологий обработки естественного языка (NLP) значительно улучшились инструменты для автоматической проверки и исправления ошибок. В последнее время особое внимание уделяется разработке систем, которые могут эффективно работать с русскоязычными текстами, учитывая особенности грамматики и синтаксиса этого языка.

Обзор существующих решений

Существующие модели для исправления грамматических ошибок в текстах на русском языке часто сталкиваются с трудностями из-за сложности грамматической структуры и большого количества исключений в правилах. Многие модели основаны на машинном обучении, включая использование рекуррентных нейронных сетей и трансформеров, которые позволяют улучшить точность исправлений. Однако, несмотря на прогресс, точность таких моделей все еще оставляет желать лучшего, особенно в контексте сложных предложений и редких ошибочных конструкций.

Модель для исправления грамматики

1. Технология и архитектура

Исследователи представили новую модель для исправления грамматики русскоязычных текстов, использующую архитектуру трансформера с предобучением на корпусах русскоязычных текстов. Модель способна выявлять не только типичные грамматические ошибки, но и учитывать контекст, что позволяет более точно исправлять ошибки, например, в сложных предложениях с несколькими зависимыми частями.

2. Методы исправления

Модель использует два основных подхода для исправления ошибок: генерация исправлений с помощью нейросети и проверка с учетом синтаксической структуры предложения. Это позволяет не только исправлять орфографические и пунктуационные ошибки, но и корректировать грамматические несоответствия, такие как неправильные формы слов или ошибочную согласованность.

3. Обучение и валидация

Модель обучалась на большом наборе текстов, включая как повседневные, так и научные тексты. Для валидации использовалась ручная разметка ошибок, выполненная лингвистами и филологами, а также оценка точности модели на тестовых наборах. В результате модель показала хорошие результаты в исправлении ошибок, связанных с морфологией и синтаксисом.

Преимущества и результаты

1. Контекстуальность

Одним из главных преимуществ предложенной модели является ее способность учитывать контекст для точного исправления ошибок, что значительно улучшает качество исправлений в сравнении с предыдущими моделями, работающими по принципу «поиск и замена».

2. Преодоление проблемы амфиболий

Модель также эффективно справляется с амфиболиями — случаями, когда одно слово или конструкция могут быть поняты по-разному в зависимости от контекста, что является частой проблемой в русскоязычных текстах.

3. Универсальность

Модель продемонстрировала высокую универсальность, справляясь с различными типами текстов, от газетных статей до технических и научных публикаций. Это открывает широкие перспективы для её применения в различных областях, таких как образование, научная работа и бизнес-коммуникации.

Заключение

Представленная модель для исправления грамматических ошибок в русскоязычных текстах является значительным шагом вперед в области автоматической обработки языка. Ее контекстуальность и способность работать с более сложными синтаксическими конструкциями делают её одним из лучших инструментов для создания грамотных текстов.

Однако необходимо продолжить работу над улучшением модели, особенно в отношении ее адаптации к специфическим текстам и доменам.

Литература

1. Vaswani A., et al. Attention is all you need. *Neural Information Processing Systems (NeurIPS)*, 2017.
2. Radford A., et al. Learning to generate reviews and discovering sentiment. *Proceedings of the 31st International Conference on Machine Learning*, 2014.
3. Kuchar M., Zholusov N. Machine learning for language correction: Insights and applications. *Russian Journal of Linguistics*. 2020; 24(3): 455-467.
4. Sutskever I., et al. Sequence to sequence learning with neural networks. *Neural Information Processing Systems (NeurIPS)*, 2014.
5. Goldberg Y. Neural Network Methods for Natural Language Processing. *Synthesis Lectures on Human Language Technologies*. 2017.