تمرین شماره ۸ شبکه عصبی

۱. ابتدا مجموعه داده مورد نظر را لود می‌کنیم و دو دسته داده آزمون و داده اعتبار سنجی که داخل خود مجموعه‌داده تعریف شده است تفکیک می‌کنیم:

```
ds, metadata = tfds.load('ted_hrlr_translate/pt_to_en', with_info=True,
                         as_supervised=True)
train_set, val_set = ds['train'], ds['validation']
```

byte pair از tfds.deprecated.text.SubwordTextEncoder از TextEncoder ارث‌بری می‌کند و به علت اینکه از encoding استفاده می‌کند کدکننده‌ها برگشت‌پذیر هستند:

```
tokenizer_en = tfds.deprecated.text.SubwordTextEncoder.build_from_corpus(
    (en.numpy() for pt, en in train_set), target_vocab_size=2**13)

tokenizer_pt = tfds.deprecated.text.SubwordTextEncoder.build_from_corpus(
    (pt.numpy() for pt, en in train_set), target_vocab_size=2**13)
```

به این صورت یک توکنایزر کلمه مخصوص از مجموعه داده آموزشی ایجاد کرده در اینجا target_covab_size برای تعیین کردن مقدار تقریبی دیکشنری است که میخواهیم بسازیم.

خروجی گرفته شده به صورت زیر می باشد:

```
Tokenized string is [7915, 128, 13, 7, 6512, 7863, 1057, 48, 7911, 1785, 3662, 6939, 104, 5, 7900, 321, 4230, 7935, 7877]
The original string: This is a translation machine from Portuguese to English.
```

به دلیل ساختار decoder_encoder ها یک نشانه شروع و پایان به ورودی و خروجی اضافه می‌کنیم.

می‌خواهیم از Dataset.map برای اعمال این تابع روی هر عنصر مجموعه داده استفاده کنیم. Dataset.map به صورت گرافی اعمال می‌شود.بنابراین نمی توانیم مستقیماً این تابع را map. کنیم: باید آن را در یک tf.py_function بپیچید.

```python
def start_end(lang1, lang2):
  lang1 = [tokenizer_pt.vocab_size] + tokenizer_pt.encode(
      lang1.numpy()) + [tokenizer_pt.vocab_size+1]

  lang2 = [tokenizer_en.vocab_size] + tokenizer_en.encode(
      lang2.numpy()) + [tokenizer_en.vocab_size+1]

  return lang1, lang2
```

```python
def tf_start_end(pt, en):
  result_pt, result_en = tf.py_function(start_end, [pt, en], [tf.int64, tf.int64])
  result_pt.set_shape([None])
  result_en.set_shape([None])

  return result_pt, result_en
```

همچنین طول هر جمله را با عملگر and منطقی کنترل می‌کنیم:

```python
MAX_LENGTH = 30
def maximum_length(x, y, max_length=MAX_LENGTH):
    return tf.logical_and(tf.size(x) <= max_length,
                          tf.size(y) <= max_length)
```

سپس روی مجموعه آموزش و مجموعه اعتبارسنجی فیلترها را اعمال کرده، ترتیب مجموعه‌داده آموزش را به صورت تصادفی تغییر میدهیم.و هردو مجموعه را به batch های ۳۲ تایی تقسیم کرده:

```python
#filter train set
train_ds = y=train_set.map(tf_start_end)
train_ds = train_ds.filter(maximum_length)

train_ds = train_ds.cache()
train_ds = train_ds.shuffle(BUFFER_SIZE).padded_batch(BATCH_SIZE)
train_ds = train_ds.prefetch(tf.data.experimental.AUTOTUNE)

#filter val set
val_ds = val_set.map(tf_start_end)
val_ds = val_ds.filter(maximum_length).padded_batch(BATCH_SIZE)
```

۲. Positional encoding اضافه می شود تا به مدل اطلاعاتی در مورد موقعیت نسبی کلمات در جمله بدهد.

بردار positional encoding به بردار embedding اضافه می‌شود. Embedding نشان دهنده یک توکن در یک فضای d بعدی هستند که در آن توکن‌های با معنای مشابه به یکدیگر نزدیکتر خواهند بود. اما موقعیت نسبی کلمات را در یک جمله رمزگذاری نمی‌کنند.

بنابراین پس از افزودن positional encoding، کلمات بر اساس شباهت معنی و موقعیت آنها در جمله، در فضای d بعدی به یکدیگر نزدیکتر خواهند شد.

تابع sin به ایندکس‌های زوج اعمال می‌شود و تابع cos به ایندکس های فرد اعمال می‌شود.

```python
def get_angles(pos, i, dim):
  angle_rates = pos * (1 / np.power(10000, (2 * (i//2)) / np.float32(dim)))
  return angle_rates
```

```python
def positional_encoding(position, dim):
  angle_rads = get_angles(np.arange(position)[:, np.newaxis],
                          np.arange(dim)[np.newaxis, :],
                          dim)
  angle_rads[:, 0::2] = np.sin(angle_rads[:, 0::2])
  angle_rads[:, 1::2] = np.cos(angle_rads[:, 1::2])
  pos_encoding = angle_rads[np.newaxis, ...]
  return tf.cast(pos_encoding, dtype=tf.float32)
```

برای تضمین اینکه مدل با padding مانند ورودی رفتار نمیکند، padding که با ۱ نمایش داده می‌شود را در خروجی به ۱ نگاشت می‌کنیم:

```python
def padding_mask(seq):
  seq = tf.cast(tf.math.equal(seq, 0), tf.float32)
  return seq[:, tf.newaxis, tf.newaxis, :]
```

۳.

- با توجه به مقاله attention is all you need از بهینه ساز Adam با 10^{-9} با $\beta_1 = 0.9, \beta_2 = 0.98, \varepsilon = 10^{-9}$ استفاده کردیم.و نرخ یادگیری را در طول دوره‌ی آموزش، طبق فرمول زیر تغییر داده:
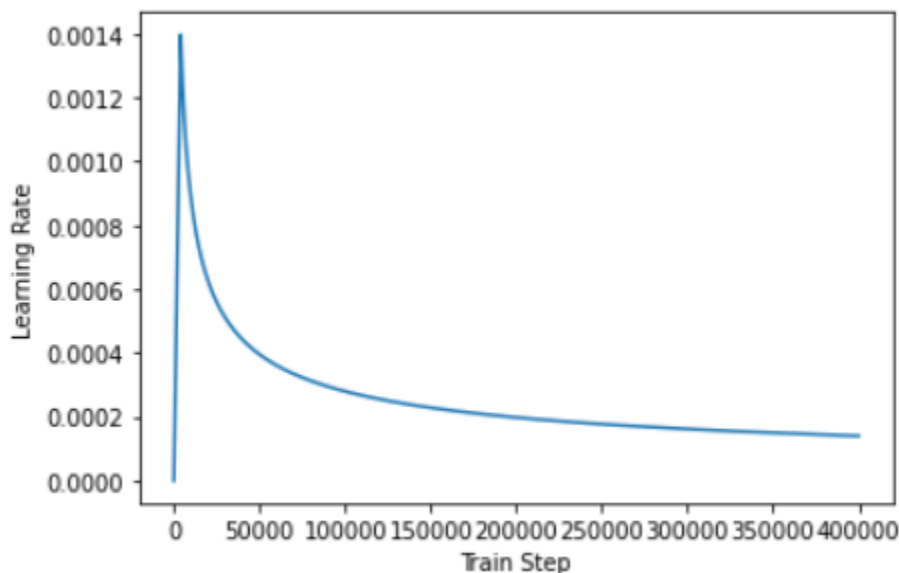
$$lrate = d_{\text{model}}^{-0.5} \cdot \min(step\_num^{-0.5}, step\_num \cdot warmup\_steps^{-1.5})$$

پیاده سازی learning rate به صورت زیر می‌باشد:

```python
class learning_rate(tf.keras.optimizers.schedules.LearningRateSchedule):
  def __init__(self, d_model, warmup_steps=4000):
    super(learning_rate, self).__init__()

    self.d_model = d_model
    self.d_model = tf.cast(self.d_model, tf.float32)

    self.warmup_steps = warmup_steps

  def __call__(self, step):
    arg1 = tf.math.rsqrt(step)
    arg2 = step * (self.warmup_steps ** -1.5)

    return tf.math.rsqrt(self.d_model) * tf.math.minimum(arg1, arg2)
```
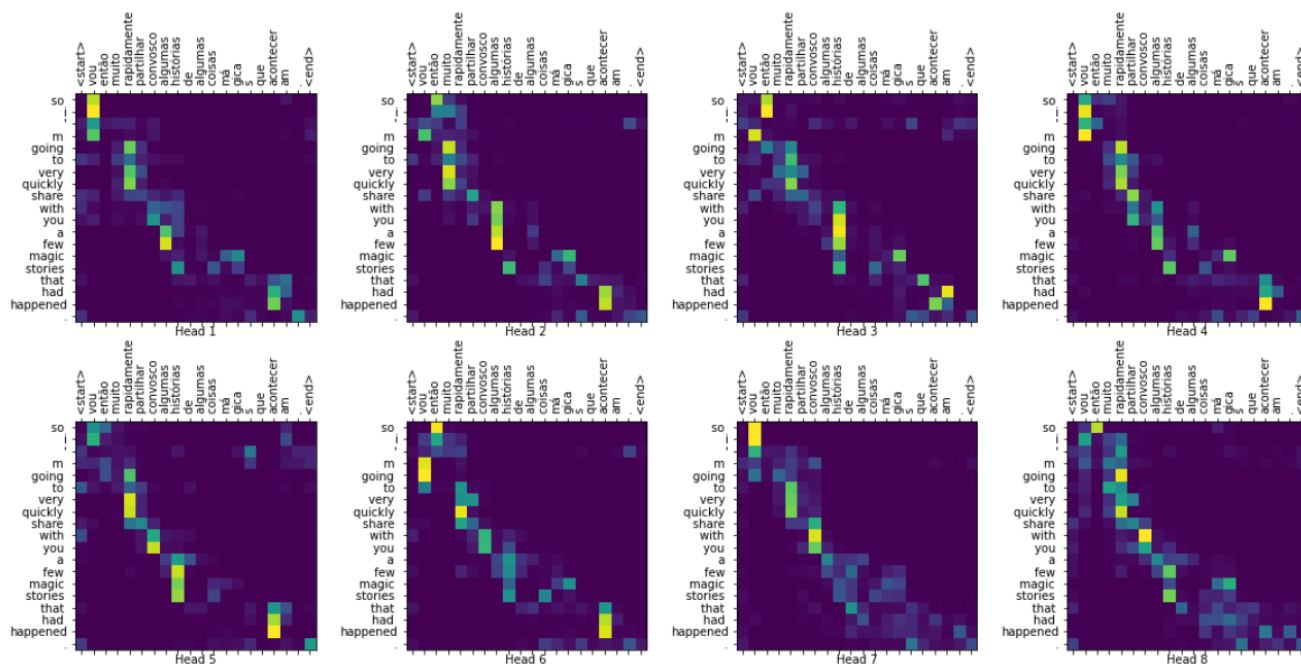
سپس از learning rate تعریف شده استفاده می‌کنیم برای adam optimizer همراه با مقادیر پارامترهای گفته شده:

```
lr_rate = learning_rate(d_model)
optimizer = tf.keras.optimizers.Adam(lr_rate, beta_1=0.9, beta_2=0.98,
                                     epsilon=1e-9)
```


- Heatmap برای جمله داده شده به شکل زیر می‌باشد:

همانطور که مشاهده می‌شود کلمات نظیر و مرتبط دارای attention weight بالاتری می‌باشند.

- ۲۰ نمونه تولید شده به صورت زیر می‌باشد:

Input: na verdade , nos anos em que eu recolhi dados , apenas uma experiÃªncia chegou perto de ser perfeita .
Predicted translation: in fact , in the years i sampled data , just a prevented some of prevention of completely perfect it .
Correct translation is : in fact , in the years that i gathered data , only one experience came close to being the perfect one .

Input: e eu era o único nos Estados Unidos escrevendo sobre derramamentos de óleo.
Predicted translation: and i was the only way to get involved that it was using his names called rathe 's oil .
Correct translation is : and i was the only one in the united states writing on oil spills .

Input: tinham comido peixe com batatas fritas ?
Predicted translation: did you hear fish with their lives fried ?
Correct translation is : did they eat fish and chips ?

Input: com essas pessoas eu ainda saio .
Predicted translation: these people still pass it out .
Correct translation is : these people i still do hang out with .

Input: mas aqui , elas simplesmente saem e voltam .
Predicted translation: but just these things just come back and back .
Correct translation is : but here , just they leave and come back .

Input: para alguns de nós pode ser música. para alguns de nós pode ser comida .
Predicted translation: for some of us can be music. for some of us can be food .
Correct translation is : for some of us it might be music . for some of us it might be food .

Input: aqui está o caso de negócios para a sustentabilidade.
Predicted translation: here 's the business case for sustainability .reverse .
Correct translation is : here is the business case for sustainability .

Input: este é o primeiro livro que eu fiz.
Predicted translation: this is the first book i did .
Correct translation is : Real translation: this is the first book i've ever done.

Input: os meus vizinhos ouviram sobre esta ideia.
Predicted translation: my neighbors had gotten about this idea .
Correct translation is : Real translation: and my neighboring homes heard about this idea .

Input: este é um problema que temos que resolver.
Predicted translation: so this is a problem that we must solve our problem in a problem is that we must solve it .
Correct translation is : Real translation: this is a problem we have to solve .

Input: stabelecemos 2020 como o ano da nossa meta para zero , para chegar ao topo do cume do monte sustentabilidade .
Predicted translation: we stereotyped 20 about our own half-version , just to almost at the top of sustainability are sustained .
correct translation is : we 've set 2020 as our target year for zero , for reaching the top , the summit of mount sustainability .

Input: eles estão em constante atualização e, em alguns casos, o governo ajuda.
Predicted translation: they 're in constant current case , and in some cases , what is scientists at all help bring humans from .
Correct translation is : they are constantly upgrading , and in a few cases , the government helps .

Input: eu fui para a china no dia 15 de fevereiro de 2006 .
Predicted translation: i went to china on the 15 day of 2004 .
Correct translation is : i made it to china on february 15 , 2006 .

Input: vou então muito rapidamente partilhar convosco algumas histórias de algumas coisas mágicas que aconteceram.
Predicted translation: so i 'm going to very quickly share with you a few magic stories that had happened .
Correct translation is : Real translation: so i 'll just share with you some stories very quickly of some magical things that have happened .

Input: mas aqui , elas simplesmente saem e voltam .
Predicted translation: but just these things just come back and back .
Correct translation is : but here , just they leave and come back .

Input: mas claro, corda não tem nada a ver com eles.
Predicted translation: but of course , it 's nothing to do with them .
Correct translation is : but of course , rope has nothing to do with them .

Input: as pessoas têm aptidões muito diferentes.
Predicted translation: people have very different susceptibilities .
Correct translation is : people have very different aptitudes .

Input: eles pensam nele como um líder.
Predicted translation: they think of them as a leader .
Correct translation is : they think of him as a leader .

Input: eles são exatamente o que parecem .
Predicted translation: they exactly what they seem .
Correct translation is : they're exactly what they sound like .

Input: outra vez .
Predicted translation: again , this one again .
Correct translation is : and again .