

预训练语言模型

汇报人：李子明

武汉大学金融系

2025.12.28

主流预训练语言模型范式

- Encoder-only: 判别式模型
 - 双向上下文建模: 同时捕捉词语在前后语境中的语义依赖关系
 - 掩码预测判别: 能够对文本语义结构的精细刻画
 - 语义表征: 适合文本分类、命名实体识别、语义匹配及信息抽取等理解型任务
 - 不包含自回归机制, 文本生成与长序列输出受限, 需要额外结构或与生成模型结合
- Decoder-only: 生成式模型
 - 单向左侧建模: 给定历史序列的条件下预测下一个词元
 - 自回归生成: 实现对语言生成过程的显式刻画
 - 生成推理: 以输入文本作为条件提示进行续写和推断, 适合 prompt 形式的任务描述
 - 精确语义对齐或判别式输出中, 需要通过提示工程或外部评估机制加以约束。
- Diffusion: 扩散模型
 - 全局分布建模: 不直接建模序列的逐步预测关系, 通过对整体数据分布的逐层逼近
 - 噪声建模与逐步去噪: 正向噪声注入与反向去噪生成
 - 复杂数据: 样本多样性与生成稳定性, 训练和推理计算开销大, 生成效率较低

BERT 在文本嵌入的优势

- 经典嵌入方法局限性
 - 传统嵌入方法
 - Word2Vec (Mikolov 等, 2015)
 - GloVe (Pennington, Socher 和 Manning, 2014)
 - FastText (Bojanowski 等, 2017; Joulin 等, 2016)
 - 依赖于分布假设, 忽略了词语的多义性和上下文依赖性。
 - 从零开始构建词嵌入, 无法利用大规模语料中的丰富语义信息。
 - 在词级嵌入方面表现良好, 但作为句子编码器效果不佳, 常常误解语境。
- 基于深度神经网络的自然语言处理方法 BERT (Devlin 等, 2018)
 - 神经网络将词语嵌入向量空间可提升 NLP 性能 (Mikolov 等, 2013)
 - BERT 比以往的方法更能够深入揭示文化关联和类别 (Kozlowski 等, 2019)