



# 第8章：预训练语言模型

课件制作：曲彦儒、吴宪泽、屠可伟

讲解人：曲彦儒、吴宪泽

# 预训练语言模型

---

- 预训练语言模型在易于获取的海量无标注语料数据上训练，不针对任何特定下游任务，而是为了获取通用的语言学知识，因此称为预训练（pretraining）
- 预训练语言模型可以分为以下三类：
  - 仅包含编码器，即自编码（autoencoding）模型
  - 仅包含解码器，即自回归（auto-regressive）模型
  - 编码器和解码器均包含，即序列到序列（seq2seq）模型

## 8.1 ELMo: 基于语言模型的上下文相关词嵌入

---

### □ 来自多层双向语言模型的词表示

- 训练时，分别训练前向和后向的多层语言模型
- 在下游任务上使用时，将所有层的前后向LSTM的输出共同作为特征

### □ 上下文相关词嵌入

- 使用整个句子作为上下文

## 语言模型架构与训练

- 设计目标：有效且高效的模型
- 两层双向LSTM
  - 每层LSTM的输入维度为512，隐状态维度为4096
- 使用残差结构
- 使用字符粒度的CNN获取词嵌入
- 分别训练前向和后向的语言模型
- 前后向语言模型共享词嵌入，且将词嵌入作为输出投影层的参数

## 微调

□ 组合每层模型输出的词表示，组合方式与下游任务相关

- 不同于仅使用最上层LSTM的隐状态的传统做法

$$\text{ELMo}_k^{\text{task}} = \gamma^{\text{task}} \sum_{j=0}^L s_j^{\text{task}} h_{k,j}^{\text{LM}}$$

$$h_{k,0}^{\text{LM}} = x_k^{\text{LM}} \quad h_{k,j}^{\text{LM}} = [\vec{h}_{k,j}^{\text{LM}}, \overleftarrow{h}_{k,j}^{\text{LM}}] \quad (j = 1, \dots, L)$$

- $\gamma^{\text{task}}$  : 控制ELMo词表示在下游任务中的作用
- $s^{\text{task}}$  : 控制每层模型的字表示在ELMo词表示中的占比

## 8.2 BERT: 基于Transformer的双向编码器

### BERT

- ELMo的缺点是没有将整个上下文进行统一表示，BERT模型通过将预训练任务从语言模型替换为掩码语言模型（masked language model, MLM）解决了这一问题
- 掩码语言模型：给定输入文本，将一定比例的词进行掩码（替换为特殊符号“[MASK]”），然后让模型预测这些被掩码的词
- 下句预测（next sentence prediction）：即输入两个句子，预测句2是否是句1的下一句话

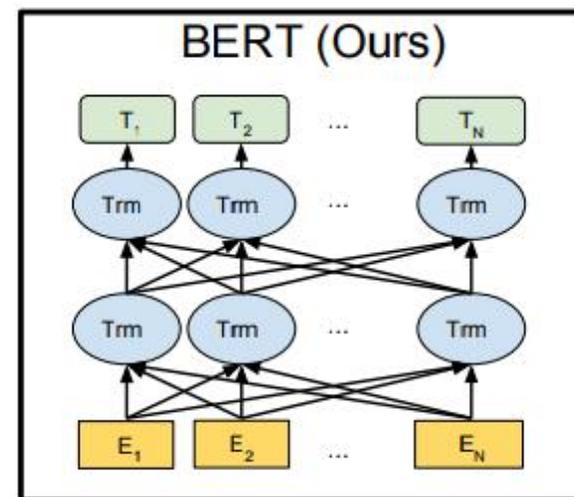
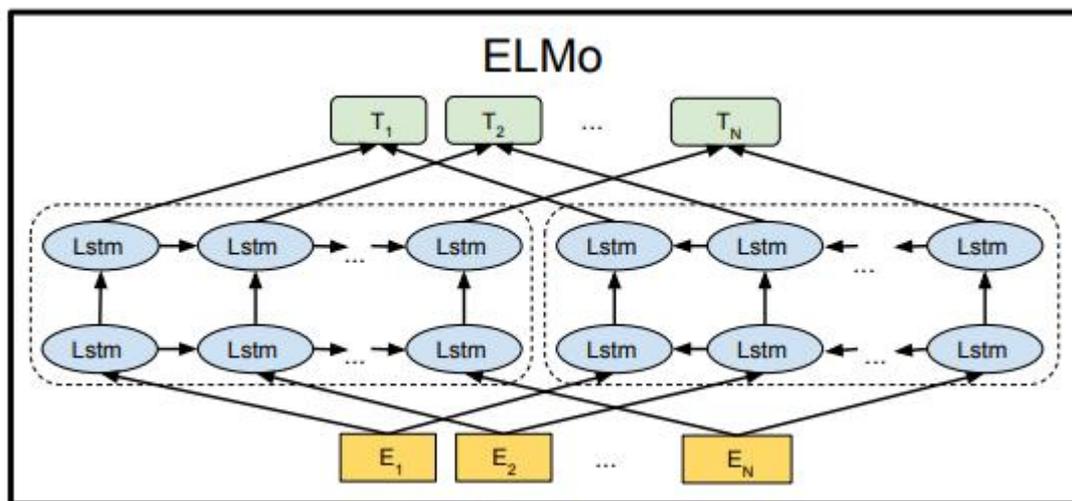
# BERT

---

## 微调与提示

- BERT的微调方式与ELMo一样，即将BERT模型与下游任务模型连接起来，一起在该任务上进行训练，BERT微调时一般使用较低的学习率
- 提示学习（prompt learning）：通过额外输入一段提示文本，将目标任务转化为掩码语言模型的形式，利用BERT直接预测掩码位置的结果

# BERT: 与ELMo的比较



- ❑ 条件语言模型，单向
- ❑ 拼接两个单向模型的表示以获取双向信息
- ❑ LSTM

- ❑ 遮挡语言模型，双向
- ❑ 单模型表示已包含双向信息
- ❑ Transformer

## 8.3 GPT: 基于Transformer的生成式预训练

---

- GPT (generative pre-trained transformer) 是由OpenAI提出的基于Transformer的生成式预训练语言模型
- 与BERT不同, GPT是一个严格意义上的语言模型, 使用注意力掩码
- GPT模型的演变:
  - GPT-1到GPT-3: GPT-1基于经典的预训练微调范式, GPT-2相比于GPT-1, GPT-3相比于GPT-2使用了更多的模型参数和训练数据。随着模型规模的增大, GPT-3涌现出了上下文内学习 (in-context learning) 的能力
  - InstructGPT: 在GPT-3的基础上, InstructGPT提出了在训练过程中引入人类反馈来让模型更好地理解用户意图, 称为基于人类反馈的强化学习 (reinforcement learning with human feedback, RLHF)

## □ GPT模型的演变：

- ChatGPT：2022年底，OpenAI发布了ChatGPT，在全球引起了轰动。ChatGPT拥有惊人的自然语言理解和对话能力，能很好地遵从用户指令，执行包括文本生成和改写，各类问答甚至代码生成和修复等各种任务
- GPT-4：GPT-4是一个大规模的多模态模型，可以接受图像和文本输入。GPT-4在各种专业和学术测试中表现出人类水平的性能

# 总结

---

## □ 预训练语言模型

- 自编码模型
- 自回归模型
- 序列到序列模型