

НЕЙРОСЕТЕВЫЕ ЯЗЫКОВЫЕ МОДЕЛИ

Рахматхужаев А.А.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь)*

Арцыменя Д.Ф. – ст. преподаватель

В статье рассматривается принцип работы и разнообразные области практического применения нейросетевых языковых моделей.

В последние годы нейросетевые языковые модели стали ключевым элементом в области обработки текста. Они явились прорывом в обработке естественного языка и стали катализатором революции в этой области, изменяя способ, которым мы взаимодействуем с текстовой информацией. Их модели основаны на искусственном интеллекте и предназначены для анализа, понимания и генерации текста на естественных языках. Они используют глубокие нейронные сети и обучаются на больших объемах текстовых данных. Принципы работы и разнообразные области практического применения нейросетевых языковых моделей рассматриваются в данной статье.

Одна из главных характеристик этих моделей – их основа на глубоком обучении. Глубокое обучение – это общий термин для методов машинного обучения, использующих "глубокие" нейронные сети. Сегодня глубокое обучение является одной из наиболее заметных областей машинного обучения, так как его успехи в таких областях, как Компьютерное зрение, обработка естественного языка, а также при применении к обучению с подкреплением, сценариям, таким как игра в игры, принятие решений и моделирование. Важным элементом успеха глубокого обучения является доступность данных, вычислений, программных платформ и сред выполнения, которые упрощают создание моделей нейронной сети и их выполнение для вывода [1].

Важной особенностью нейросетевых языковых моделей является их способность анализировать контекст. Они не только учитывают отдельные слова, но и понимают их значения в контексте предложения. Благодаря этому они способны понимать смысл предложения в целом. Кроме того, нейросетевые языковые модели могут переносить знания. Обученные на одном наборе данных, они могут применять свои знания к различным задачам обработки текста. Это делает их гибкими и универсальными инструментами [2].

Работа нейросетевых языковых моделей основана на нескольких ключевых принципах:

1. Архитектура глубоких нейронных сетей: NLP-модели включают в себя множество слоев нейронных сетей. Эти сети строятся на принципах сверточных или рекуррентных нейронных сетей, которые позволяют модели анализировать и извлекать паттерны из текстовых данных.

2. Обучение на множестве данных: Для успешной работы модели необходимо обучить на больших объемах текстовых данных. Это позволяет модели понимать разнообразные языковые структуры и особенности.

3. Внимание к контексту: NLP-модели способны учитывать контекст в анализе текста. Это означает, что они не только анализируют отдельные слова, но и учитывают семантические и синтаксические связи между ними.

4. Перенос знаний: NLP-модели обучаются на основе большого объема текста, что позволяет им "переносить знания" о языке и его структурах. Это делает их способными адаптироваться к разнообразным задачам обработки текста.

5. Генерация и классификация: Модели способны не только анализировать текст, но и генерировать новый текст, а также классифицировать текст по различным параметрам (например, тональность, тема и другие).

Принципы работы нейросетевых языковых моделей обеспечивают им способность адаптироваться к разнообразным задачам и делают их мощными инструментами для обработки текста на естественных языках [3].

Нейросетевые языковые модели (НЛМ) находят широкое практическое применение в русском языке в различных областях и задачах. Представляются следующие конкретные примеры использования:

1. Машинный перевод: НЛМ применяются для автоматического перевода текста с одного языка на другой. Они основаны на глубоких нейронных сетях, обучаемых на параллельных корпусах текстов.

Продвинутые архитектуры, такие как трансформеры, используются для моделирования длинных зависимостей и обеспечения высокого качества переводов.

2. Генерация текста: НЛМ могут быть применены для автоматической генерации текстового контента. Они обучаются на больших объемах данных и используются для создания статей, новостей, рекламных текстов и других текстовых материалов. Генерация текста основана на моделировании вероятностного распределения слов и последовательностей слов.

3. Автокоррекция и автодополнение: НЛМ применяются в системах автокоррекции и автодополнения текста. Они используют контекстную информацию и предсказывают вероятные варианты исправления ошибок или дополнения предложений. Для этого модели обучаются на больших корпусах текстов и используют контекстуальные эмбединги слов.

4. Анализ тональности и сентимента: НЛМ применяются для анализа эмоциональной окраски текста и определения его тональности или сентимента. Они обучаются на размеченных данных, где каждый текст имеет метку положительной, отрицательной или нейтральной тональности. Модели используются для классификации текстов по этим категориям.

5. Вопросно-ответные системы: НЛМ могут быть использованы для создания систем вопросно-ответной обработки текста. Они обучаются на парах вопрос-ответ и могут генерировать ответы на новые вопросы. Это достигается путем моделирования контекстуальных представлений вопросов и используя их для генерации соответствующих ответов.

6. Анализ и обработка текстовых данных: НЛМ применяются для классификации текстовых данных по различным категориям. Это может включать классификацию отзывов, новостей, комментариев и других текстовых материалов. Обучение моделей основано на размеченных данных и использует методы глубокого обучения для извлечения признаков из текста.

7. Автоматическая генерация резюме: НЛМ могут быть применены для автоматической генерации резюме на основе входного текста. Модели обучаются извлекать ключевую информацию из текста и формулировать ее в сжатом и информативном формате, соответствующем требованиям резюме [4].

Таким образом, нейросетевые языковые модели (НЛМ) представляют собой мощный инструмент с широким спектром практического применения в русском языке. Они демонстрируют потенциал в различных областях, включая машинный перевод, генерацию текста, автокоррекцию и автодополнение, анализ тональности и сентимента, вопросно-ответные системы, анализ и обработку текстовых данных, а также автоматическую генерацию резюме.

Список использованных источников:

1. Что такое глубокое обучение? [Электронный ресурс] / Microsoft Ignite. – Минск, 2021. – Режим доступа: <https://learn.microsoft.com/ru-ru/dotnet/machine-learning/deep-learning-overview> – Дата доступа: 14.10.2023.
2. Attention Is All You Need. [Электронный ресурс] / Кристиан Лин. – Минск, 2023. – Режим доступа: <https://medium.com/@crlc112358/transformer-attention-is-all-you-need-8e617307874a>. – Дата доступа: 18.10.2023
3. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [Электронный ресурс] / Нихил Верма. – Минск, 2022. – Режим доступа: <https://lih-verma.medium.com/pre-training-of-deep-bidirectional-transformers-for-language-understanding-bert-2a8fcc61a22a>. – Дата доступа: 25.10.2023.
4. Russian SuperGLUE [Электронный ресурс] / A Russian Language Understanding Evaluation Benchmark. – Минск, 2023. – Режим доступа: <https://publications.hse.ru/pubs/share/direct/419126604.pdf> – Дата доступа: 05.11.2023.