# 使用预训练模型Ernie3.0进行3场景10情绪分类

## 任务介绍

为鉴别抑郁症患者的情绪及其来源或针对的范围，我们借助大语言模型对文本进行分类，分类为3种场景（'家庭','学校','社会环境'）和10种情绪（'恐惧','愤怒','厌恶','悲伤','羞愧','内疚','嫉妒','羡慕','爱'',快乐'）。

在此次任务中遇到的最大问题是数据集的问题，由于缺乏现有数据，我们使用现有的大语言模型生成数据以构建数据集。模型在数据集的表现较优，但验证集缺乏人类直接产生的数据，这仍然是一个相当大的空缺。

## 为什么选用这个模型？

据说性能优异。

（摘自论文《ERNIE 3.0: Large-scale Knowledge Enhanced Pre-training for Language Understanding and Generation》）

【摘要】：经过预训练的模型在各种自然语言处理任务中取得了最先进的结果。最近的工作，如 T5和 GPT-3已经表明，扩大预先训练的语言模型可以提高他们的泛化能力。特别是，具有1750亿参数的gpt-3模型显示了其强大的任务在无样本/少样本的学习能力。尽管它们取得了成功，这些大规模的模型是在纯文本上训练的，没有引入知识，如语言知识和世界知识。此外，大多数大模型都是以自回归的方式进行训练。因此，这种传统的微调方法在解决下游语言理解任务时表现出相对较弱的性能。为了解决上述问题，我们提出了一个统一的框架 ERNIE 3.0，用于预先训练大规模知识增强模型。它融合了自回归网络和自编码网络，使得训练后的模型能够很容易地适应自然语言理解和文本生成任务在0样本，少样本或者微调。在大规模知识图谱组成的4TB语料库上，用100亿个参数对模型进行了训练。实证结果表明， 该模型在 54 个中文自然语言处理任务中的表现优于国内最先进的模型，其英文版在 SuperGLUE 基准(2021年7月3日)中排名第一，超过人类表现 + 0.8% (90.6% 对 89.8%)。

论文链接：[2107.02137.pdf (arxiv.org)](https://arxiv.org/pdf/2107.02137.pdf)

## 模型型号选择



即6个transformer encoder层，每层中间输出768维特征，以及包含多头注意力机制层。

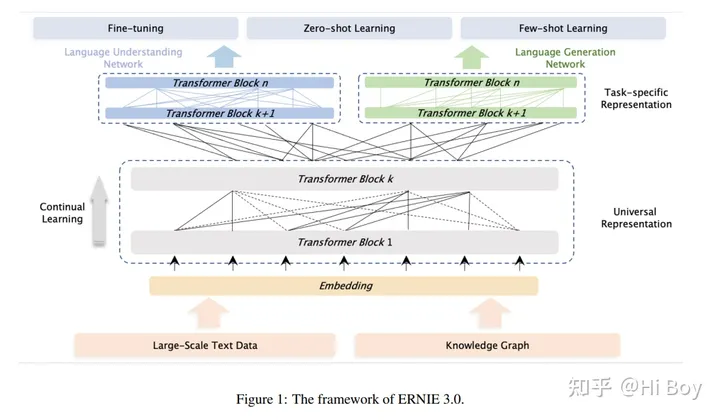
（注意力头是用来计算不同词之间的关系的，采用12个头相当于算12次然后再线性组合成一个值作为最终值）

**参考项目代码：**

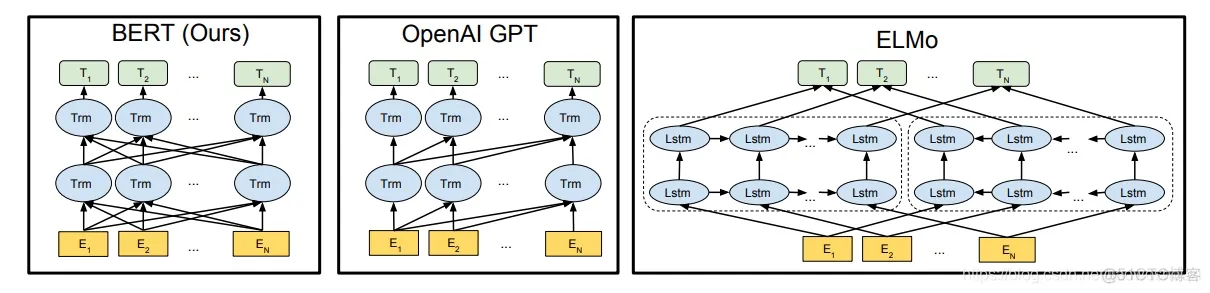
<https://github.com/PaddlePaddle/PaddleNLP/tree/develop/applications/text_classification/multi_class>

## Ernie3.0模型简介

Ernie3.0模型，结合了BERT和GPT，既有文本理解的功能，又有文本生成的功能。在本项目中，仅用到了文本理解（文本分类）的部分，即下文中属于左边淡蓝色的进行微调的部分（右边的淡绿色部分是用于文本生成的）。



由下图和上图对照可知，Ernie3.0模型是由embedding层，全局双向transformer层，加上后续的局部双向transformer层和单向transformer层构成的。（做个比喻，就是在BERT的树干上左边嫁接BERT的枝条，右边嫁接GPT的枝条。）



可见ernie3.0对于本项目的任务而言相当于一个BERT。

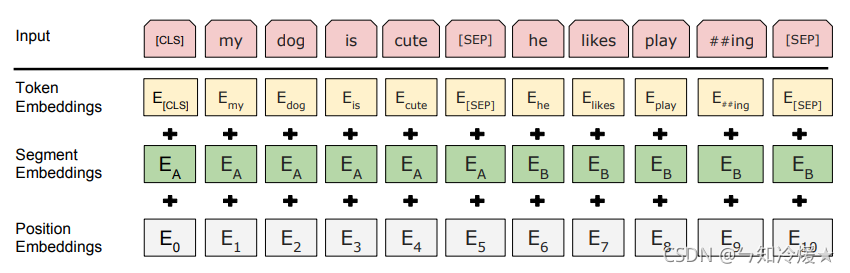
以下简要介绍BERT原理：

Bert是基于Transformer的模型，并且是一个迁移能力很强的通用语义表示模型，但是Bert只是运用了Transformer的Encoder部分。Bert的全称是Bidirectional Encoder Representation from Transformers，即双向Transformer的Encoder。（表现在上图中，就是箭头可以指向左边。GPT只指向右边，这样信息只能往句子的后边传递，因而只能是生成式的）

BERT层次结构：

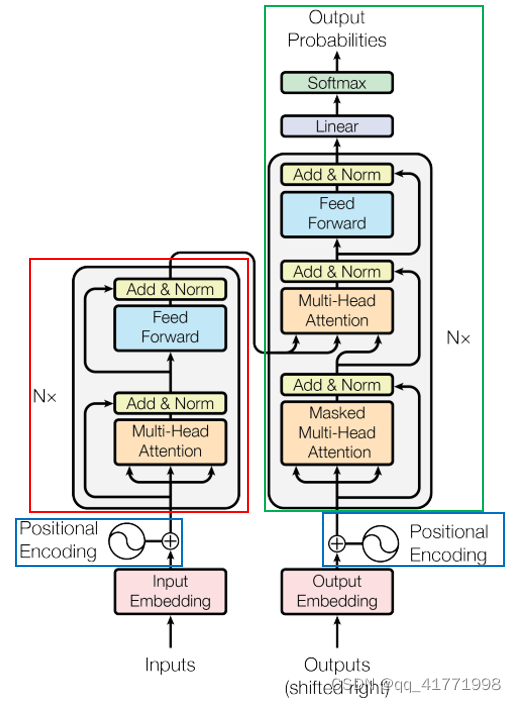
负一层：分词，将字符串分割为token串。

最底层：embedding，将输入的token串变成向量，在BERT中，最终的向量是由词的本身的词向量（淡黄色层）、所在句子的向量（淡绿色层）和位置向量（白色层）加和成的。

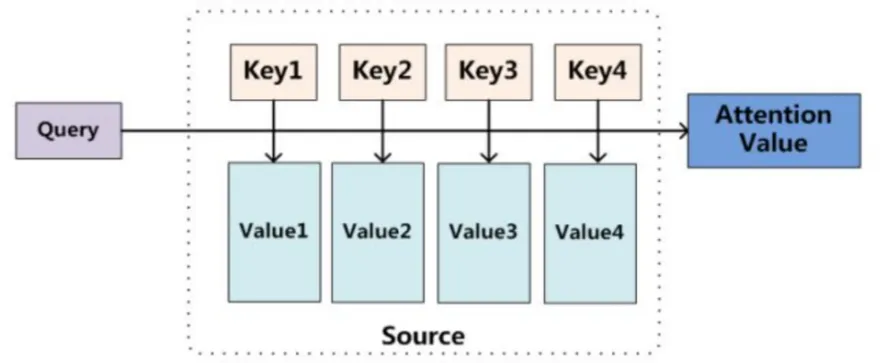


中间层：双向transformer encoder的深度网络，transformer encoder采用多头注意力机制（Multi-Head Attention）和前馈神经网络(Feed Forward)，前馈神经网络即相当于CNN中的全连接层。

输出层：最后一层transformer encoder的输出值作为输出



注意力机制分为查询（query）、键（key）、值（value）三部分。结果为query经过key和value的一系列运算得到。自注意力机制即查询、键、值都来自于同一个输入x（经过一些线性变换）。多头注意力机制只不过是多个注意力结果进行线性组合。



## 训练过程

BERT训练过程：

包括预训练和微调（或者说叫应用到下游任务）两个部分。

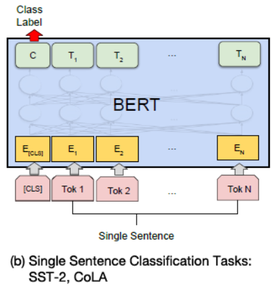
预训练：

任务1：Bert的第一个预训练任务是Masked LM，即在句子中随机遮盖一部分单词，然后同时利用上下文的信息预测遮盖的单词，这样可以更好地根据全文理解单词的意思。

任务2：BERT的第二个预训练任务是Next Sentence Prediction（NSP），下一句预测任务，这个任务主要是让模型能够更好地理解句子间的关系。

微调：

下游任务，比如说本项目的文本分类，一般使用C作为判断依据进行分类。



微调时一般只训练（非常小）部分的参数。有的时候只训练最后新加的全连接层的参数，有的时候是冻结模型的很多层的参数进行训练，因而微调的训练代价可以大大减小。

**ERNIE3.0训练过程：**

ERNIE3.0模型进行预训练与BERT略有不同，它采用了三种方式，第一种是单词预测（即BERT的训练任务1），第二种是单词重排任务，第三种是知识感知预训练任务。

ERNIE3.0的具体微调方法网上没找到现成资料，猜想大概是同BERT的方法。

## 训练数据

为Chat-GPT、文心一言、星火大模型、智谱AI、紫东太初这些模型生成的数据，按条目数分别占总数的10%、40.7%、0.9%、17%、31.3%。

数据大小：69.82MB

数据数量：11.8万

数据平均长度：204个字

例如：

在家庭场景下的恐惧情绪：

假设你是一个高中生，并有抑郁问题，这个问题是因父母教养方式粗暴产生的导致的,你感到恐惧，请你以这个学生的口吻描述一下你遇到的问题。如果可以，编个小故事,举一些例子，事情，场景，别人说的一些话和观点，体现这个学生的情绪。另外，请你口语化、通俗化、情绪化、更随便一些，以更有真实感。请你简洁一点，大概在100个字以内。

结果大致如下：

你知道吗？我感觉好难过。这都是因为我父母的教养方式太粗暴了！他们总是大喊大叫，从不理解我的感受。举个例子吧，有一次我考试成绩不好，我很难过，就想跟他们聊聊，结果他们却说：“你怎么这么没用！”我的心情更糟糕了。就是这样啊，我感觉自己像个没有价值的失败者。我真希望他们能多关心我一点，理解我的心情，而不是总是责骂我

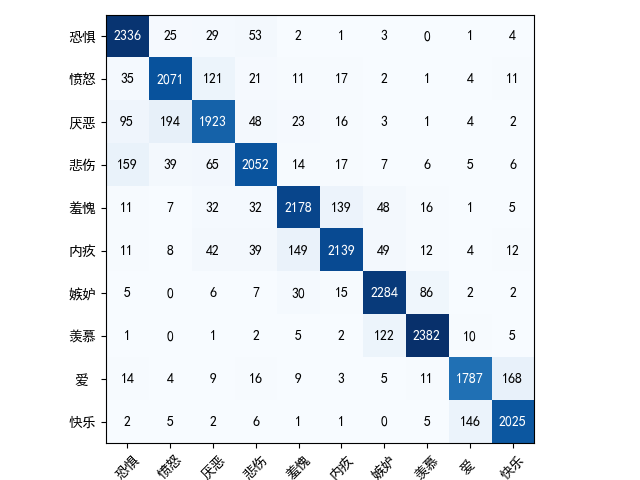
## 训练结果

情绪分类：

测试集总个数:23547

正确率：89.94%

混淆矩阵如下图所示：



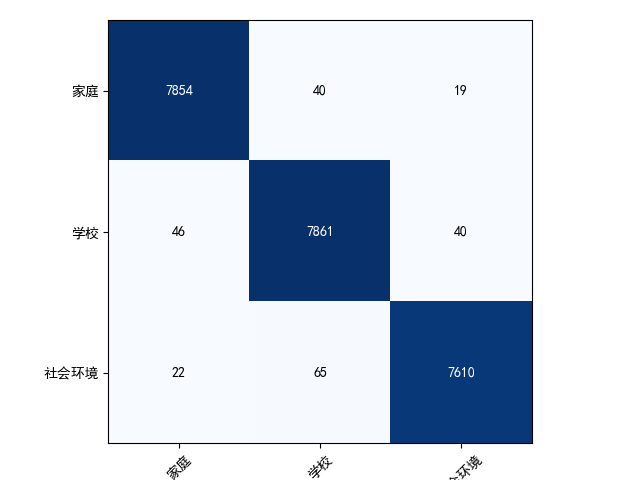
可以观察到一些情绪之间有轻微的粘连现象。

场景分类：

测试集总个数:23547

正确率：99.02%

混淆矩阵如下图所示：



## 总结

单看在样本集上的表现而言，模型的表现是比较优秀的。但是由于缺乏真正可靠的测试集，模型在真实场景下的表现仍不可保证。

有人建议：把真实的数据（不一定适用，不一定相关）通过大模型加工（比如加上‘愤怒’，‘社会因素’等提示），可能可以缓解训练数据和真实数据的异质化问题。