

Большие языковые модели

Корнет Мария Евгеньевна
старший преподаватель кафедры
Инженерная кибернетика



Pre-training

Это начальная фаза, когда модель обучается на огромном количестве текстовых данных (например, из интернета) для изучения общего языка и знаний.

Самообучение (Self-supervised learning): Основной метод предобучения LLM.
Модель обучается предсказывать часть текста по другой части.

- **Авторегрессионное обучение** (как в GPT): модель учится предсказывать следующее слово (токен) на основе предыдущих.
- **Маскированное языковое моделирование** (как в BERT): модель учится предсказывать маскированные (скрытые) слова в предложении.

Задачи Pre-training

Masked Language Modeling (MLM): В этом подходе (используемом, например, в BERT) часть слов в предложении скрывается, и модель должна восстановить их на основе контекста.

Causal Language Modeling (CLM): Этот метод (используемый в GPT) обучает модель предсказывать следующий токен на основе предыдущих.

Next Sentence Prediction (NSP): В некоторых моделях (например, BERT) используется задача предсказания, идёт ли одно предложение после другого.

Fine-tuning

Полная тонкая настройка (Full Fine-tuning): Обучение всех (или большинства) параметров модели на целевом датасете.

Недостатки: требует больших ресурсов и может приводить к катастрофическому забыванию.

Эффективная настройка с сохранением параметров (Parameter-Efficient Fine-tuning, PEFT): Методы, которые обновляют только небольшую часть параметров модели, что значительно снижает вычислительные затраты и требования к памяти.

Примеры:

LoRA (Low-Rank Adaptation): как мы уже обсуждали, добавляет обучаемые низкоранговые матрицы к некоторым слоям.

Адаптеры (Adapters): Вставляют небольшие обучаемые модули между слоями трансформера, замораживая основную модель.

Настройка префикса (Prefix Tuning) / Prompt Tuning: Добавляет обучаемые векторы (префиксы) к входу модели, в то время как сама модель заморожена.

Instruction tuning

Supervised Fine-Tuning (SFT): обучение на датасетах вида (инструкция – правильный ответ).

Датасет запросов и эталонных ответов на них: $\{(\text{ans}_i, \text{prompt}_i)\}_{i=1}^l$

$$\sum_i \log p_\theta(\text{ans}_i \mid \text{prompt}_i) \rightarrow \max_\theta$$

Reinforcement learning (RL)

1. Обучение с подкреплением на основе человеческих предпочтений (RLHF):

Состоит из нескольких шагов:

↑
• **Сбор данных предпочтений:** Люди оценивают несколько ответов модели, выбирая лучший.

• **Обучение модели вознаграждения (Reward Model):** Обучается на собранных данных предпочтений, чтобы предсказывать, какой ответ понравится человеку.

• **Настройка с помощью RL:** Исходная языковая модель оптимизируется для генерации ответов, которые получают высокие оценки от модели вознаграждения, при этом не слишком отклоняясь от исходной модели (чтобы сохранить качество и избежать "галлюцинаций").

2. Прямая оптимизация предпочтений (Direct Preference Optimization, DPO):

Он напрямую оптимизирует модель на данных предпочтений, используя теоретический вывод, что модель вознаграждения может быть выражена через оптимальную политику.

Reinforcement learning (RL)

1. Обучение с подкреплением на основе человеческих предпочтений (RLHF):

Состоит из нескольких шагов:

↑
• **Сбор данных предпочтений:** Люди оценивают несколько ответов модели, выбирая лучший.

• **Обучение модели вознаграждения (Reward Model):** Обучается на собранных данных предпочтений, чтобы предсказывать, какой ответ понравится человеку.

• **Настройка с помощью RL:** Исходная языковая модель оптимизируется для генерации ответов, которые получают высокие оценки от модели вознаграждения, при этом не слишком отклоняясь от исходной модели (чтобы сохранить качество и избежать "галлюцинаций").

2. Прямая оптимизация предпочтений (Direct Preference Optimization, DPO):

Он напрямую оптимизирует модель на данных предпочтений, используя теоретический вывод, что модель вознаграждения может быть выражена через оптимальную политику.

Другие подходы

- 1. Смесь экспертов (Mixture of Experts, MoE)**
- 2. Контекстное обучение (In-context Learning)**
- 3. Цепочки размышлений (Chain-of-Thought, CoT)**
- 4. Мультимодальное обучение.**



Transformer

- **Data:** WMT 2014 English-German (4.5M пар предложений) + WMT 2014 English-French (36M пар предложений)
- **Batch Size:** 25,000 tokens \approx 64-128 sequences \times \sim 200-400 length
- **Training Time:**
 - English-German: 100,000 steps (\sim 12 часов)
 - English-French: 300,000 steps (\sim 3.5 дня)
- **Optimizer:** Adam, $\beta_1=0.9$, $\beta_2=0.98$, $\epsilon=10^{-9}$
- **Warmup Steps:** 4,000 (линейный рост LR, затем обратный корень decay)
- **Dropout:**
 - Residual dropout = 0.1
 - Attention dropout = 0.1
 - Embedding dropout = 0.1 (только в большей модели)
- **Label Smoothing:** $\varepsilon_{ls} = 0.1$

BERT представление входа

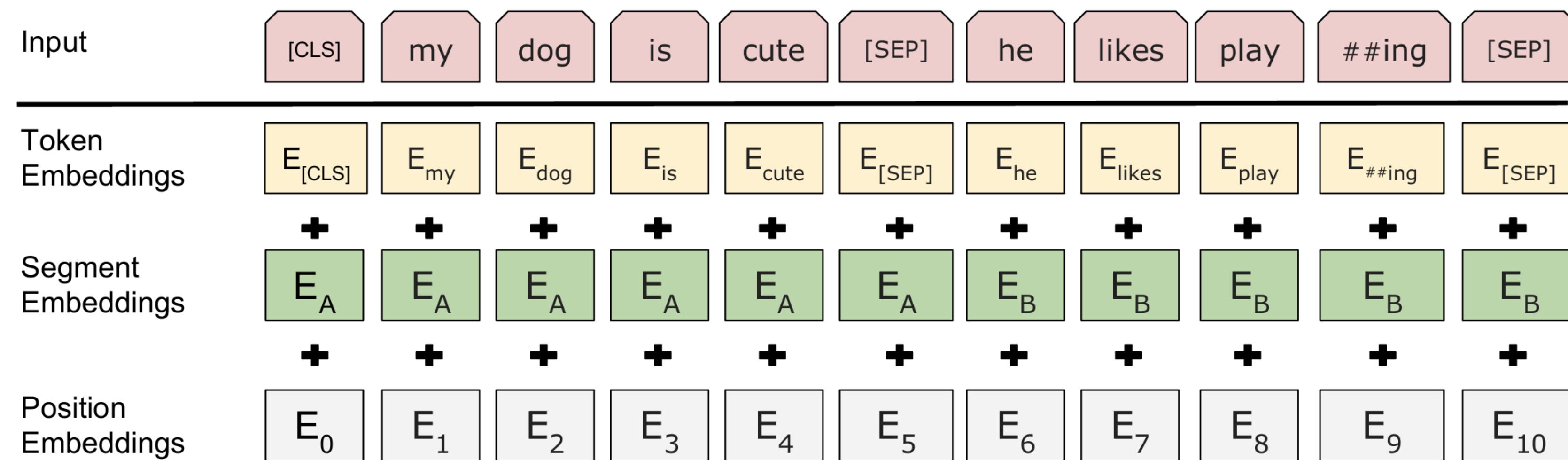


Figure 2: BERT input representation. The input embeddings are the sum of the token embeddings, the segmentation embeddings and the position embeddings.

BERT

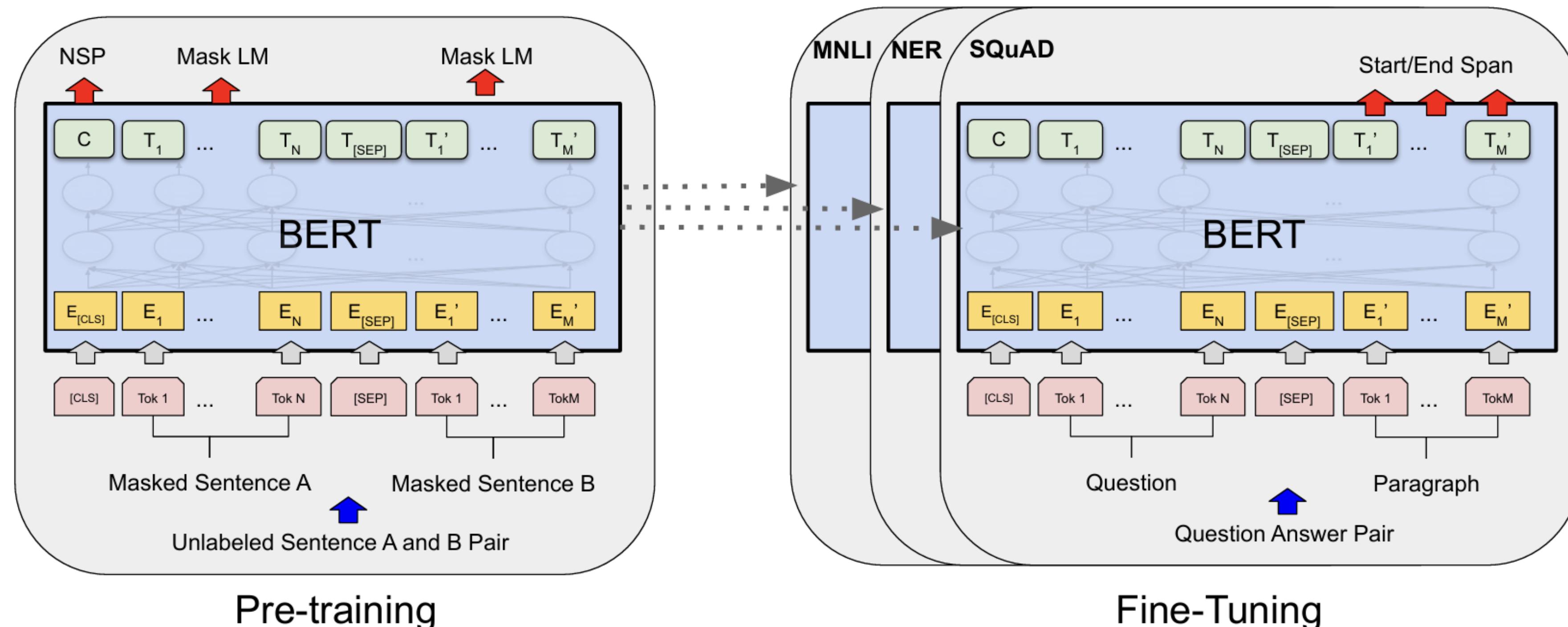


Figure 1: Overall pre-training and fine-tuning procedures for BERT. Apart from output layers, the same architectures are used in both pre-training and fine-tuning. The same pre-trained model parameters are used to initialize models for different down-stream tasks. During fine-tuning, all parameters are fine-tuned. [CLS] is a special symbol added in front of every input example, and [SEP] is a special separator token (e.g. separating questions/answers).

Задачи

Название задачи

MNLI

Multi-Genre Natural Language Inference

QQP

Quora Question Pairs

QNLI

Question NLI

STS-B

Semantic Textual Similarity Benchmark

MRPC

Microsoft Research Paraphrase Corpus

RTE

Recognizing Textual Entailment

SWAG

Situations With Adversarial Generations

Тип задачи

Natural Language Inference
(NLI)

Дублирование вопросов

NLI на вопрос+текст

Регрессия

Классификация перефраз

Natural Language Inference
(NLI)

Завершение ситуации
(multiple choice)

Отношение между посылкой и гипотезой

Являются ли два вопроса дубликатами

Содержит ли текст ответ на вопрос

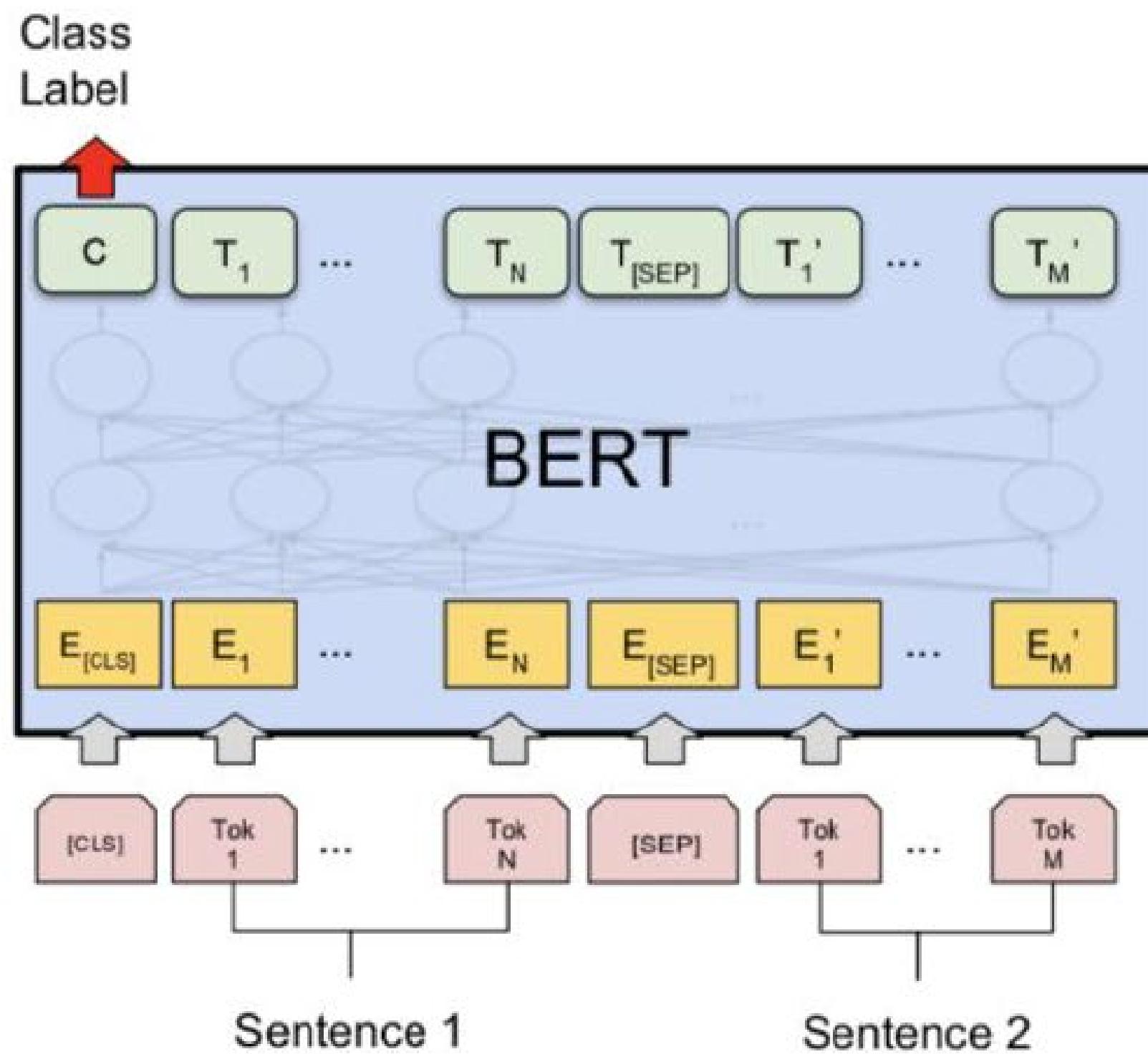
Оценить степень семантической близости (0–5)

Перефраз ли одно предложение другого

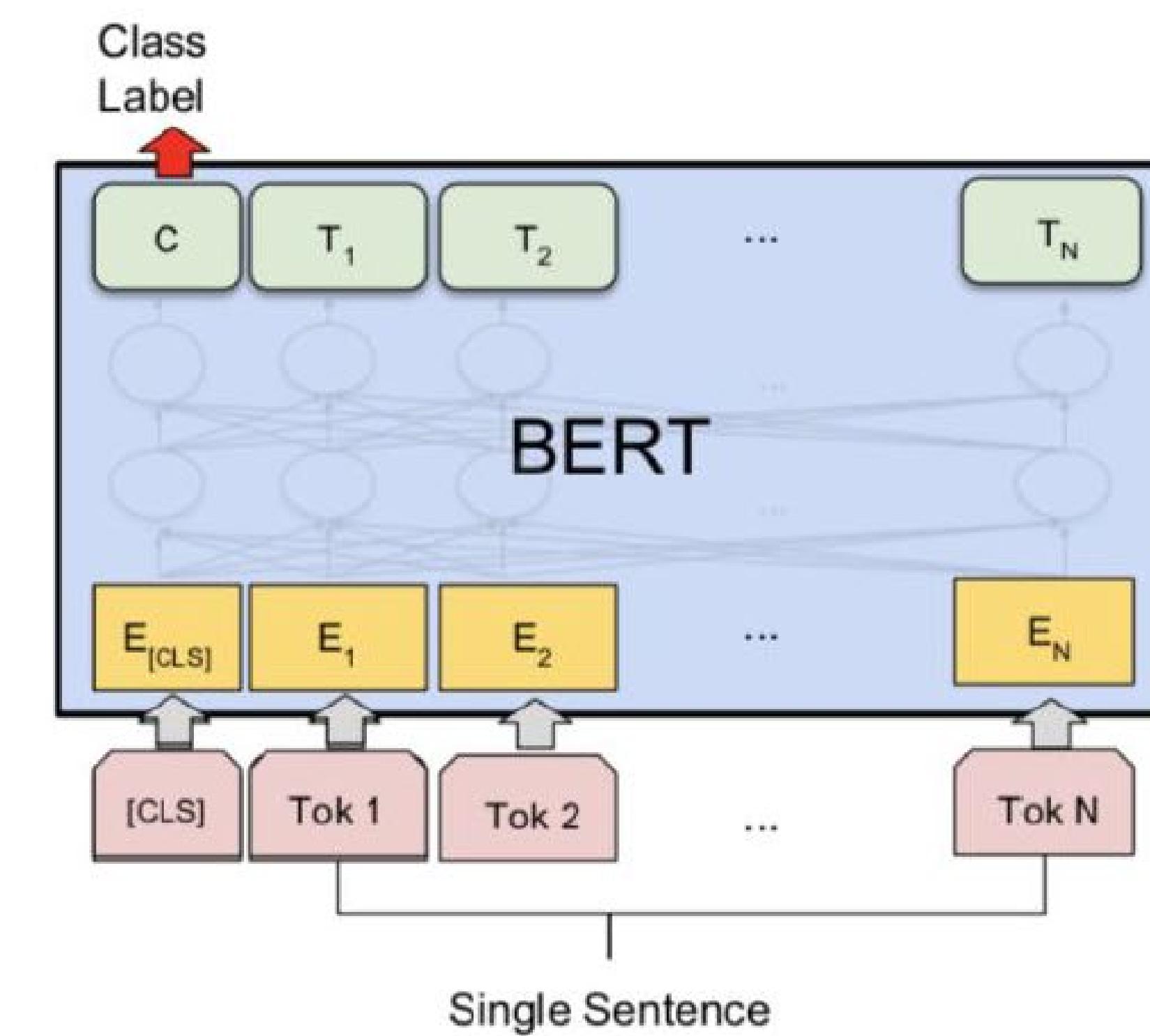
Следует ли гипотеза из посылки

Выбрать логичное продолжение из 4-х вариантов

BERT

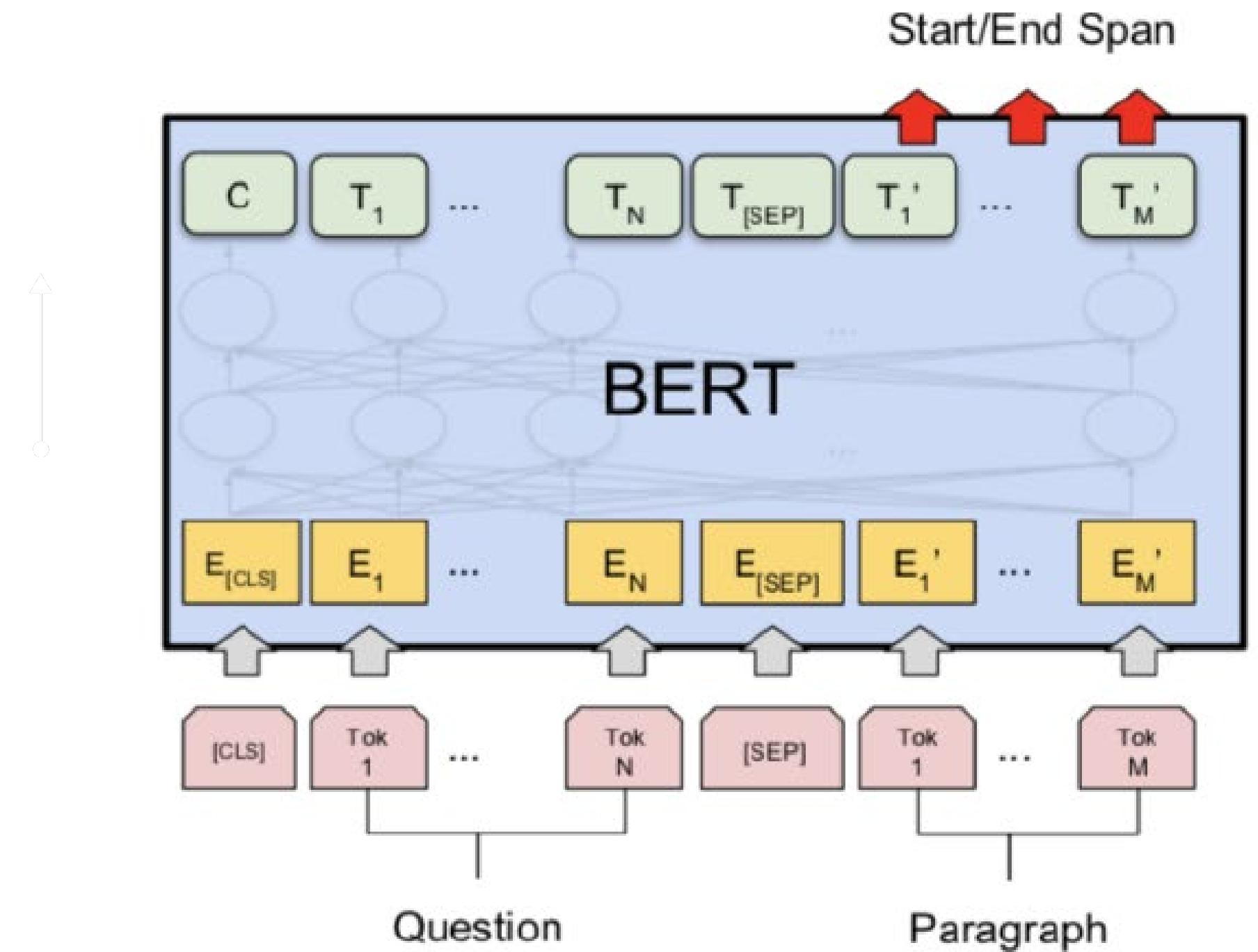


(a) Sentence Pair Classification Tasks:
 MNLI, QQP, QNLI, STS-B, MRPC,
 RTE, SWAG

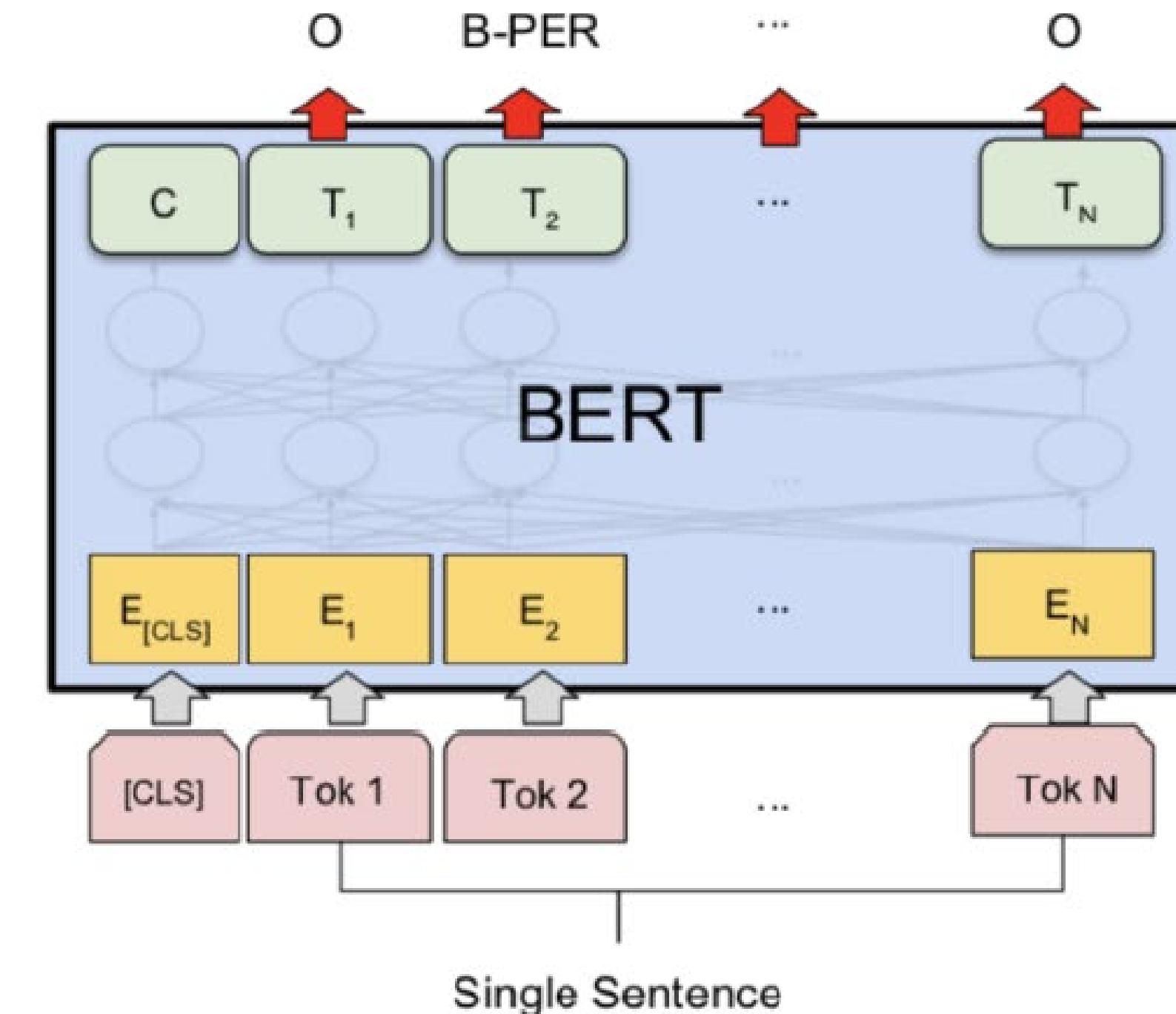


(b) Single Sentence Classification Tasks:
 SST-2, CoLA

BERT



(c) Question Answering Tasks:
SQuAD v1.1



(d) Single Sentence Tagging Tasks:
CoNLL-2003 NER



**Спасибо
за внимание!**

