

Выполнил(а) Казарин А. М., № группы P3108, оценка _____
Фамилия И.О. студента не заполнять

Название статьи/главы книги/видеолекции О методах позиционного кодирования в Transformer		
ФИО автора статьи (или e-mail) murat_apishev	Дата публикации (не старше 2020 года) "12" декабря 2023 г.	Размер статьи (от 400 слов) 2247
Прямая полная ссылка на источник или сокращённая ссылка (bit.ly, tr.im и т.п.) https://habr.com/ru/articles/780116/		
Теги, ключевые слова или словосочетания Машинное обучение, Искусственный интеллект, Natural Language Processing, кодирование		
Перечень фактов, упомянутых в статье (минимум три пункта) <ol style="list-style-type: none"> 1. В работе "Relative, 2018" предложен метод работы с позициями, основанный на относительном кодировании, где используются пары позиций на разных расстояниях для подсчёта self-attention. 2. Модель T5, созданная в 2020 году, кодирует позиционную информацию с использованием скаляра, прибавляемого к self-attention перед применением softmax. 3. NoPE (No Positional Encoding), разработанная в 2023 году, представляет собой модель, которая обучается без явной позиционной информации, особенно эффективна для декодировщиков. 		
Позитивные следствия и/или достоинства описанной в статье технологии (минимум три пункта) <ol style="list-style-type: none"> 1. Возможность выделения абсолютной позиционной информации на первом слое модели декодировщика, что теоретически позволяет обходиться без явной позиционной информации в авторегрессионных моделях. 2. Расширение окна улавливаемой позиционной информации с каждым последующим слоем модели, что способствует обобщению на более длинные контексты. 3. Интерполяционные методы кодирования позиционной информации (например, RoPE) выглядят изящно и перспективно, обеспечивая успешное обучение моделей. 		
Негативные следствия и/или недостатки описанной в статье технологии (минимум три пункта) <ol style="list-style-type: none"> 1. NoPE, несмотря на свою эффективность, может столкнуться с проблемами при росте длины контекста, что ограничивает её применимость. 2. Отсутствие единого стандарта в области кодирования позиционной информации, что может затруднить выбор подхода при обучении модели. 3. В статье не упоминаются конкретные результаты экспериментов, оценивающих эффективность различных методов кодирования позиций на разных задачах, что может затруднить сравнение и выбор оптимального метода в конкретном контексте. 		
Ваши замечания, пожелания преподавателю или анекдот о программистах¹ Собрался как-то один студент сдать первую аннотацию в срок, но не вышло, так его анекдот в группу «Студенты Балакшина П.В.» выложили. Не верите – попробуйте снова выложить в группу, затем проверьте снова, пост должен появиться =)		

¹ Наличие этой графы не влияет на оценку