Домашняя работа по лекции 3

ФКН НИУ ВШЭ

25 ноября 2024 г.

1 Задание

Рассмотрим механизм самовнимания в трансформерной модели.

1. Пусть заданы матрицы запросов $Q \in \mathbb{R}^{T \times d_k}$, ключей $K \in \mathbb{R}^{T \times d_k}$ и значений $V \in \mathbb{R}^{T \times d_v}$. Выведите выражение для матрицы выхода внимания $Z \in \mathbb{R}^{T \times d_v}$, используя механизм самовнимания:

$$Z = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V.$$

   Объясните пошагово, как вычисляются веса внимания и как масштабирующий фактор $\frac{1}{\sqrt{d_k}}$ влияет на стабильность градиентов при обучении.

2. Опишите, как позиционные кодировки добавляются к входным эмбеддингам и почему они необходимы в трансформерных моделях, которые не используют рекуррентность.

3. Для заданной последовательности длины T сравните вычислительную сложность самовнимания в трансформерах с рекуррентными нейронными сетями (RNN) и сетью долговременной краткосрочной памяти (LSTM). Покажите, как масштабируется время вычислений и использование памяти в зависимости от T.

4. Предложите модификацию стандартного механизма самовнимания для снижения вычислительной сложности при обработке очень длинных последовательностей. Опишите математически предложенный подход и обсудите его потенциальные преимущества и недостатки.

Баллы
1 задание - 2 балла
2 задание - 3 балла
3 задание - 2 балла
4 задание - 3 балла