

東 北 大 學

语言分析与机器翻译作业

学 号：	2101790
姓 名：	孙庆轩
导 师：	朱靖波
学科类别：	<input checked="" type="checkbox"/> 工学 <input type="checkbox"/> 全日制专业学位
学科/工程领域：	计算机科学与技术
所属班级：	计硕 2104
所属学院：	计算机科学与工程学院
拟选题目：	基于 Transformer 模型实现，使用预处理好的 IWSLT14 De-En 数据集进行训练，输入德文输出对应的英文

# 1 背景

在机器翻译任务中，序列-序列模型被广泛应用，将源语言句子作为输入序列，将目标语言句子作为输出序列，便可以快速准确的得到翻译结果。这种模型使用编码器将输入序列转换为一个向量，使用解码器对向量进行解码，来得到另一个序列，巧妙的解决了输入与输出都为序列的问题。在自然语言处理的发展中，也产生了多个序列-序列的模型，比如循环神经网络，Transformer 等。

在自然语言处理任务中，编码的方向是沿着某一个方向的，这种处理问题的方式是循环神经网络中经常出现的，这类网络也被称之为自回归网络。

与之相对，Transformer 会将待处理的单词遮盖掉，使用其他的所有单词的信息来计算被遮盖的单词的信息。因此，Transformer 可以利用几乎全部的语句的单词信息来进行计算，并且不受时间的影响，因此在相当多的任务上都取得了瞩目的成绩。

# 2 任务要求

基于 Transformer 模型实现，使用预处理好的 IWSLT14 De-En 数据集进行训练，输入德文输出对应的英文。

# 3 模型架构

Transformer 的结构如图 1 所示。在原文中，其中编码器由 6 个完全相同的层组成。每一层可以拆分成两部分。第一部分被称为多头注意力机制。第二部分是一个按照位置的前馈神经网络。每一部分都使用了层归一化和残差连接。

在解码器部分，也使用了 6 个完全相同的层构成。每一层都是由三个部分构成，但是在第一层中，针对注意力机制模块做了掩盖，可以防止编码的时候访问到未来的目标单词，造成训练和解码的差异。在第二部分使用了多头注意力机制，作用类似于 RNN 的注意力机制，使得解码器能够根据上下文来进行解码。第三部分与编码器的第二部分相同，使用前馈神经网络来计算得到编码器的输出。

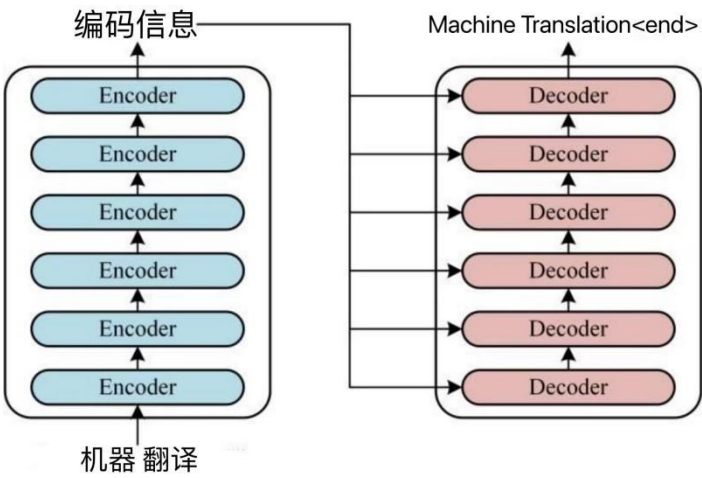


图 1 Transformer 架构

下面是 Transformer 中提出的几个重要组成部分。

1) 自注意力机制: 图 2 是 Self-Attention 的结构, 进行计算时, 需要用到矩阵 Q(查询),K(键值),V(值)。在机器翻译任务中, Self-Attention 接收的是单词的表示向量  $x$  组成的矩阵  $X$  或者是上一个 Encoder 的输出。Q,K,V 是通过 Self-Attention 的输入进行线性变换得到的。

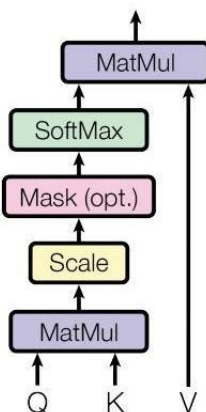


图 2 自注意力机制

2) 多头注意力机制: 图 3 是多头注意力机制的结构, 包含多个 Self-Attention 层, 在实际应用中, 为了增强模型的表达能力, 通常会将多个缩放的点积注意力机制进行组合, 这种组合后的注意力机制被称作多头注意力机制。首先将输入矩阵  $X$  分别传递到  $h$  个不同的 Self-Attention 中, 计算得到  $h$  个输出矩阵  $Z$ , 这样做的目的是为了使模型学习到不同方面的信息, 它将  $h$  个缩放的点积注意力机制的计算结果拼接在一起, 然后通过一个线性变换, 得到最终的输出向量。

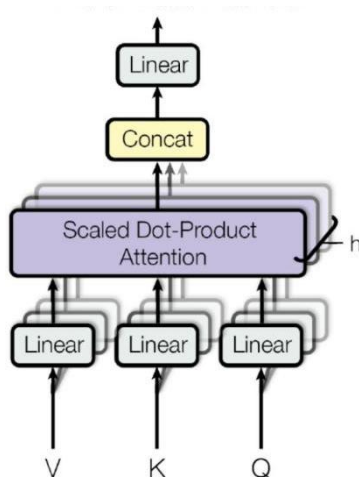


图 3 多头注意力机制

3) 编码: 单词的 Embedding 可以在 Transformer 中训练得到, 也可以采用 Word2Vec、Glove 等算法预训练得到。除此以外, Transformer 不同于 RNN, 不能利用单词的顺序信息, 训练中需要使用位置 Embedding 来表示单词出现在句子中的位置。这样才能使用全局信息, 通过位置 Embedding 保存单词在序列中的相对或绝对位置。

位置 Embedding 用 PE 表示，其中 PE 的维度与单词 Embedding 是一样的。PE 可以通过训练得到，也可以使用某种公式计算得到。在 Transformer 中采用了后者，计算公式如下：

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d})$$
$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d})$$

其中, pos 表示单词在句子中的位置, d 表示 PE 的维度 (与词 Embedding 一样), 2i 表示偶数的维度, 2i+1 表示奇数维度 (即  $2i \leq d, 2i+1 \leq d$ )。

这样做可以使 PE 能够适应比训练集里面所有句子更长的句子，可以让模型容易地计算出相对位置，因为  $\sin(A+B) = \sin(A)\cos(B) + \cos(A)\sin(B)$ ,  $\cos(A+B) = \cos(A)\cos(B) - \sin(A)\sin(B)$ 。所以对于固定长度的间距 k,  $PE(pos+k)$  可以用  $PE(pos)$  计算得到。这为模型捕捉单词之间的相对位置关系提供了非常大的便利。

Transformer 的提出抛弃了在 NLP 中经常使用的最根本的 RNN 或者 CNN，并且取得了非常不错的效果，通过位置编码，有效的解决了之前的长期依赖问题。同时 Transformer 可以根据硬件设备，实现良好的并行性。目前 Transformer 不仅仅应用在机器翻译领域等自然语言处理方向上的任务，甚至可以利用在计算机视觉等领域，可以说是火出圈来了。但是 Transformer 模型丧失了捕捉局部特征的能力，同时失去的位置信息是自然语言处理任务中非常重要的一环，这是一个词袋模型，Transformer 结构上的固有缺陷需要后期的工作来提升，这也为探索其发展潜力带来了机会。

## 4 实验

### 4.1 实验配置

本实验是在实验室远程服务器的条件下完成的，实验配置如下：

操作系统：CentOS 7.3

显卡：NVIDIA GeForce GTX TITAN X 12G

CPU：Intel(R) Core(TM) i7-5930K CPU @ 3.50GHz

### 4.2 数据集

本次实验的数据集是已经预处理过的 iwslt14 de-en，内容为图 4 所示。

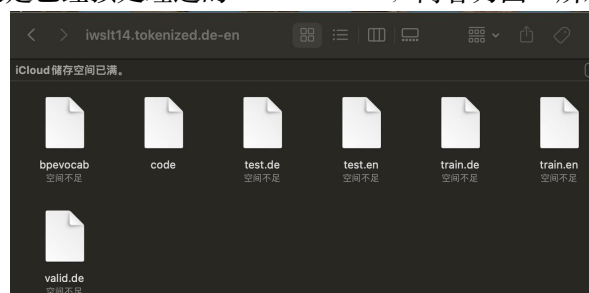


图 4 数据集格式

对数据集的处理要进行 BPE 子词划分，需要下载 subnmt，可以使用 pip 安装。

结合原语料进一步使用 apply bpe，结合词汇表，不使用低频词，以减少低词频影响。因为过滤掉低词频，可能会出现 oov 问题，如出现 oov 问题，则将原词切分为更小的词。

算法的执行步骤为：将原始语料每一行读入，以空格切分，形成每个词组。对词组中的每个词得到对应的 word pair。每次在这个 word pairs 选择在 bpe code 权重最大的 pair，组合并形成新的 word，如果最终 pair 数为 1 即没有合并的可能或者最大的 pair 也不再 bpe\_codes 中，则停止循环。为了便于区分，处理结束后在文件名称其中插入了 .bpe。。如下图 5 所示。

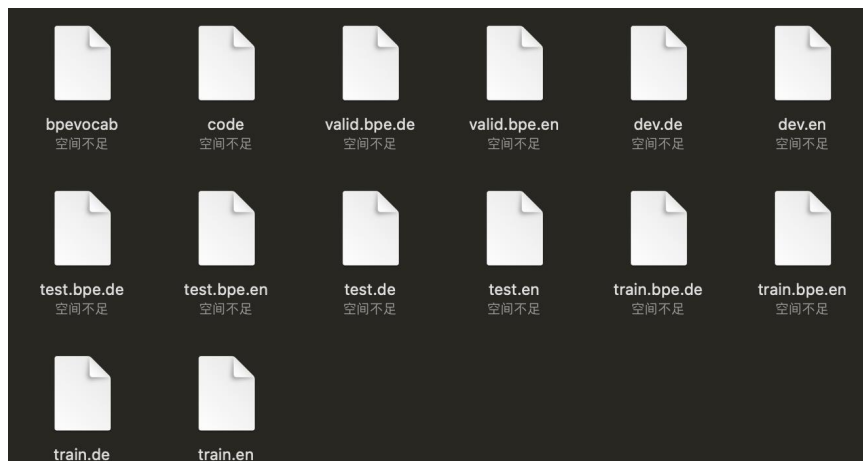


图 5 处理之后格式

## 4.3 训练

使用 3 张 NVIDIA GeForce GTX TITAN X 12G 显卡进行训练，训练 60 轮，warmup steps 为 10000，encoder 和 decoder 都为 6 层，betas 为 0.9 0.98。大约花费 4 小时。训练过程中，动态的计算 loss 和困惑度，在每一轮训练结束，将每一轮的 loss 和最小的困惑度打印出来，并保存每一轮的模型。同时在验证集上进行检验，如果这一轮的模型比之前训练的模型效果更佳，就将其指定为最佳模型。

训练过程如下图 6 所示：

```
Epoch 31 ::: Train
[step 1075/1075] loss: 2.505, nll loss: 1.515, ppl: 2.059, time: 244.2s

Epoch 31 ::: Validation
nll loss: 1.597, ppl: 3.025, best ppl: 3.025
[*] Model is saved in 'models/de-en/model31.pth'.
[*] Best model is changed!

Epoch 32 ::: Train
[step 1075/1075] loss: 2.487, nll loss: 1.495, ppl: 2.019, time: 244.3s

Epoch 32 ::: Validation
nll loss: 1.589, ppl: 3.009, best ppl: 3.009
[*] Model is saved in 'models/de-en/model32.pth'.
[*] Best model is changed!

Epoch 33 ::: Train
[step 1075/1075] loss: 2.472, nll loss: 1.477, ppl: 2.784, time: 244.8s

Epoch 33 ::: Validation
nll loss: 1.580, ppl: 3.006, best ppl: 3.006
[*] Model is saved in 'models/de-en/model33.pth'.
[*] Best model is changed!

Epoch 34 ::: Train
[step 1075/1075] loss: 2.457, nll loss: 1.460, ppl: 2.751, time: 239.3s

Epoch 34 ::: Validation
nll loss: 1.505, ppl: 3.000, best ppl: 3.000
[*] Model is saved in 'models/de-en/model34.pth'.
[*] Best model is changed!

Epoch 35 ::: Train
[step 1075/1075] loss: 2.442, nll loss: 1.443, ppl: 2.719, time: 239.6s

Epoch 35 ::: Validation
nll loss: 1.575, ppl: 2.979, best ppl: 2.979
[*] Model is saved in 'models/de-en/model35.pth'.
[*] Best model is changed!

Epoch 36 ::: Train
[step 1075/1075] loss: 2.428, nll loss: 1.427, ppl: 2.689, time: 229.5s

Epoch 36 ::: Validation
nll loss: 1.500, ppl: 2.990, best ppl: 2.979
[*] Model is saved in 'models/de-en/model36.pth'.
```

## 4.4 实验结果及分析

生成所有的输出结果后，删除掉子词划分，使用 fairseq score 计算 BLEU，最后在测试集上面进行 BLUE 计算所得 BLUE4=32.76。翻译结果样例如下：

样例一

输入：wir durchlaufen initiations rituale .

标准翻译：we go through initiation rites .

实际翻译：we go through initiation rites .

样例二

输入：all diese völker lehren uns , dass es andere möglichkeiten gibt , zu denken , andere arten sich selbst in der erde zu orientieren .

标准翻译：all of these peoples teach us that there are other ways of being , other ways of thinking , other ways of orienting yourself in the earth .

实际翻译：all these people teach us that there are other ways of thinking , different ways of orienting on earth .

样例三

输入：wir müssen uns mit der unerbittlichen trennung des todes befassen , also sollte es uns nicht überraschen , dass wir alle singen , wir alle tanzen , wir alle haben kunst .

标准翻译 we have to deal with the inexorable separation of death , so it shouldn't surprise us that we all sing , we all dance , we all have art .

实际翻译：we have to deal with the unstoppable separation of death , and therefore it shouldn't surprise us that we all sing , dancing and art .

样例四：

输入：aber was interessant ist , ist die einzigartige kadenz des stückes , der rhythmus des tanzes in jeder kultur .

标准翻译：but what 's interesting is the unique cadence of the song , the rhythm of the dance in every culture .

实际翻译：but what 's interesting is the unique inflection of the song , the rhythm of dance in every culture .

样例五：

输入：und ob es der penan in den wäldern von borneo ist , oder voodoo acolytes in haiti , oder die kriegler in der kaisut von nordkenia , der curandero in den bergen der wüste , oder einer karavanavanavien in der mitte der sahara .

标准翻译：and whether it is the penan in the forests of borneo , or the voodoo acolytes in haiti , or the warriors in the kaisut desert of northern kenya , the curandero in the mountains of the andes , or a caravanserai in the middle of the sahara -- this is incidentally the fellow that i traveled into the desert with a month ago -- or indeed a yak herder in the slopes of qomolangma , everest , the goddess mother of the world .

实际翻译: it doesn't matter if it's the penan in the borneo forests , or the voodoo acolyths in haiti , or the warriors in the caisut-desert of northern kenya , the curandero in the middle of a carawana desert .

#### 样例六

输入: und das ist eine idee , wenn sie darüber nachdenken , kann sie nur mit hoffnung füllen .

标准翻译: and this is an idea , if you think about it , can only fill you with hope .

实际翻译: and that is an idea that , when you think about it , can only fulfill you with hope .

#### 样例七

输入: nun , zusammen bilden die myriaden kulturen der welt ein netz aus spirituellem leben und einem kulturellen leben , das den planeten umschließt , und das wohlbefinden der welt genauso wichtig ist wie das biologische netz des lebens , das sie als biosphäre kennen .

标准翻译: now , together the myriad cultures of the world make up a web of spiritual life and cultural life that envelops the planet , and is as important to the well-being of the planet as indeed is the biological web of life that you know as a biosphere .

实际翻译: together , the countless cultures of the world form a web of spiritual and cultural life that surrounds the earth and is as important to the welfare of the earth as the biological life web you know as a biosphere .

#### 样例八

输入: und man könnte sich dieses kulturelle netz als ethnosphäre vorstellen , und man könnte die ethnosphäre als die summe aller gedanken und träume definieren , mythen , ideen , inspirierungen , intuitionen , die von der menschlichen vorstellungskraft seit dem beginn des bewusstseins bestimmt werden .

标准翻译: and you might think of this cultural web of life as being an ethnosphere , and you might define the ethnosphere as being the sum total of all thoughts and dreams , myths , ideas , inspirations , intuitions brought into being by the human imagination since the dawn of consciousness .

实际翻译: you can think of this cultural lifestyle web as an ethnosphere , and ethnosphere can be defined as the total amount of all thoughts and dreams , myths , inspiration and intuitions that have been defined by human imagination since the beginnings of consciousness .

#### 样例九

输入: die ethnosphäre ist das große vermächtnis der menschheit .

标准翻译: the ethnosphere is humanity's great legacy .

实际翻译: the ethnosphere is the great legacy of humanity .

#### 样例十

输入: wissen sie , eine der intensiven vergnügen des reisens und eines der lieferungen ethnografischer forschung ist die möglichkeit , unter denen zu leben , die nicht die alte art und weise vergessen haben , die immer noch ihre vergangenheit im wind spüren , es mit steinen polierten regen anfassen , es schmeckt in den bitteren blättern .

标准翻译: you know , one of the intense pleasures of travel and one of the delights of



ethnographic research is the opportunity to live amongst those who have not forgotten the old ways , who still feel their past in the wind , touch it in stones polished by rain , taste it in the bitter leaves of plants .

实际翻译: you know , one of the great sensitivity in traveling , and one of the pleasures for ethnographic research is to live with the people who can still remember the old days , who still feel their past in the wind , to touch them on the rain shiny stones , in the plants .

在翻译过程中也出现了翻译的没有逻辑的句子, 比如图 7:

```
simply knowing that jaguar shamanaş are still traveling beyond the milky way , or the importance of
the myths of inuit is still very successful , or that in the himalayas , the buddhists are still
tracking the breath of the dharma , which means the most unfolding of the world , is to the most
inuit , to the most inuit , to the most unlogic of us , to the most revelocity of us , to the most
revelocity , of the most uncidharma , of us , of the most unce of the most untarian , of the most
untact of us , of the world , of us , of the most untantaneal life , of us , of being , of us , of
being , of being , of the world , is still , of the most unfolding of us , of the most unfolding of
us , of us , of us , of us , of us , of the most unfolding of the world , of the most unfolding of
the world , of the world , is still , of us , of the world , of us , of us , of us , of the most of
being in the most of the most of the most of the most of the most of us , of us , of the
and of course , we all share the same adaptation .
```

图 7 错误翻译

根据翻译样例可以看出翻译的句子相比较于标准答案还是差很多的, 也存在较多的语法错误。由于本模型只实现了贪婪搜索, 应该是在推理时选择了错误的内容, 导致后续的内容也错误。

## 5 心得体会

在语言分析和机器翻译这门课上, 我学习到了很多知识。首先是在肖老师和胡驰学长的讲解下度过了理论课的内容和知识, 学习了机器翻译和语言建模的基础知识, 包括统计机器翻译和神经机器翻译, 尤其是基于循环神经网络的模型和基于自注意力机制的模型, 这在我之前都没有深入去了解, 使我的基础知识更加牢固。

在机器翻译的作业中, 基于 Transformer 模型, 实现德英翻译。由于个人原因, 前期没有充分利用好时间, 导致最后模型训练的结果也不是特别的满意, 在推理阶段也只是完成了贪婪搜索方法, 事实证明贪婪搜索算法确实效果不好, 尤其在推理比较长的源语言句子上。并且在训练过程中遇到了许多问题, 比如训练收敛的速度很慢很慢; 使用 nohup 指令出现问题 OSError: [Errno 25] Inappropriate ioctl for device, 而且只是对这个模型训练有问题, 查了许多资料修改之后也没有解决, 只能老老实实挂着训练了。在模型结构上, 刚开始按照论文给出的计算公式来写, 发现计算完结果不对, 是中间的一些计算越界了, 超出 float 范围, 根据 github 别人的代码改了计算公式, 也手动推导了一下, 确定计算公式是等价的。

最后, 通过本次实验, 我也深刻的体会到了自己的不足, 自己之前就是查找 baseline, 修改, 报错, 修改, 再报错, 最后一步一步的完成实验。如果不上手实验的话, 是不会彻底搞懂的, 只有上手之后才会发现很多地方自己不会, 需要借鉴和比较别人的代码。也能发现很多差距, 甚至之前得过且过的内容也能豁然开朗。同时刚进入这个领域学习, 发现自己真的是亿丢丢的菜, 自己的文献阅读量太少, 仅仅针对于一篇论文去实现论文的话, 有些地方写的就不是很好, 我也明白了理论结合实际的重要性, 以后要多读多做, 也要解决之前出现的问题。在往日的学习中, 得到了许晨学长以及小组成员的大量帮助, 也让我发现了自身存在的诸多问题, 自己也在失去自信和建立自信中不断的交替, 但是有一点可以确定的是我越来越喜欢我的研究小组, 越来越觉得实验室是一个温暖的大家庭, 我也越来越爱我的研究方向, 并愿意为之而努力。