

# UniLM

参考: <https://zhuanlan.zhihu.com/p/113380840>

中文UniLM语言模型参考: <https://github.com/YunwenTechnology/Unilm>

## 模型原理图

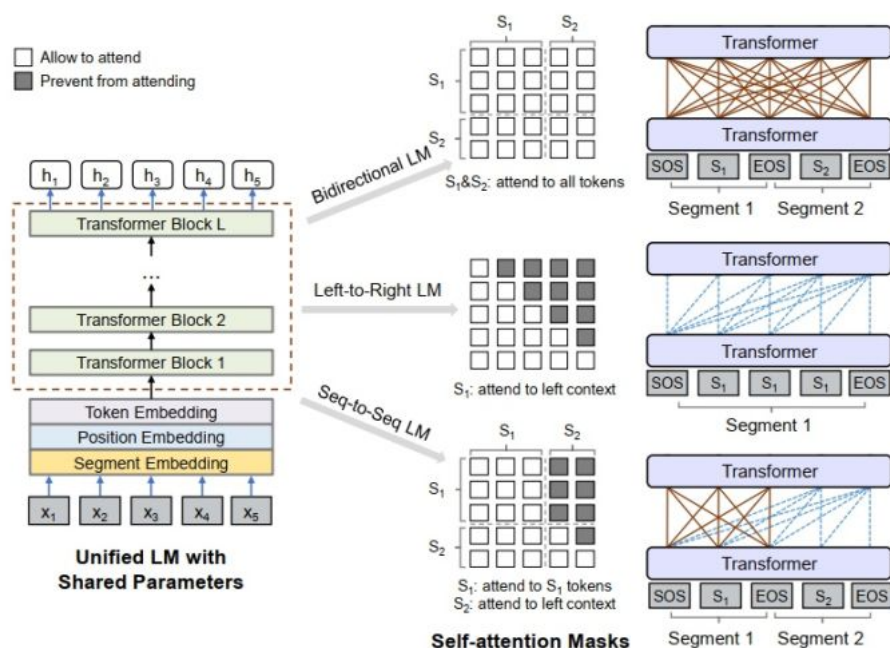


Figure 1: Overview of unified LM pre-training. The model parameters are shared across the LM objectives (i.e., bidirectional LM, unidirectional LM, and sequence-to-sequence LM). We use different self-attention masks to control the access to context for each word token. The right-to-left LM is similar to the left-to-right one, which is omitted in the figure for brevity. 知乎 @刘聪NLP

## 模型创新点

让预训练语言模型既可以应用于nlu任务, 也可以应用于nlg任务

## 模型思想

1. 句子对喂给模型, 如 [SOS]  $S_1$  [EOS]  $S_2$  [EOS], 其中 $s_2$ 是 $s_1$ 的答案
  2. MARK大法: 通过不同的掩码方式实现三种不同目标函数的语言模型
  3. 双向语言模型, 如图中中间列第一个图, 预测token时可看见 $s_1, s_2$ 中所有token;
  4. 单向语言模型, 如图中中间列第二个图, left->right, 预测当前token时只能看到它之前的token;
  5. seq-to-seq, 如图中中间列第三个图, 预测token出现在 $s_1$ 中, 只看见 $s_1$ 中所有的token; 如果预测token出现在 $s_2$ 中, 预测当前token, 只能看到 $s_1$ 中全部token和 $s_2$ 中预测token之前的token.
  6. NSP任务, 和bert一样
- 3、模型结构和bert-large一致, 由训练好的bert-large模型初始化